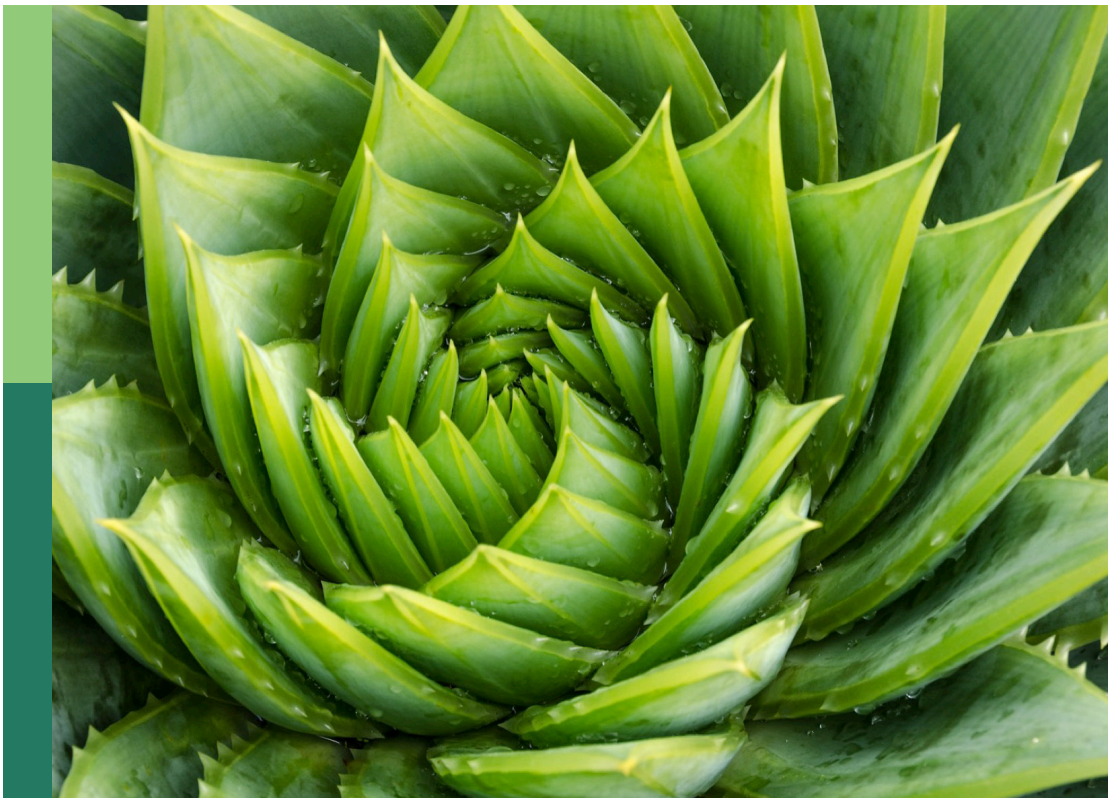
frontiers | Research Topics

Improving yield and quality of cereal crops: exploring and utilizing genes for green and efficient traits

Edited by
Hao Zhou, Yuqing He and Ming Luo

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-6495-0
DOI 10.3389/978-2-8325-6495-0

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Improving yield and quality of cereal crops: exploring and utilizing genes for green and efficient traits

Topic editors

Hao Zhou — Sichuan Agricultural University, China

Yuqing He — Huazhong Agricultural University, China

Ming Luo — Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

Citation

Zhou, H., He, Y., Luo, M., eds. (2025). *Improving yield and quality of cereal crops: exploring and utilizing genes for green and efficient traits*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6495-0

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 10 June 2025
DOI 10.3389/fpls.2025.1617702

OPEN ACCESS

EDITED AND REVIEWED BY
Peng Wang,
Jiangsu Province and Chinese Academy of Sciences, China

*CORRESPONDENCE
Hao Zhou
zhouhao666@foxmail.com

RECEIVED 24 April 2025
ACCEPTED 14 May 2025
PUBLISHED 10 June 2025

CITATION
Zhou H, He Y and Luo M (2025) Editorial: Improving yield and quality of cereal crops: exploring and utilizing genes for green and efficient traits. *Front. Plant Sci.* 16:1617702. doi: 10.3389/fpls.2025.1617702

Editorial: Improving yield and quality of cereal crops: exploring and utilizing genes for green and efficient traits

Hao Zhou[1]*, Yuqing He[2] and Ming Luo[3]

[1]State Key Laboratory of Cop Gene Exploration and Utilization in Southwest China, Rice Research Institute, Sichuan Agricultural University, Chengdu, China, [2]National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, China, [3]Agriculture and Food, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, ACT, Australia

KEYWORDS

cereal crop, green and efficient trait, gene discovery, molecular network, breeding utilization

Editorial on the Research Topic
Improving yield and quality of cereal crops: exploring and utilizing genes for green and efficient traits

In the face of escalating global challenges—population growth, climate change, and limited arable land—enhancing the yield and quality of cereal crops has become a cornerstone of sustainable agriculture. This Research Topic delves into the genetic and molecular mechanisms underpinning stress resilience, nutritional improvement, and breeding innovation in staple crops such as rice (*Oryza sativa*), wheat (*Triticum aestivum*), and maize (*Zea mays*). By integrating cutting-edge genomic technologies, transcriptomic analyses, and artificial intelligence-driven methodologies, the studies presented here illuminate pathways toward precision breeding for green and efficient traits (Figure 1). Below, we contextualize these contributions within broader scientific and agricultural landscapes.

Unlocking stress resilience: from genes to fields

Environmental stressors, including soil compaction, pathogen attacks, and herbicide pressure, persistently threaten cereal production. The featured studies leverage multi-omics approaches to decode plant responses to these challenges. For instance, genome-wide association studies (GWAS) in rice (Klinsawang et al.) identified 'F-box' genes (*Os10g0124700*, *Os10g0126600*, and *Os10g0128200*) that modulate ethylene signaling and root architecture under compacted soil conditions. These findings offer actionable targets for breeding rice varieties with enhanced root plasticity, a trait critical for nutrient acquisition in degraded soils. Similarly, transcriptomic profiling of Fusarium head blight (FHB)-resistant wheat landrace *Wuyangmai* revealed miRNA-mediated regulation of glutathione metabolism and phenylpropanoid biosynthesis, pathways pivotal for oxidative stress mitigation and pathogen defense (Wu et al.). Such insights not only

deepen our understanding of plant immunity but also highlight miRNAs as novel tools for engineering disease-resistant crops.

Equally compelling is the GWAS-based dissection of herbicide resistance in rice, which uncovered geographically localized alleles (e.g., *RGlu6* and *RGly8*) in European *japonica* varieties (Xu et al.). These results underscore the importance of preserving genetic diversity in germplasm collections to address region-specific agricultural challenges. Collectively, these studies exemplify how genetic diversity and molecular networks can be harnessed to fortify crops against biotic and abiotic threats.

Quality traits: balancing nutrition and palatability

Crop quality is a multidimensional trait shaped by complex metabolic and regulatory networks. The functional characterization of *OsG6PGH1* in rice exemplifies this complexity (Peng et al.). By interacting with *OsAAP6*, this gene enhances protein body formation and grain protein content while reducing chalkiness, thereby improving both nutritional and sensory properties. Notably, *OsG6PGH1* also contributes to salt stress tolerance, illustrating the pleiotropic roles of genes at the intersection of quality and resilience. Another breakthrough comes from the triple mutant *sbe2b/sbe1/OE-Wxa*, which elevates resistant starch (RS) content to 4.63%—a fivefold increase over wild-type levels—without compromising yield (Chen et al.). This achievement demonstrates the power of stacking mutations in starch biosynthesis genes to achieve synergistic improvements in health-promoting traits and agronomic performance.

Innovations in breeding technology: from data to decisions

Translating genetic discoveries into breeding outcomes requires robust technological frameworks. The development of *TAL-SRX*, a machine learning model for KASP primer evaluation, exemplifies this transition (Chen et al.). By integrating deep learning (Transformer, LSTM) and traditional algorithms, this tool achieves 92.83% accuracy in genotyping, significantly accelerating marker-assisted selection. Similarly, genomic prediction models (e.g., rrBLUP, LightGBM) applied to maize hybrids in Southwest China identified Reid+ × Suwan+ as the optimal heterotic pattern, aligning computational predictions with empirical breeding outcomes (Xiang et al.). These advances highlight the transformative potential of AI in democratizing precision breeding.

Complementing these tools is the novel strategy for constructing "trait-customized core collections" in winter wheat (Berkner et al.). By balancing phenotypic extremes and genetic diversity, this approach maximizes statistical power in association studies, enabling efficient mining of rare alleles for traits like yellow rust resistance. Such methodologies are indispensable for leveraging genebank resources in the era of big data.

The trade-off conundrum: immunity vs. agronomic performance

A recurring theme in crop improvement is the trade-off between stress resistance and yield or quality. The rice *spl-A*mutant, harboring a *LRD6-6* mutation, epitomizes this dilemma

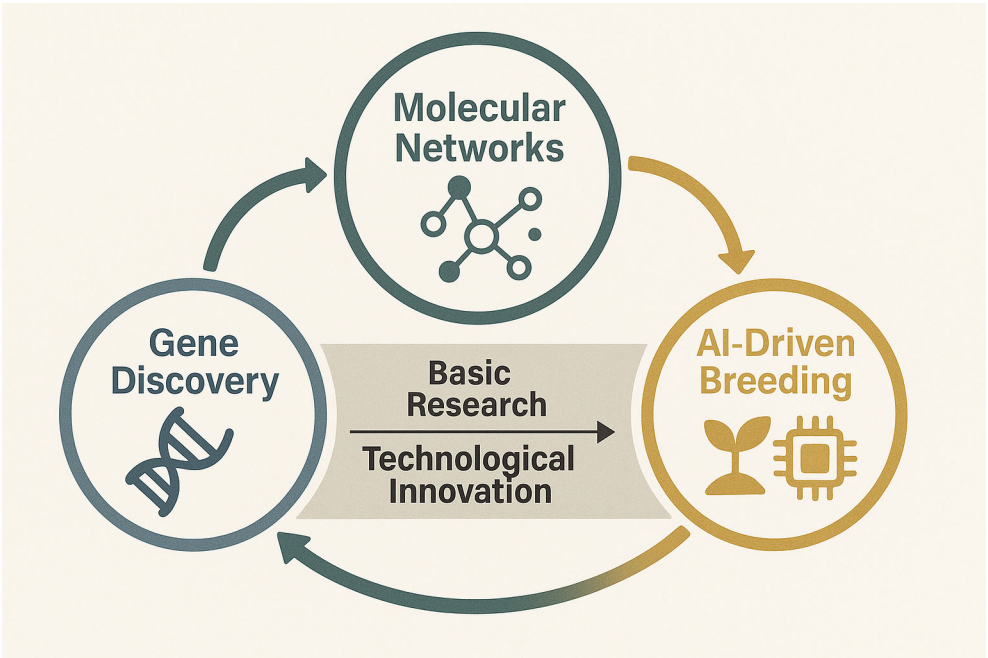


FIGURE 1
A conceptual diagram of this topic.

(Yin et al.). While the mutation confers broad-spectrum disease resistance, it disrupts vesicle trafficking via ESCRT-III interactions, leading to compromised grain quality. This study serves as a cautionary tale: single-gene interventions may incur unintended costs, necessitating strategies such as tissue-specific expression or epistatic modulation to decouple desirable traits from fitness penalties.

Future directions: integrating disciplines for sustainable agriculture

The studies in this Research Topic collectively underscore the importance of interdisciplinary integration—genomics, bioinformatics, and agronomy—to address food security challenges. Future efforts should prioritize:

1. Precision editing: Combining CRISPR-based allele replacement with multi-omics data to fine-tune trait expression.
2. Dynamic network analysis: Elucidating crosstalk between stress-response, metabolic, and developmental pathways.
3. Climate-resilient prediction models: Leveraging AI to predict genotype × environment × management (G×E×M) interactions.

By bridging fundamental research and translational breeding, we can accelerate the development of "smart" crop varieties that meet the dual demands of productivity and sustainability.

Author contributions

HZ: Writing – original draft, Writing – review & editing. YH: Writing – review & editing. ML: Writing – review & editing.

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.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript.

We used DeepSeek to correct and polish the language of this article.

Publisher's note

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 July 2024
DOI 10.3389/fpls.2024.1436998

OPEN ACCESS

EDITED BY
Hao Zhou,
Sichuan Agricultural University, China

REVIEWED BY
Xianjin Qiu,
Yangtze University, China
Wen Yao,
Henan Agricultural University, China

*CORRESPONDENCE
Bo Peng
pengbo@xynu.edu.cn
Yanfang Sun
370380460@qq.com
Quanxiu Wang
wangquanxiu@xynu.edu.cn

RECEIVED 23 May 2024
ACCEPTED 25 June 2024
PUBLISHED 10 July 2024

CITATION
Peng B, Liu Y, Qiu J, Peng J, Sun X, Tian X, Zhang Z, Huang Y, Pang R, Zhou W, Zhao J, Sun Y and Wang Q (2024) *OsG6PGH1* affects various grain quality traits and participates in the salt stress response of rice.
Front. Plant Sci. 15:1436998.
doi: 10.3389/fpls.2024.1436998

OsG6PGH1 affects various grain quality traits and participates in the salt stress response of rice

Bo Peng[1*], Yan Liu[1], Jing Qiu[1], Jing Peng[2], Xiaoyu Sun[1], Xiayu Tian[1], Zhiguo Zhang[3], Yaqin Huang[4], Ruihua Pang[1], Wei Zhou[1], Jinhui Zhao[1], Yanfang Sun[1*] and Quanxiu Wang[1*]

[1]College of Life Sciences, Xinyang Normal University, Xinyang, China, [2]College of Agronomy, Xinyang Agriculture and Forestry University, Xinyang, China, [3]Henan Lingrui Pharmaceutical Company Limited, Xinyang, China, [4]School of Pharmacy, Xinyang Agriculture and Forestry University, Xinyang, China

Cytoplasmic 6-phosphogluconate dehydrogenase (G6PGH) is a key enzyme in the pentose phosphate pathway that is involved in regulating various biological processes such as material metabolism, and growth and development in plants. However, it was unclear if *OsG6PGH1* affected rice grain quality traits. We perform yeast one-hybrid experiments and reveal that *OsG6PGH1* may interact with *OsAAP6*. Subsequently, yeast *in vivo* point-to-point experiments and local surface plasmon resonance experiments verified that *OsG6PGH1* can bind to *OsAAP6*. *OsG6PGH1* in rice is a constitutive expressed gene that may be localized in the cytoplasm. *OsAAP6* and protein-synthesis metabolism-related genes are significantly upregulated in *OsG6PGH1* overexpressing transgenic positive endosperm, corresponding to a significant increase in the number of protein bodies II, promoting accumulation of related storage proteins, a significant increase in grain protein content (GPC), and improved rice nutritional quality. *OsG6PGH1* positively regulates amylose content, negatively regulates chalkiness rate and taste value, significantly affects grain quality traits such as appearance, cooking, and eating qualities of rice, and is involved in regulating the expression of salt stress related genes, thereby enhancing the salt-stress tolerance of rice. Therefore, *OsG6PGH1* represents an important genetic resource to assist in the design of high-quality and multi-resistant rice varieties.

KEYWORDS

rice, *OsG6PGH1*, grain quality, proteome, salt stress

Introduction

Rice (*Oryza sativa* L.) is a globally important food crop, upon which more than half of the world's and more than two-thirds of China's populations rely as a dietary staple (Wang et al., 2018; Zhao et al., 2020; Li et al., 2022a). With continued improvement in societal living standards, the demand for high-quality rice is increasing (Zhou et al., 2019). Rice

grain quality is measured in terms of its physical, chemical, and nutritional characteristics during growth, processing, storage, and consumption. These characteristics affect the nutritional value, taste, and processing performance of rice, and its appearance, nutritional, milling, processing, cooking, and eating qualities (Li et al., 2018; Custodio et al., 2019; Zhou et al., 2020; Zhao et al., 2022a).

Rice appearance is mainly affected by grain size and shape, chalkiness, and translucency. Nutritional quality is mostly affected by protein, starch, fatty and amino acid, vitamin, and mineral contents. The protein in rice is high in quality, and easily digested and absorbed in the gut (Peng et al., 2021; Li et al., 2022b; Gopika et al., 2022; Riaz et al., 2018). Grain protein is the second largest nutritional component after carbohydrates, and its content largely affects nutritional quality. An increase in grain protein content (GPC) improves the nutritional quality of rice. Based on solubility, grain proteins can be divided into glutelins, albumins, prolamins, and globulins (He et al., 2013; Peng et al., 2014; Chen et al., 2018). Milling and processing qualities mainly affect the performance of rice during processing, and are based on percentages of brown, milled, and head rice (Balindong et al., 2018). The brown rice rate directly affects processing quality, and indirectly its edible quality (Tong et al., 2019). The head rice rate is a grading indicator for grain quality. Cooking and eating qualities are the sensory and physicochemical properties of rice during cooking (e.g., extensibility, water absorption, swelling, gelatinization, color, form, aroma, and palatability) (Zhou et al., 2019; Shi et al., 2022), and are measured using amylose content, gelatinization temperature, and gel consistency (Zhou et al., 2019; Bao et al., 2023). Aspects of the cooking and eating quality are of greatest concern to consumers.

Amino acid permease (AAP) belongs to a family of amino acid transporters. *AAPs* play important roles in amino acid transport throughout plant tissues (Tegeder and Hammes et al., 2018). A total of 19 *OsAAPs* have been identified from rice, and most of them have been cloned and functionally identified. Of them, *OsAAP1* and *OsAAP4* regulate absorption and redistribution of neutral amino acids, thereby regulating rice growth, development, and yield (Ji et al., 2020; Fang et al., 2021; Erinaldo et al., 2022); *OsAAP3* and *OsAAP5* mainly transport alkaline amino acids, and have a negative effect on rice tillering and yield (Lu et al., 2018; Wang et al., 2019); *OsAAP6* positively regulates GPC, thereby affecting the nutritional quality of rice (Peng et al., 2014); *OsAAP7* and *OsAAP16* can transport most amino acids (excepting aspartic acid and β-alanine) (Taylor et al., 2015); *OsAAP8* mutation significantly increases GPC and amino acid contents (Peng et al., 2024); *OsAAP10* and *OsAAP11* regulate starch and protein contents in rice grains by affecting the transport of acidic and neutral amino acids, respectively (Wang et al., 2020; Yang et al., 2023a); *OsAAP13* is mainly involved in amino acid transportation in above-ground parts of rice, and plays an important role in rice growth, development, and stress regulation (Nie et al., 2023); *OsAAP14* is involved in nitrogen metabolism and positively regulates rice axillary bud growth, tillering, and yield (Yang et al., 2022); *OsAAP15* affects rice yield by regulating panicle branch development (Yang et al., 2023b); and *OsAAP17*-related alleles positively regulate yield per plant in *indica* rice (Itishree et al., 2023). Thus, *OsAAPs* play important roles in regulating rice yield and quality traits.

The pentose phosphate pathway (PPP) is important for sugar metabolism in plants; its main physiological function is to produce reduced NADPH and generate pentose phosphate to participate in nucleic acid metabolism. PPP is the main source of NADPH in eukaryotic cells, and is involved in regulating metabolic intermediates in the biosynthesis process, which can participate in the synthesis of substances such as fatty and amino acids (Wang et al., 2017; Li et al., 2022c). 6-phosphogluconate dehydrogenase (G6PGH/6PGDH) is a key enzyme in the PPP, catalyzing the production of 5-phosphoribose and NADPH from 6-phosphogluconate. *G6PGH* has been successfully isolated and cloned from plants such as soybeans, corn, tomatoes, apples, *Arabidopsis*, rice, and wheat (Hanks et al., 1983; Tanksley and Kuehn, 1985; Bailey-Serres and Nguyen, 1992; Huang et al., 2003; Hölscher et al., 2016; Su et al., 2016). *G6PGH* participates in the plant's response to various abiotic stresses, and plays an important role in responses to temperature extremes, salt, heavy metals, and osmotic and other stresses. In maize, *6PGDH* can significantly increase yield under high temperature stress. After a double mutation of *6pgd1* and *6pgd2* in corn, the homozygous mutant showed delayed growth (Bailey-Serres and Nguyen, 1992); after the *PGD3* mutation, the mutant exhibited severe granule filling and defective embryonic phenotype (Ribeiro et al., 2020). Overexpression of *6PGDH* in *Arabidopsis* significantly enhances cold tolerance (Tian et al., 2021), and can accelerate growth of callus tissue in apples (Zhu et al., 2019). Transcription levels of two types of *OsG6PGH* isolated and cloned in rice (cytoplasmic 6-phosphogluconate dehydrogenase (*OsG6PGH1*) and plastid 6-phosphogluconate dehydrogenase (*OsG6PGH2*)) were significantly upregulated in rice seedlings under drought, high salt, cold, and abscisic acid treatment, indicating that *OsG6PGH* plays an important role in the response of rice to abiotic stress (Huang et al., 2003; Hou et al., 2007). However, it remained unclear if *OsG6PGH* was involved in biological processes such as rice growth, development, and substance metabolism, or if it affected grain quality.

We identified a major quantitative trait locus gene *OsAAP6* that positively regulated GPC and affected rice nutritional quality (Peng et al., 2014, 2018). To gain a deeper understanding of the molecular mechanism by which *OsAAP6* regulated GPC, we determined *OsG6PGH1* to possibly interact with it (Peng et al., 2023). We further validated the ability of *OsG6PGH1* to bind to *OsAAP6* through *in vivo* point-to-point experiments in yeast and localized surface plasmon resonance experiments. In overexpressing *OsG6PGH1* plants, we report *OsG6PGH1* as capable of regulating the expression of *OsAAP6* to affect the contents of the main storage substances in rice grains, and to have important effects on grain quality traits. Simultaneously, *OsG6PGH1* participates in biological processes such as the salt-stress response. These results are important for molecular design and breeding of high-quality and multi-resistant rice varieties.

Materials and methods

Point-to-point validation in yeast cells

We first simultaneously transformed the plasmid pGADT7-Rec2-*OsG6PGH1* carrying the target gene *OsG6PGH1* and the bait plasmid pHIS2-*OsAAP6* into yeast receptive cell Y187. Transformed yeast cells were cultured on SD-Trp-Leu solid medium at 30°C for 3–5 d (Peng et al., 2023). Select yeast monoclonal antibodies from this medium were inoculated into SD-Trp-Leu liquid medium. Cultivation was continued at 30°C for 12 h. The OD600 value of the yeast solution was adjusted for all samples to 0.4 by adding ddH_2O to ensure consistency in experimental conditions. Next, the standardized yeast solution onto SD-Leu-Trp-His+150 mmol·L^{-1} 3AT medium, and cultured at 30°C for 3–4 d to further validate the interaction between yeast clones and bait proteins (Peng et al., 2023; Zhang et al., 2024).

Localized surface plasmon resonance

Binding analysis of *OsG6PGH1* and *OsAAP6* was performed using the OpenSPR surface plasmon resonance biosensor (Nicoya Life Science Inc., Kitchener, Canada). Using a standard coupling reaction of 1-ethyl-3-(3-dimethylpropyl)-carbodiimide (EDC) and N-hydroxysuccinimide (NHS), the *OsAAP6* was fixed at a concentration of 50 μg·ml^{-1} onto the activated COOH sensor chip (Peng et al., 2023). A blocking solution was then injected into the sensor chip to reduce the possibility of non-specific binding. Then, prokaryotic total cell lysate protein solutions at concentrations of 15, 22.5, and 30 μg·ml^{-1} were introduced to the sensor chip. As the sample flows through the chip, *OsAAP6* contained in it binds specifically to previously fixed antibodies, and is captured by the sensor (Liu et al., 2019). To detect the interaction between *OsAAP6* and *OsG6PGH1*, we injected 50 μg·ml^{-1} IgG as a negative control and *OsG6PGH1* antibody as a target protein into the sensor chip. By comparing the response signals of IgG (non-specific binding) and *OsG6PGH1* antibody (with potential specific binding), it can be determined if *OsG6PGH1* has interacted with *OsAAP6*.

Creation of *OsG6PGH1* transgenic rice lines

To create mutants targeting the deletion of *OsG6PGH1* based on CRISPR/Cas9 gene-editing tool, we commissioned Wuhan Astronomical Biotechnology Co., Ltd. to conduct genetic transformation experiments. PCR and Sanger sequencing techniques were used to detect and analyze the mutant sequence (Peng et al., 2024), ensuring the successful knockout of *OsG6PGH1*.

The cDNA fragment of *OsG6PGH1* was cloned onto the intermediate vector PMD-18T to obtain the PMD-18T-*OsG6PGH1* recombinant plasmid. Sequencing validation was then performed on the recombinant plasmid to ensure the accuracy of the *OsG6PGH1* sequence. After verifying the accuracy, the target fragment was cut from PMD-18T-*OsG6PGH1* and connected to the pCMBIA1301s overexpression vector to construct the overexpression plasmid pCMBIA1301s-*OsG6PGH1*. Wuhan Astronomical Biotechnology Co., Ltd. was again entrusted to use *Agrobacterium*-mediated genetic transformation to introduce this overexpression plasmid into rice cells, to overexpress the *OsG6PGH1*.

RNA extraction and real-time quantitative polymerase chain reaction

Total RNA was extracted using TRIzol reagent according to manufacturer instructions (TAKARA) (Peng et al., 2024). A reverse transcription kit (R202-02 HyperScript III RT SuperMix for qPCR with gDNA Remover) was used to reverse transcribe RNA into cDNA. qRT-PCR was performed using 2 SYBR Green Fast qPCR Master Mix (YiFeiXue Bio Technology) on quantitative analyzer (ABI7300), with a reaction program involving a predenaturation step at 95°C for 30 s, and an amplification reaction of 40 cycles, each including 95°C denaturation for 15 s, and a 60°C annealing extension for 30 s. To accurately evaluate changes in expression of target genes, *β-actin* was used as an internal reference gene, and relative expression levels were standardized using the $2^{-\Delta\Delta CT}$ calculation method (Wang et al., 2022). Three biological replicates were performed for each reaction using primers detailed in Supplementary Tables 1–4.

Expression patterns

Based on the cDNA sequence of *OsG6PGH1*, we designed and detected primers using the NCBI website. Total RNA was extracted from root, stem, leaf, sword leaf, glume, stem node, and endosperm tissues, and reverse transcribed into cDNA. Expression levels of *OsG6PGH1* in tissues were detected by qRT-PCR (Li et al., 2023). *β-Actin* was used as an internal reference gene in each replicated experiment.

Subcellular localization

Protoplasts were first extracted from rice seedlings cultured in darkness for 14 d. The subcellular localization vector *OsG6PGH1-GFP* and plasma membrane (*PAT2-RFP*) and nuclear (*Ghd7-RFP*) markers were co-transformed into the protoplasts of rice seedlings. Fluorescence signals in protoplasts were detected using confocal fluorescence microscopy (Peng et al., 2023); all fluorescence experiments were independently observed at least three times.

Plant growth status and trait determination

Field experiments were performed on rice plants at Xinyang Normal University, Henan Province. Planting density was 16.5 × 26.0 cm, with the field managed in accordance with local agricultural standards. Before measurement, harvested grains were air dried, and

then stored at room temperature for at least 3 months. Fully grouted grains were used to detect rice quality and yield traits (Li et al., 2014). A near-infrared grain analyzer (INFRAEC Nova) was used to detect protein, amylose, and free fatty acid contents, and taste value (Peng et al., 2023). The gelatinization temperature of grain was measured using the alkaline digestion method (Peng et al., 2024), and polished rice gel consistency was determined following People's Republic of China (GB/T17891-1999, high-quality rice) national standards.

Transmission electron microscopy

The endosperm of 10 DAF rice was fixed in 2.5% glutaraldehyde with 0.2 mol L^{-1} phosphate buffer (pH 7.2) for at least 12 h. Fixed endosperm tissue was cut into 1 mm sections using an ultra-thin slicer (Power Tome XL, RMC). Sections were examined by TEM (H-7650, Hitachi). ImageJ software (NIH) was used to calculate the quantity and area of PBs in each TEM sample image (Peng et al., 2014).

Observation by optical and scanning electron microscopy

The middle part of the grain was gently tapped with the back of a knife to allow natural breakage; a sharp blade was then used to cut the broken part to form a thin sheet of ~2–3 mm thickness. Slices were then divided into two: one part to observe the chalky area beneath a regular optical microscope; the other to be fixed on a copper sample stage with conductive adhesive, sprayed with platinum using a HUS-5GB vacuum coating machine, and examined by scanning electron microscope (Hitachi S-4800) (Xu et al., 2022).

Proteomic analysis

Both *OsG6PGH1*(OX) positive and negative plants grew in the same environment. Endosperms were simultaneously collected, and then ground for proteomic sequencing (replicated three times per sample). Total protein was first extracted, and its concentration, purity, and integrity were examined to ensure sample usability. Obtained data were filtered to obtain clean readings aligned with the reference proteome sequence. Statistical tests were performed to obtain DEPs based on the annotation file of the proteome and reading position (i.e., expression level) compared with the proteome. Finally, Clusters of Orthologous Genes functional annotation, Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and Pfam enrichment analyses were performed on DEPs to identify biological processes and metabolic pathways of protein enrichment (Xiong et al., 2019).

Salt stress test

After peeling and disinfecting *OsG6PGH1* transgenic seeds, they were inoculated into 1/2 MS solid culture medium. After 7 d, rice seedlings with consistent growth were selected and placed into a hydroponic medium for acclimation for 7 d. Seedlings cultured for 14 d were inoculated in 50 mmol L^{-1} and 100 mmol L^{-1} concentrations of NaCl medium, and placed into a 28°C constant temperature light incubator. Nutrient solutions were changed every 2 d to prevent bacterial contamination (Zhao et al., 2022b). Seedling growth was monitored daily.

Results

OsG6PGH1 can bind to *OsAAP6*

A total of 92 regulatory factors that may interact with *OsAAP6* were screened through yeast one- hybrid experiments (Peng et al., 2023). One regulatory factor, the plasmid pGADT7-Rec2-*G6PGH1*, along with the bait plasmid pHIS2-*OsAAP6*, were transferred into yeast receptive cell Y187. On solid culture media, with the addition of 3AT inhibitors, the *OsG6PGH1* yeast colony and positive control yeast colony (+) both grew well, but there were significant differences between them and the negative control (−) (Figure 1A). This indicates that *OsG6PGH1* can interact with OsAAP6 in yeast.

To verify the interaction between *OsG6PGH1* and *OsAAP6*, prokaryotic expression vectors pET-28a-*OsG6PGH1* and pET-28a were transfected into *Escherichia coli* competent cells BL21 (DE3). In local surface plasmon resonance experiments, the curves for three concentrations all significantly trended upward at 280 s. This change indicated that an interaction occurred between *OsG6PGH1* and *OsAAP6*, forming a complex (*OsG6PGH1–OsAAP6*). At 1650 s, there was a significant downward trend in curves for each concentration, indicating that the *OsG6PGH1–OsAAP6* complex had undergone dissociation (Figure 1B). These dynamic changes indicate that *OsG6PGH1* can bind to *OsAAP6 in vitro*.

OsG6PGH1 is a constitutive expression gene and the encoded protein may be localized in the cytoplasm

To investigate expression levels of *OsG6PGH1* in different rice tissues, RNA was extracted from root, stem, stem node, leaf, sword leaf, endosperm, and glume tissues. Expression patterns were analyzed using qRT-PCR. *OsG6PGH1* was expressed in each tissue type (Figure 2A), and this result was in accordance with the distinct tissue expressions of *OsG6PGH1* (LOC_Os06g02144) on CREP (Supplementary Figure 1), indicating it is a constitutive expression gene. Expression levels of *OsG6PGH1* were higher in roots, and lower in the endosperm and glume (Figure 2A).

To investigate localization of *OsG6PGH1* in cells, *OsG6PGH1* was fused with the green fluorescent protein gene (*GFP*) (*OsG6PGH1-GFP*) and simultaneously co-expressed with the cell-membrane-localization marker gene *PAT2*-RFP and nuclear marker gene *Ghd7*-RFP in rice protoplasts, and observed using laser confocal electron microscopy. *OsG6PGH1*-GFP was not co-localized with membrane marker protein PAT2-RFP or nuclear marker protein Ghd7-RFP (Figure 2B), indicating that *OsG6PGH1*

10.3389/fpls.2024.1436998

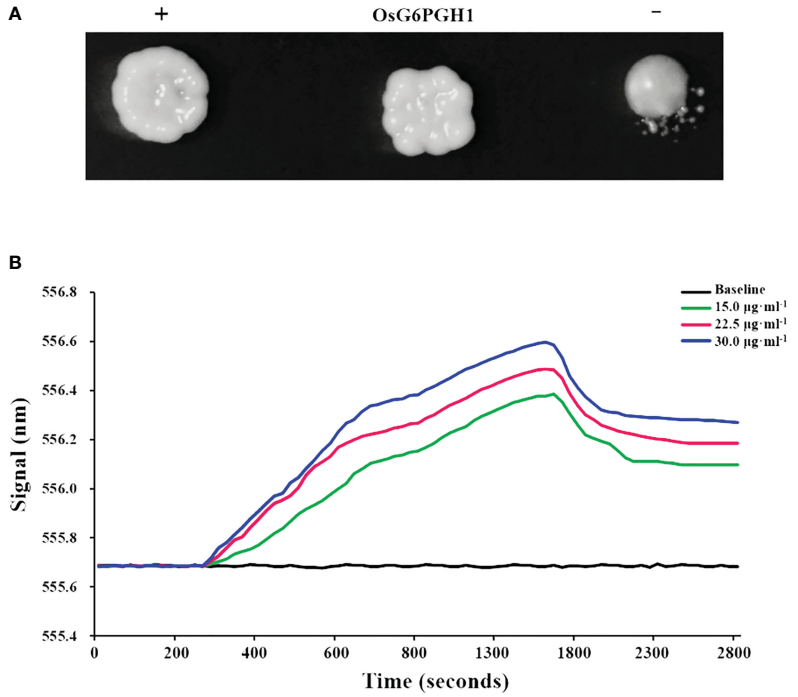


FIGURE 1
Verification of interactions between *OsG6PGH1* and OsAAP6. **(A)** Yeast *in vivo* point-to-point experiment: +, positive control; –, negative control. **(B)** *In vitro* validation using localized surface plasmon resonance experiments.

was not localized in either the cell membrane or nucleus, but that it may be localized in the cytoplasm.

Creation of *OsG6PGH1* transgenic rice lines

To investigate the biological function of *OsG6PGH1* in rice, we created CRISPR/Cas9 knockout and overexpressing transgenic lines of *OsG6PGH1*. Based on the wild-type Zhonghua 11 (ZH11), a target site for *OsG6PGH1* was designed. CRISPR/Cas9 gene editing tool was used to target and induce the fusion of sgRNA and Cas9 protein, and specific splicing was performed on the two PAM sites of *OsG6PGH1*. A gene edited expression vector was constructed, and 20 CRISPR/Cas9 gene edited plants were obtained through *Agrobacterium*-mediated genetic transformation (including 16 *OsG6PGH1* positive mutants) (Supplementary Figure 2A). DNA was extracted from the leaves of the *OsG6PGH1* mutants, and PCR amplification and DNA sequencing were performed; two types of mutations were found in the cDNA target sequence of *OsG6PGH1*: *Osg6pgh1-1* (+1 bp) and *Osg6pgh1-2* (+2 bp) (Supplementary Figure 2B). The target sequence of *Osg6pgh1-1* has a 1 bp insertion and 3 bp substitution in its second exon, resulting in a frameshift mutation, and *Osg6pgh1-2* has a 2 bp insertion and 4 bp substitution in its first and second exons, which also causes a frameshift mutation.

To obtain transgenic plants with *OsG6PGH1* overexpression, we connected the cDNA of *OsG6PGH1* to the intermediate vector PMD-18T to obtain PMD-18T-*OsG6PGH1*. After sequencing verification, the target fragment was connected to the pCMBIA1301s overexpression vector to obtain the *OsG6PGH1* overexpression plasmid pCMBIA1301s-*OsG6PGH1*. Through *Agrobacterium*- mediated genetic transformation, 20 transgenic plants with *OsG6PGH1* overexpression (*OsG6PGH1*(OX)) were obtained (including 15 positive plants) (Supplementary Figure 2C).

Overexpression of *OsG6PGH1* affects the content of major nutrients in rice

We performed qRT-PCR analysis on expression levels of *OsAAP6* in the endosperm of *OsG6PGH1* mutants (M_2) and overexpressing transgenic F_2 lines. There was no significant change in expression levels of *OsAAP6* in the endosperm of *OsG6PGH1* homozygous mutants, while expression levels significantly increased in the *OsG6PGH1* overexpressing positive endosperm (Figure 3A).

Because of the influence of *OsAAP6* on the distribution of amino acids in grains and its positive regulation of grain protein content, we detected and analyzed amino acid contents and GPC in *OsG6PGH1* mutants (M_2) and overexpressing transgenic F_2 grains. *OsG6PGH1* affected the content of some amino acids (e.g., alanine, glutamic acid, leucine, lysine, methionine) (Supplementary Figure 3A). Total and essential amino acid contents in grains of *OsG6PGH1*(OX) positive plants significantly increased, while there was no difference in grains of *OsG6PGH1* homozygous mutants (Figures 3B, C). Additionally, the GPC in *OsG6PGH1*(OX) positive plants increased significantly, while there was no significant

10.3389/fpls.2024.1436998

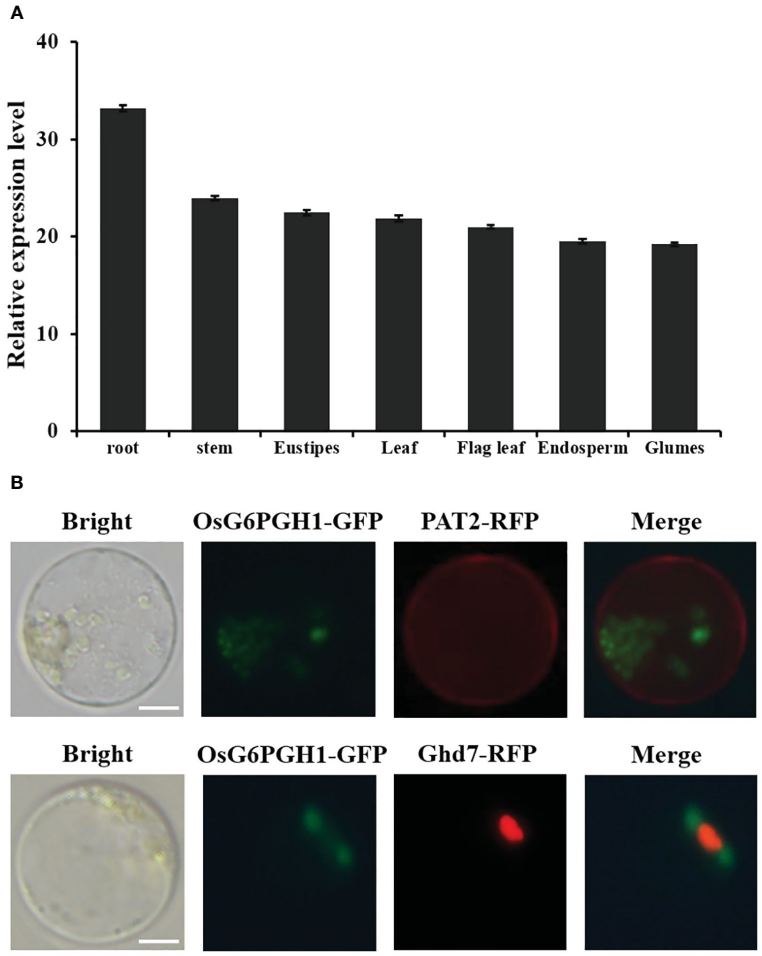


FIGURE 2
Analysis of *OsG6PGH1* expression patterns and subcellular localization in rice. **(A)** Detection and analysis of expression levels in different tissues. **(B)** *OsG6PGH1*-GFP is not co-localized with either the membrane marker protein PAT2-RFP or nuclear marker protein Ghd7-RFP. Scale bars, 5 μm; error bars, standard error of the mean (SEM).

difference in *OsG6PGH1* mutants (Figure 3D). To further investigate if *OsG6PGH1*(OX) affected the protein bodies in rice endosperm, we observed the endosperm of *OsG6PGH1*(OX) 10 d after fertilization (DAF) using transmission electron microscopy (TEM). Protein bodies I (PB I) and protein bodies II (PB II) were clearly visible under TEM: PB I had relatively regular circular or elliptical shapes, and PB II had irregular shapes and stained uniformly (Figure 3E). Average cross-sectional areas of PB I and II in *OsG6PGH1*(OX) positive endosperms did not differ significantly from those in transgenic negative endosperms, while numbers of PB II significantly increased (Figure 3F; Supplementary Figures 3B–D). This indicates that overexpression of the *OsG6PGH1* significantly increased the number of PB II in rice endosperm, leading to increased GPC.

To further investigate if *OsG6PGH1* affected the accumulation of nutrients such as starch and free fatty acids in rice, starch and free fatty acid content was measured in *OsG6PGH1* mutants (M_2) and overexpressing transgenic F_2 plants. Total starch contents in *OsG6PGH1*(OX) positive grains were significantly reduced, while total starch contents in *OsG6PGH1* homozygous mutants did not change significantly (Figure 3G). There was no significant change in free fatty acid content in homozygous mutants of *OsG6PGH1* or *OsG6PGH1*(OX) positive grains (Supplementary Figure 3E). Therefore, overexpression of *OsG6PGH1* helps to increase GPC, but not the accumulation of starch in grains.

OsG6PGH1 participates in regulating expression of genes related to protein and starch metabolism in grains

Because of the significant impact of *OsG6PGH1*(OX) on accumulation of protein and starch in grain, we examined expression levels of protein-and starch-metabolism-related genes in *OsG6PGH1* mutants and overexpressed transgenic F_2 endosperm. Gene expression levels correlated positively with grain protein synthesis and accumulation metabolism, with *10KD Prolamine*, *GluA1*, *11S Globulin*, *Glutelin 4*, *GluB1*, *AlaAT*, *GluA3*, *GluA2*, *GluB4* significantly upregulated in the positive endosperm of *OsG6PGH1*(OX), but not in *OsG6PGH1* homozygous mutants

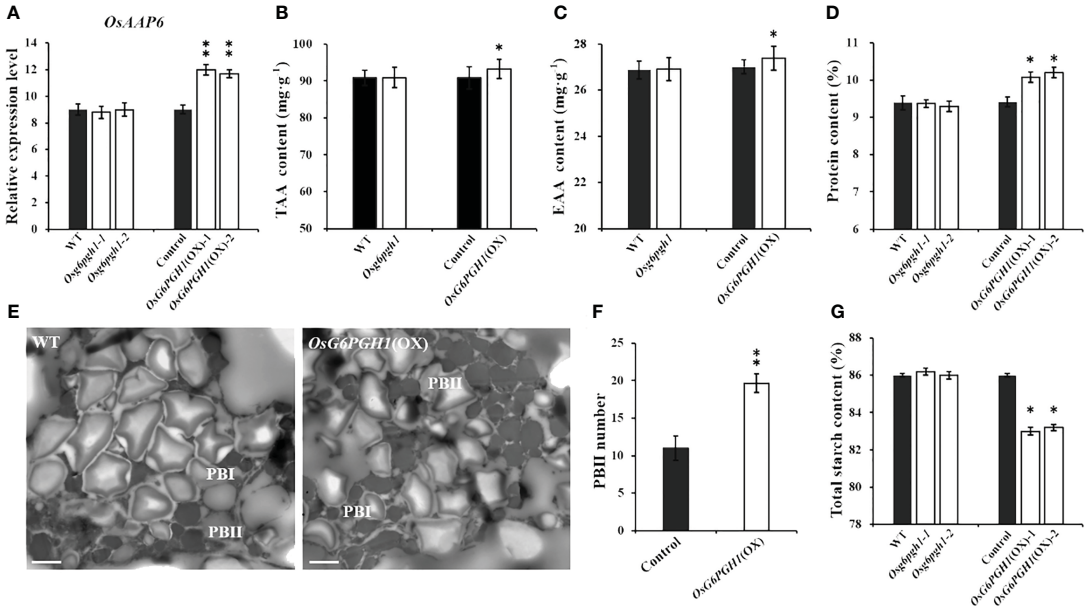


FIGURE 3
Detection and analysis of major nutrients in transgenic rice grains with *OsG6PGH1*. **(A)** Detection and analysis of expression levels of *OsAAP6* in *OsG6PGH1* transgenic endosperm. *OsG6PGH1* transgenic plant: **(B)** total amino acids, **(C)** essential amino acids, and **(D)** grain protein content. **(E)** transmission electron microscopy image of endosperm of *OsG6PGH1*(OX) 10 days after fertilization; and analysis of **(F)** protein body II, and **(G)** total starch content. Significant differences are based on two tailed *t*-tests, where $^{**}P \leq 0.01$, $^{*}P \leq 0.05$. WT, Wild type; Control, *OsG6PGH1*(OX) negative control. Scale bars, 2 μm; error bars, standard error of the mean (SEM), n ≥ 50.

(Figures 4A–J). Expression levels of starch-synthesis-related genes (*Susy2*, *Susy3*) and starch-degradation metabolism-related genes (*AMY3A*, *AMY3B*, *ISA1*) were significantly downregulated in *OsG6PGH1* homozygous mutants, and significantly upregulated in *OsG6PGH1*(OX) positive plants. Expression levels of genes related to starch synthesis (*SS1*, *SSIVa*, *SBE*) were significantly upregulated in *OsG6PGH1* homozygous mutants, while expression levels were the opposite in *OsG6PGH1*(OX) positive plants (Figures 4K–S). The granule-bound starch synthase gene (*GBSSI*) related to amylose synthesis was significantly downregulated in *OsG6PGH1* homozygous mutants and significantly upregulated in *OsG6PGH1*(OX) positive plants (Figure 4T). These results are consistent with a phenotype with significant increases in GPC and a significant decrease in starch content in rice grains with *OsG6PGH1*(OX).

To investigate the effect of changes in *OsG6PGH1* on the activity of upstream and downstream enzymes in the PPP, we measured the activity of 6-phosphogluconate dehydrogenase (G6PGH) and its upstream rate-limiting enzyme glucose-6-phosphate dehydrogenase (G6PDH) in the endosperm of rice 10 DAF. Activity of G6PGH in the endosperms of *OsG6PGH1* transgenic F_2 plants changed significantly: the activity of G6PGH in endosperm of *OsG6PGH1* homozygous mutants was significantly reduced, and significantly enhanced in *OsG6PGH1*(OX) positive endosperms (Figure 4U); the activity of the upstream rate-limiting enzyme G6PDH did not change significantly in *OsG6PGH1* transgenic plants (Figure 4V).

OsG6PGH1 affects the cooking and processing quality of rice

To investigate if *OsG6PGH1* affects the cooking and eating quality of rice, we tested amylose content, taste value, gel consistency, and gelatinization temperature of *OsG6PGH1* mutants and overexpressing transgenic F_2 plants. Compared with wild type rice, the amylose content of *OsG6PGH1* homozygous mutants was significantly lower, and taste value and gelatinization temperature were significantly higher. Amylose contents of *OsG6PGH1*(OX) positive grains were significantly higher, the taste value was significantly lower, and the gelatinization temperature did not change significantly (Figures 5A–C). Gel consistency of *OsG6PGH1* mutants and overexpressing transgenic rice showed no significant changes (Supplementary Figure 4A).

Because amylose contents of *OsG6PGH1* mutants and overexpressing transgenic rice changed significantly, we analyzed their milling quality traits. The head rice rate of *OsG6PGH1*(OX) positive F_2 rice significantly decreased (Figure 5D), while the brown and milled rice rates showed no significant change (Supplementary Figures 4B, C). Compared with wild type rice, there was no significant change in milling quality traits of *OsG6PGH1* homozygous mutant rice (Figure 5D; Supplementary Figures 4B, C). These results indicate that upregulation of *OsG6PGH1* expression does not improve rice cooking or eating quality, decreases the head rice rate, and affects the grain processing quality.

10.3389/fpls.2024.1436998

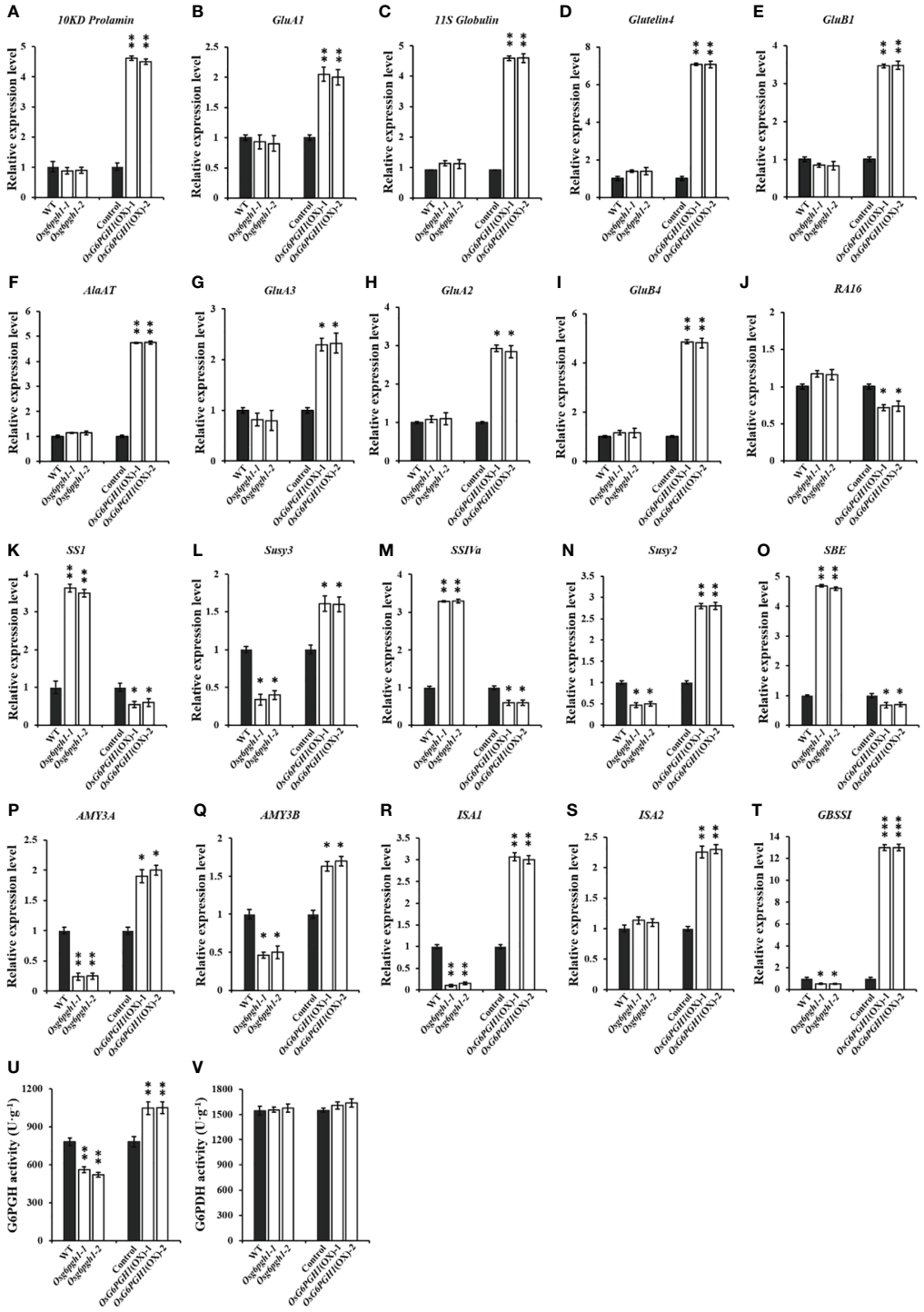


FIGURE 4
Expression and enzyme activities of related genes in *OsG6PGH1* transgenic rice endosperm. **(A–J)** qRT-PCR analysis of protein-metabolism-related genes; **(K–T)** qRT-PCR analysis of starch-metabolism-related genes; and detection and analysis of enzyme activity of endosperm **(U)** 6PGDH and **(V)** G6PDH in *OsG6PGH1* transgenic rice lines 10 DAF. Significant differences are based on two tailed *t*-tests, where ***$P \leq 0.001$, **$P \leq 0.01$, *$P \leq 0.05$. WT, Wild type; Control, *OsG6PGH1*(OX) negative control. Error bars, standard error of the mean (SEM).

OsG6PGH1 negatively regulates rice chalkiness rate

Chalkiness rates of *OsG6PGH1* homozygous mutants increased significantly, while the chalkiness rate of *OsG6PGH1*(OX) rice was significantly reduced (Figures 5E, F). There were no significant changes in chalkiness area or degree in *OsG6PGH1* mutants or overexpressing transgenic rice. The morphology and arrangement of starch granules in the endosperm of *OsG6PGH1* mutants and overexpressing transgenic positive rice showed no significant

10.3389/fpls.2024.1436998

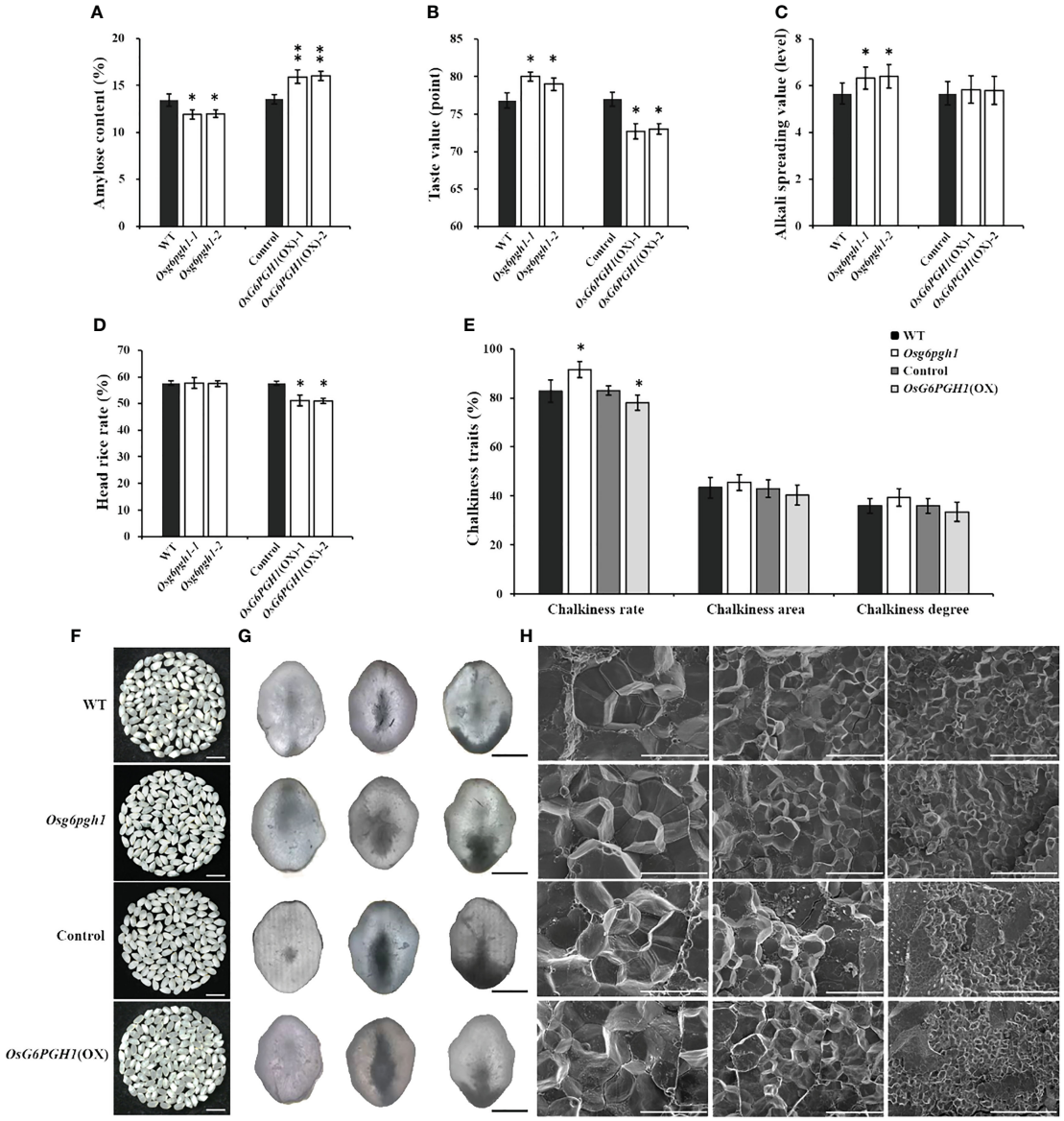


FIGURE 5
Analysis of cooking and eating, processing, and appearance quality traits of *OsG6PGH1* transgenic rice. Detection and analysis of: **(A)** amylose content, **(B)** taste value, **(C)** gelatinization temperature; **(D)** head rice rate, **(E)** chalkiness traits, and **(F)** endosperm chalkiness characteristics. Scale bars, 1 cm. Images of: **(G)** chalkiness under optical microscope; columns from left (transparent), central (heart white), and right (belly white); scale bars, 1 mm; **(H)** Scanning electron microscopy of endosperm; scale bars, left to right: 1 mm, 10 μm, 20 μm, and 100 μm. Significant differences are based on two-tailed *t*-tests, where $^{**}P \leq 0.01$, $^{*}P \leq 0.05$. WT, Wild type; Control, *OsG6PGH1* overexpression transgenic negative control. Error bars, standard error of the mean (SEM).

change; both presented a densely arranged polygonal structure (Figures 5G, H). Therefore, *OsG6PGH1* negatively regulated the chalkiness rate of grain, but did not affect chalkiness area or degree.

OsG6PGH1 does not affect rice yield per plant

To investigate if *OsG6PGH1* affected rice yield, we tested yield-related traits of *OsG6PGH1* mutants (M_2) and their overexpressing transgenic F_2 plants. *OsG6PGH1* homozygous mutant plant height was significantly reduced (Supplementary Figures 5A, B); there were no significant changes in the number of tillers per plant, number of spikelets per panicle, seed setting rate, thousand grain weight, or yield per plant (Supplementary Figures 5C–G). Numbers of spikelets per panicle of *OsG6PGH1*(OX) positive plants significantly decreased (Supplementary Figure 5D), and the setting rate significantly increased (Supplementary Figure 5E); other yield-related traits showed no significant changes, leading to no overall significant change in rice yield per plant (Supplementary Figure 5G). Because *OsG6PGH1*(OX) significantly affected grain protein and starch contents, we examined the grain type of *OsG6PGH1* mutants and overexpressed transgenic seeds. There were no significant changes in grain length, width, or thickness of

OsG6PGH1 homozygous mutants and overexpressing transgenic positive grains (Supplementary Figure 6). Therefore, although *OsG6PGH1* affected plant height, number of spikelets per panicle, and seed setting rate, it did not affect rice yield per plant.

Proteomic analysis of *OsG6PGH1*(OX) transgenic endosperms

The GPC in *OsG6PGH1*(OX) positive grains significantly increased. To evaluate the accumulation of different storage proteins in grains, we performed proteomic analysis on *OsG6PGH1*(OX) endosperm. Upregulated proteins were screened using two methods: differential multiple (|log2 (Fold change)| ≥ 1.50) and significance level (*P*-value ≤ 0.05). Downregulated proteins were screened through two methods: differential multiple (|log2 (Fold change)| ≤ 1/1.50) and significance level (*P*-value ≤ 0.05). A total of 156 differentially expressed proteins (DEPs) were screened (Supplementary Figure 7A).

To understand the structure and function of these DEPs, functional annotation was performed using the COG database. Functions of DEPs were mainly related to amino acid transport and metabolism, carbohydrate transport and metabolism, and protein post-translational modification (Supplementary Figure 7B). GO functional enrichment analysis of these DEPs revealed significant differential enrichment in *OsG6PGH1*(OX) positive endosperms to occur mainly in cellular components and organelles, participation in metabolic processes, cellular processes, and other biological processes, with molecular functions such as binding and catalytic activity (Supplementary Figure 7C). KEGG pathway analysis on these DEPs revealed carbon metabolism, glycolysis/gluconeogenesis, starch and sucrose metabolism, and carbon fixation by photosynthetic organisms, to be pathways where DEPs were more concentrated in *OsG6PGH1*(OX) positive endosperms (Supplementary Figure 7D). Domain annotation and enrichment were performed on these DEPs through the Pfam database; protease inhibitors, seed storage, and LTP family were mostly enriched in *OsG6PGH1*(OX) positive endosperms (Supplementary Figure 7E). Results of proteomic analysis are consistent with upregulation of protein- and amylose-synthesis-related genes, and downregulation of amylopectin-metabolism-related genes in the positive endosperm of *OsG6PGH1*(OX) (Figure 4; Supplementary Table 5). This suggests that overexpression of *OsG6PGH1* may promote protein synthesis and accumulation in the endosperm, while inhibiting corresponding starch synthesis and metabolic pathways.

OsG6PGH1 enhances rice tolerance to salt stress

Because *OsG6PGH1*, as a key gene in the PPP, may be involved in the salt stress response of rice, we performed salt stress treatment experiments on *OsG6PGH1* mutants (M_2) and *OsG6PGH1*(OX) transgenic F_2 seedlings. Rice seedlings grown for 14 d in culture media were exposed to concentrations of 50 mmol·L^{-1} and 100 mmol·L^{-1} NaCl for 7 d. The survival rate of *OsG6PGH1* homozygous mutant seedlings was significantly reduced compared with wild-type rice at 100 mmol·L^{-1} NaCl, but that of *OsG6PGH1*(OX) positive seedlings increased significantly in both salt stress treatments. This indicates that *OsG6PGH1* improves salt stress resistance of rice seedlings (Figures 6A–C).

To further explore the molecular mechanism of *OsG6PGH1* regulating salt stress resistance in rice seedlings, we examined expression levels of six salt stress-related genes (*OsPEX11*, *OsSalt*, *OsWsi18*, *OsJRL*, *OsHKT1*, *OsATK1*) in *OsG6PGH1* mutants and *OsG6PGH1*(OX) seedlings. qRT-PCR results revealed expression levels of these genes to be significantly increased in *OsG6PGH1*(OX) positive seedlings, while *OsSalt* was significantly reduced in *OsG6PGH1* homozygous mutant plants (Figures 6D–I). These results are consistent with observations that salt stress resistance of *OsG6PGH1* homozygous mutant seedlings is significantly reduced, while the salt stress resistance of transgenic positive rice seedlings is significantly enhanced.

Discussion

AAPs are involved in various physiological metabolic processes in plants, and affect rice growth and development (Fang et al., 2021). In the rice genome, 19 *OsAAPs* (*OsAAP1–19*) have been identified, with most of them isolated, cloned, and identified (Peng et al., 2014; Taylor et al., 2015; Lu et al., 2018; Wang et al., 2019; Ji et al., 2020; Wang et al., 2020; Fang et al., 2021; Erinaldo et al., 2022; Yang et al., 2022; Yang et al., 2023a, b; Peng et al., 2024). Among them, *OsAAP6* is a major quantitative trait locus gene that positively regulates GPC and nutritional quality of rice (Peng et al., 2014, 2019). To further elucidate the regulatory mechanism of *OsAAP6*, a total of 92 potential regulatory factors that may interact with it were screened through yeast one-hybrid experiments (Peng et al., 2023). The ability of *OsG6PGH1* to bind to *OsAAP6* was then validated through yeast *in vivo* point-to-point rotation experiments and local surface plasmon resonance experiments (Figure 1); it was involved in regulating expression of *OsAAP6* in *OsG6PGH1*(OX). These results provide key clues for the in-depth analysis of *OsAAP6* regulation of GPC, but also open up new possibilities for using *OsAAP6* as a gene resource in molecular design breeding.

G6PGH is a key gene in the PPP, with multiple biological functions. After the double mutation of *6pgd1* and *6pgd2* in corn, the growth of homozygous individuals with the mutation was delayed (Bailey-Serres and Nguyen, 1992); overexpression of *G6PGH* in apples can accelerate growth of callus tissue (Zhu et al., 2019). We obtained *OsG6PGH1* mutants using CRISPR/Cas9 gene-editing tool and created *OsG6PGH1*(OX) transgenic plants. *OsG6PGH1* homozygous mutant plant height was significantly reduced, as were the number of spikelets per panicle in *OsG6PGH1*(OX) transgenic positive plants, while the seed setting rate increased significantly. Notably, there was no significant change in rice yield per plant in *OsG6PGH1* homozygous mutants or overexpressing transgenic positive rice. Therefore. *OsG6PGH1* may be involved in regulating rice growth and development, but not yield per plant.

10.3389/fpls.2024.1436998

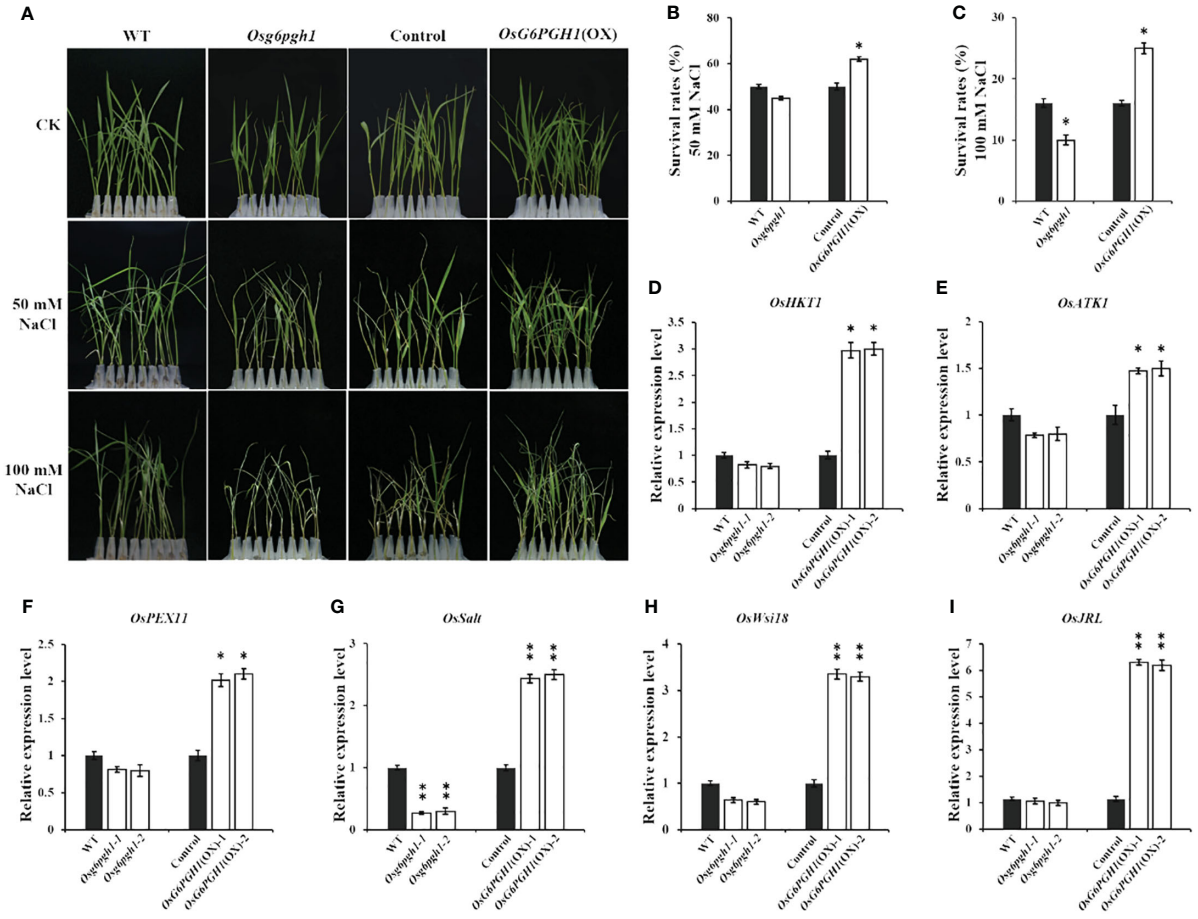


FIGURE 6
Salt stress tests and expression analysis of related genes in *OsG6PGH1* transgenic rice seedlings. **(A)** Salt stress test. Survival rate of rice seedlings at **(B)** 50 mmol·L^{-1} NaCl, and **(C)** 100 mmol·L^{-1} NaCl. **(D–I)** qRT-PCR detection and analysis of salt stress-related genes. Significant differences are based on two-tailed *t*-tests, where $^{**}P \leq 0.01$, $^{*}P \leq 0.05$. WT, Wild type; Control, *OsG6PGH1* overexpression transgenic negative control. Error bars, standard error of the mean (SEM).

Rice grains are rich in amino acids, with a relatively balanced composition (20 types of amino acids, including 8 essential amino acids) (Yang et al., 2022). Grain protein is also high-quality, and easily digestible and absorbed (Peng et al., 2023). GPC is a key factor in the nutritional quality of rice (Xia et al., 2022). Therefore, exploring genes related to GPC and analyzing their functions is important for molecular genetic breeding of high-quality, genetically improved rice. We report the positive regulatory factor *OsAAP6* for GPC to be significantly upregulated in *OsG6PGH1* (OX) transgenic positive plants, resulting in a significant increase in GPC and a significant decrease in starch content. This indicates that *OsG6PGH1* is involved in regulating accumulation of major storage substances in rice grains. Further observation of protein bodies in the endosperm of *OsG6PGH1*(OX) revealed no significant changes in the area and quantity of protein body I, and area of protein body II, but the quantity of protein body II increased significantly. This indicates that overexpression of *OsG6PGH1* significantly increases the number of protein bodies II, leading to a significant increase in GPC (Figure 3). Detection and analysis of expression levels of protein- and starch-metabolism-related genes in *OsG6PGH1* mutants and overexpressed transgenic endosperm (Figure 4) revealed expression levels of protein- and starch-synthesis metabolism-related genes in rice grains to be consistent with their corresponding main storage substances (protein and starch) contents. Proteomic analysis of *OsG6PGH1*(OX) endosperm revealed DEPs to mainly contain structural domains related to seed storage, have biological functions in metabolic processes, and participate in biological pathways such as carbon metabolism, and starch and sucrose metabolism (Supplementary Figure 6). Therefore, genes related to protein synthesis and metabolism are upregulated in *OsG6PGH1*(OX) transgenic positive endosperm, which can significantly increase numbers of protein bodies II, promote accumulation of related storage proteins in the endosperm, and increase GPC and nutritional quality.

Starch, an important component in rice grains, accounts for approximately 80%–90% of rice dry weight. This starch is composed of amylose and amylopectin (Zhao et al., 2022a), the composition and proportion of which directly affect grain quality. In rice endosperm, amylose is synthesized by the *Waxy* (*Wx*) gene encoding GBSSI (Zhang et al., 2019), and amylose content plays a decisive role in the eating quality of rice (Pang et al., 2016; Zhang et al., 2022). Inhibition of the branched starch synthesis related gene

SSIIa leads to an increase in chalkiness characters, significantly affecting the appearance of rice (Zhang et al., 2011; Zhao et al., 2022b). The PPP is important for sugar metabolism in plants, and *G6PGH* is involved in biosynthesis and metabolism of plant starch. The cytoplasmic form of *G6PGH* is necessary for normal starch biosynthesis in maize endosperm. The *PGD3* mutation of the maize exhibits severe granule filling and defective embryonic phenotype (Ribeiro et al., 2020). We report the amylose content in rice grains of *OsG6PGH1* mutants to be significantly reduced, while the chalkiness rate, taste value, and gelatinization temperature all increase significantly. However, the amylose content of *OsG6PGH1*(OX) positive grains significantly increased, while the chalkiness rate, taste value, and head rice rate all decreased significantly (Figure 5). Therefore, *OsG6PGH1* can affect rice quality traits such as appearance, cooking, eating, and processing qualities by regulating contents of main storage substances (protein and starch).

G6PGH plays an important role in the response of plants to abiotic stress (Chen et al., 2020). Transgenic *G6PGH* in maize can significantly improve heat tolerance, thereby increasing maize yield under high temperature stress (Ribeiro et al., 2020). Overexpression of *6PGDH* in *Arabidopsis* significantly enhances cold tolerance (Tian et al., 2021). In rice seedlings treated with drought, cold, high salt, and abscisic acid, the transcription levels of *OsG6PGH1* and *OsG6PGH2* were upregulated (Hou et al., 2007). Under salt stress, expression of *OsG6PGH* in rice stems is upregulated (Huang et al., 2003), indicating that OsG6PGH actively participates in the response of rice to abiotic stress. We report *OsG6PGH1* to be neither localized on the cell membrane nor in the rice nucleus (Figure 2), but possibly in the cytoplasm, consistent with previous results. Analysis of expression patterns reveals *OsG6PGH1* to be a constitutive expression gene that is expressed in all rice tissues, consistent with expression of *OsG6PGH* in root, stem, and leaf tissues of cucumber (Wei et al., 2010). We report salt resistance in rice seedlings with *OsG6PGH1* homozygous mutants to be significantly reduced, but for salt resistance in *OsG6PGH1*(OX) positive seedlings to be significantly enhanced (Figure 6), consistent with expression levels in corresponding salt stress-related genes. Therefore, under salt stress conditions, *OsG6PGH1* can positively regulate expression of genes related to the salt stress response, thereby improving rice salt stress resistance.

In summary, *OsG6PGH1* was screened, identified, and validated to interact with *OsAAP6* through yeast one-hybrid technology, and *in vivo* point-to-point rotation and local surface plasmon resonance tests. *OsG6PGH1* is a constitutive expressed gene, and its encoded *OsG6PGH1* may be located in the cytoplasm. Expression levels of *OsAAP6* were significantly upregulated in *OsG6PGH1*(OX) transgenic positive plants, resulting in a significant increase in grain protein content (GPC), and improved nutritional quality of rice. *OsG6PGH1* positively regulates amylose content, and negatively regulates rice chalkiness rates and taste values, significantly affecting various quality traits such as appearance, cooking, and eating qualities of rice. Additionally, *OsG6PGH1* is involved in regulating the expression of salt-stress-related genes, enhancing the salt-stress tolerance of rice. Therefore, *OsG6PGH1* affects multiple quality traits, participates in the salt stress response of rice, and provides important genetic resources for molecular design and breeding of high-quality and multi-resistant rice varieties.

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

BP: Writing – original draft, Writing – review & editing, Data curation, Funding acquisition, Investigation, Methodology, Project administration. YL: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft. JQ: Data curation, Formal analysis, Methodology, Writing – original draft. JP: Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft. XS: Conceptualization, Data curation, Methodology, Writing – original draft. XT: Data curation, Methodology, Resources, Writing – original draft. ZZ: Formal analysis, Methodology, Resources, Writing – original draft. YH: Funding acquisition, Investigation, Methodology, Writing – original draft. RP: Formal analysis, Investigation, Supervision, Writing – original draft. WZ: Conceptualization, Data curation, Methodology, Writing – original draft. JZ: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. YS: Data curation, Formal analysis, Investigation, Writing – review & editing. QW: Data curation, Investigation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by National Natural Science Foundation of China (U2004141; 12305399), Key Project of Science and Technology in Henan Province (222102110141; 242102110266; 242102110270), Key R&D and Promotion Project of Xinyang-Science and Technology Research Project (20230025), Nanhu Scholars Program for Young Scholars of XYNU, and Student Research Fund Project of XYNU (202410477040).

Acknowledgments

We thank EditSprings (https://www.editsprings.cn) for providing the expert linguistic services.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1436998/full#supplementary-material

References

Bailey-Serres, J., and Nguyen, M. T. (1992). Purification and characterization of cytosolic 6-phosphogluconate dehydrogenase isozymes from maize. *Plant Physiol.* 100, 1580–1583. doi: 10.1104/pp.100.3.1580

Balindong, L. J., Ward, M. R., Rose, J. T., Liu, L., Raymond, C. A., Snell, P. J., et al. (2018). Rice grain protein composition influences head rice yield. *Cereal Chem.* 95, 253–263. doi: 10.1002/cche.10031

Bao, J. S., Deng, B. W., and Zhang, B. L. (2023). Molecular and genetic bases of rice cooking and eating quality: an updated review. *Cereal Chem.* 100, 1220–1233. doi: 10.1002/cche.10715

Chen, L., Kuai, P., Ye, M. F., Zhou, S. X., Lu, J., and Lou, Y. G. (2020). Overexpression of a cytosolic 6-phosphogluconate dehydrogenase gene enhances the resistance of rice to *Nilaparvata lugens*. *Plants* 9, 1529. doi: 10.3390/plants9111529

Chen, P. L., Shen, Z. K., Ming, L. C., Li, Y. B., Dan, W. H., Lou, G. M., et al. (2018). Genetic basis of variation in rice seed storage protein (albumin, globulin, prolamin, and glutelin) content revealed by genome-wide association analysis. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00612

Custodio, M. C., Cuevas, R. P., Ynion, J., Laborte, A. G., Velasco, M. L., and Demont, M. (2019). Rice quality: How is it defined by consumers, industry, food scientists and geneticists? *Trends Food Sci. Technol.* 92, 122–137. doi: 10.1016/j.tifs.2019.07.039

Erinaldo, G. P., Cassia, P. C. B., Carlos, A. B., Leandro, A. S., Joviana, L., Claudete, S. C., et al. (2022). The amino acid transporter OsAAP1 regulates the fertility of spikelets and the efficient use of n in rice. *Plant Soil.* 480, 507–521. doi: 10.1007/S11104-022-05598-9

Fang, Z. M., Wu, B. W., and Ji, Y. Y. (2021). The amino acid transporter OSAAP4 contributes to rice tillering and grain yield by regulating neutral amino acid allocation through two splicing variants. *Rice* 14, 2. doi: 10.1186/s12284-020-00446-9

Gopika, J., Aarti, B., Prince, C., Melinda, F., and Fogarasi, S. (2022). A narrative review on rice proteins: current scenario and food industrial application. *Polymers* 14, 3003. doi: 10.3390/polym14153003

Hanks, J. F., Schubert, K., and Tolbert, N. E. (1983). Isolation and characterization of infected and uninfected cells from soybean nodules role of uninfected cells in ureide synthesis. *Plant Physiol.* 71, 869–873. doi: 10.1104/pp.71.4.869

He, Y., Wang, S. Z., and Ding, Y. (2013). Identification of novel glutelin subunits and a comparison of glutelin composition between japonica and indica rice (*Oryza sativa* L.). *J. Cereal Sci.* 57, 362–371. doi: 10.1016/j.jcs.2012.12.009

Hou, F. Y., Ji, H., Yu, S. L., and Zhang, H. S. (2007). The 6-phosphogluconate dehydrogenase genes are responsive to abiotic stresses in rice. *J. Integr. Plant Biol.* 49, 655–663. doi: 10.1111/j.1744-7909.2007.00460.x

Hölscher, C., Lutterbey, M. C., Lansing, H., Meyer, T., and Schaewen, A. V. (2016). Defects in peroxisomal 6-phosphogluconate dehydrogenase isoform PGD2 prevent gametophytic interaction in *Arabidopsis thaliana*. *Plant Physiol.* 171, 192–205. doi: 10.1104/pp.15.01301

Huang, J., Zhang, H., Wang, J., and Yang., J. (2003). Molecular cloning and characterization of rice 6-phosphogluconate dehydrogenase gene that is up-regulated by salt stress. *Mol. Biol. Rep.* 30, 223–227. doi: 10.1023/A:1026392422995

Itishree, N., Bijayalaxmi, S., Chinmay, P., Cayalvizhi, B., Seenichamy, R. P., Jawahar, L. K., et al. (2023). Genome-wide analysis of amino acid transporter gene family revealed that the allele unique to the aus variety is associated with amino acid permease 17 (OsAAP17) amplifies both the tiller count and yield in indica rice (Oryza sativa l.). *Agronomy* 13. doi: 10.3390/AGRONOMY13102629

Ji, Y. Y., Huang, W. T., Wu, B. W., Fang, Z. M., and Wang, X. L. (2020). The amino acid transporter AAP1 mediates growth and grain yield by regulating neutral amino acid uptake and reallocation in Oryza sativa. *J. Exp. Bot.* 71, 4763–4777. doi: 10.1093/jxb/eraa256

Li, P., Chen, Y. H., Lu, J., Zhang, C. Q., Liu, Q. Q., and Li, Q. F. (2022a). Genes and their molecular functions determining seed structure, components, and quality of rice. *Rice* 15, 18. doi: 10.1186/s12284-022-00562-8

Li, Y. B., Fan, C. C., Xing, Y. Z., Yun, P., Luo, L. J., Yan, B., et al. (2014). Chalk5 encodes a vacuolar h(+)-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* 46, 398–404. doi: 10.1038/ng.2923

Li, Q. P., Zhang, C. L., Wen, J. C., Chen, L. J., Shi, Y. T., Yang, Q. H., et al. (2023). Transcriptome analyses show changes in gene expression triggered by a 31-bp indel within OsSUT3 5'UTR in rice panicle. *Int. J. Mol. Sci.* 24, 10640. doi: 10.3390/ijms241310640

Li, Y. S., Peng, K. K., Tian, Y., Ren, Z. P., Cang, J., Xu, Q. H., et al. (2022c). Study on lncRNA1808 regulating cold resistance response of PPP rate-limiting enzyme (Ta6PGDH) in winter wheat. *J. Triticeae Crops* 42, 1047–1058. doi: 10.7606/j.issn.1009-1041.2022.09.01

Li, Y., Xiao, J., Chen, L., Huang, X., Cheng, Z., Han, B., et al. (2018). Rice functional genomics research: past decade and future. *Mol. Plant* 11, 359–380. doi: 10.1016/j.molp.2018.01.007

Li, Y. F., Yang, Z., Yang, C. K., Liu, Z. H., Shen, S. Q., Zhan, C. S., et al. (2022b). The NET locus determines the food taste, cooking and nutrition quality of rice. *Sci. Bull.* 67, 2045–2049. doi: 10.1016/j.scib.2022.09.023

Liu, Y., Zhang, N., Li, P., Yu, L., Chen, S. M., Zhang, Y., et al. (2019). Low-cost localized surface plasmon resonance biosensing platform with a response enhancement for protein detection. *Nanomaterials* 9, 1019. doi: 10.3390/nano9071019

Lu, K., Wu, B. W., Wang, J., Zhu, W., Nie, H. P., Qian, J. J., et al. (2018). Blocking amino acid transporter OsAAP3 improves grain yield by promoting outgrowth buds and increasing tiller number in rice. *Plant Biotechnol. J.* 16, 1710–1722. doi: 10.1111/pbi.12907

Nie, S. S., Li, H. Y., Zhang, L., Yang, G., Hang, J. N., Huang, W. T., et al. (2023). Analysis of spatiotemporal expression characteristics of rice amino acid permeability enzyme gene *OsAAP13*. *Mol. Plant Breed.* 21, 7667–7676. doi: 10.13271/j.mpb.021.007667

Pang, Y. L., Ali, J., Wang, X. Q., Franje, N. J., Revilleza, J. E., Xu, J. L., et al. (2016). Relationship of rice grain amylose, gelatinization temperature and pasting properties for breeding better eating and cooking quality of rice varieties. *PloS One* 11, e0168483. doi: 10.1371/journal.pone.0168483

Peng, B., He, L. L., Tian, X. Y., Xin, Q. Q., Sun, Y. F., Song, X. H., et al. (2019). Expression analysis of *OsAAP6* gene under different nitrogen conditions and its effects on grain nutritional quality in rice. *Southwest China J. Agric. Sci.* 32, 1973–1979. doi: 10.16213/j.cnki.scjas.2019.9.002

Peng, B., Kong, D. Y., Huang, Y. J., Jia, Y. Y., Ma, C. X., Song, H., et al. (2018). Bioinformatics analysis of *OsAAP6* gene structure and function in rice. *Southwest China J. Agric. Sci.* 31, 429–436. doi: 10.16213/j.cnki.scjas.2018.3.001

Peng, B., Sun, X. Y., Tian, X. Y., Kong, D. Y., He, L. L., Peng, J., et al. (2023). *OsNAC74* affects grain protein content and various biological traits by regulating *OsAAP6* expression in rice. *Mol. Breed.* 43, 87. doi: 10.1007/s11032-023-01433-w

Peng, B., Wang, L., Q. Fan, C. C., Jiang, G. H., Luo, L. J., Li, Y. B., et al. (2014). Comparative mapping of chalkiness components in rice using five populations across two environments. *BMC Genet.* 15, 49. doi: 10.1186/1471-2156-15-49

Peng, B., Zhang, Q. X., Liu, Y., Zhao, Q., Zhao, J. H., Zhang, Z. G., et al. (2024). *OsAAP8* mutation leads to significant improvement in the nutritional quality and appearance of rice grains. *Mol. Breed.* 44, 34. doi: 10.1007/s11032-024-01473-w

Peng, B., Zhang, Q. X., Tian, X. Y., Sun, Y. F., Huang, X. H., Pang, R. H., et al. (2021). Influencing factors of grain nutritional quality and its genetic improvement strategy in rice. *J. Biotechnol. Res.* 7, 1–11. doi: 10.32861/jbr.16

Riaz, M., Akhter, M., Iqbal, M., Ali, S., Khan, R. A. R., Raza, M., et al. (2018). Estimation of amylose, protein and moisture content stability of rice in multi locations. *Afr. J. Agric. Res.* 13, 1213–1219. doi: 10.5897/AJAR

Ribeiro, C., Hennen-Bierwagen, T. A., Myers, A. M., Cline, K., and Settles, A. M. (2020). Engineering 6-phosphogluconate dehydrogenase improves grain yield in heat-stressed maize. *Proc. Natl. Acad. Sci. United States America* 117, 33177–33185. doi: 10.1073/pnas.2010179117

Shi, S. J., Wang, E. T., Li, C. X., Cai, M. L., Cheng, B., Cao, C. G., et al. (2022). Use of protein content, amylose content, and RVA parameters to evaluate the taste quality of rice. *Front. Nutr.* 8, 758547. doi: 10.3389/fnut.2021.758547

Su, L., Liu, X., An, J. P., Li, H. H., Zhao, J., Hao, Y. J., et al. (2016). Molecular cloning and functional analysis of a 6-phosphogluconate dehydrogenase gene *Md6PGDH1* in apple. *Acta Hortic. Sin.* 43, 1225–1235. doi: 10.16420/j.issn.0513-353x.2016-0202

Tanksley, S. D., and Kuehn, G. D. (1985). Genetics, subcellular localization, and molecular characterization of 6-phosphogluconate dehydrogenase isozymes in tomato. *Biochem. Genet.* 23, 441–454. doi: 10.1007/BF00499085

Taylor, M. R., Reinders, A., and Ward., J. M. (2015). Transport function of rice amino acid permeases (AAPs). *Plant Cell Physiol.* 56, 1355–1363. doi: 10.1093/pcp/pcv053

Tegeder, M., and Hammes, U. Z. (2018). The way out and in: phloem loading and unloading of amino acids. *Curr. Opin. Plant Biol.* 43, 16–21. doi: 10.1016/j.pbi.2017.12.002

Tian, Y., Peng, K. K., Bao, Y. Z., Zhang, D., Meng, J., Wang, D. J., et al. (2021). Glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase genes of winter wheat enhance the cold tolerance of transgenic Arabidopsis. *Plant Physiol. Biochem.* 161, 86–97. doi: 10.1016/j.plaphy.2021.02.005

Tong, C., Gao, H., Luo, S., Liu, L., and Bao, J. (2019). Impact of postharvest operations on rice grain quality: a review. *Compr. Rev. Food Sci. Food Saf.* 18, 626–640. doi: 10.1111/1541-4337.12439

Wang, J. W., Tian, H., Yu, X., and Zheng, L. P. (2017). Glucose-6-phosphate dehydrogenase plays critical role in artemisinin production of Artemisia annua under salt stress. *Biol. Plantarum* 61, 529–539. doi: 10.1007/s10535-016-0674-7

Wang, J., Wu, B. W., Lu, K., Qian, J. J., Chen, Y. P., and Fang, Z. M. (2019). The amino acid permease 5 (OsAAP5) regulates tiller number and grain yield in rice. *Plant Physiol.* 180, 1031–1045. doi: 10.1104/pp.19.00034

Wang, S. Y., Yang, Y. H., Guo, M., Zhong, C. Y., Yan, C. J., and Sun, S. Y. (2020). Targeted mutagenesis of amino acid transporter genes for rice quality improvement using the CRISPR/Cas9 system. *Crop J.* 8, 457–464. doi: 10.1016/j.cj.2020.02.005

Wang, Z., Zhou, Y., Ren, X.-Y., Wei, K., Fan, X. L., Huang, L. C., et al. (2022). Co-overexpression of two key source genes, *OsBMY4* and *OsISA3*, improves multiple key traits of rice seeds. *J. Agric. Food Chem.* 71, 615–625. doi: 10.1021/acs.jafc.2c06039

Wang, Y. P., Wei, Z. Y., and Xing, S. C. (2018). Stable plastid transformation of rice, a monocot cereal crop. *Biochem. Biophys. Res. Commun.* 503, 2376–2379. doi: 10.1016/j.bbrc.2018.06.164

Wei, Y., Cheng, X. Y., Li, Z. L., Wang, Y. P., Shi, J. L., Wu, Z. M., et al. (2010). Cloning of 6-phosphogluconate dehydrogenase gene cDNA fragments from cucumber and expression analysis. *J. Nanjing Agric. Univ.* 33, 37–42.

Xia, D., Wang, Y. P., Shi, Q. Y., Wu, B., Yu, X. M., Zhang, C. Q., et al. (2022). Effects of *Wx* genotype, nitrogen fertilization, and temperature on rice grain quality. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.901541

Xiong, Q. X., Zhong, L., Shen, T. H., Cao, C. H., He, H. H., and Chen, X. R. (2019). iTRAQ-based quantitative proteomic and physiological analysis of the response to N deficiency and the compensation effect in rice. *BMC Genomics* 20, 681. doi: 10.1186/s12864-019-6031-4

Xu, H. Y., Li, S. F., Kazeem, B. B., Ajadi, A. A., Luo, J. J., Yin, M., et al. (2022). Five rice seed-specific *NF-YC* genes redundantly regulate grain quality and seed germination via interfering gibberellin pathway. *Int. J. Mol. Sci.* 23, 8382. doi: 10.3390/ijms23158382

Yang, G., Wei, X. L., and Fang, Z. M. (2022). Melatonin mediates axillary bud outgrowth by improving nitrogen assimilation and transport in rice. *Front. Plant Sci.* 13, 900262. doi: 10.3389/fpls.2022.900262

Yang, X. Y., Yang, G., Wei, X. L., Huang, W. T., and Fang, Z. M. (2023a). OsAAP15, an amino acid transporter in response to nitrogen concentration, mediates panicle branching and grain yield in rice. *Plant Sci.* 330, 111640. doi: 10.1016/j.plantsci.2023.111640

Yang, Y. H., Zhang, Y., Sun, Z. X., Shen, Z. Y., Li, Y. G., Guo, Y. F., et al. (2023b). Knocking out OsAAP11 to improve rice grain quality using CRISPR/Cas9 system. *Int. J. Mol. Sci.* 24, 14360. doi: 10.3390/ijms241814360

Zhang, C. Q., Zhu, J. H., Chen, S. J., Fan, X. L., Li, Q. F., Lu, Y., et al. (2019). Wxlv, the ancestral allele of rice waxy gene. *Mol. Plant* 12, 1157–1166. doi: 10.1016/j.molp.2019.05.011

Zhang, G. Y., Cheng, Z. J., Zhang, X., Guo, X. P., Su, N., Jiang, L., et al. (2011). Double repression of soluble starch synthase genes SSIIa and SSIIIa in rice (Oryza sativa l.) uncovers interactive effects on the physicochemical properties of starch. *Genome* 54, 448–459. doi: 10.1139/g11-010

Zhang, W., Liu, Y. X., Luo, X. L., and Zeng, X. F. (2022). Pasting, cooking, and digestible properties of; japonica; rice with different amylose contents. *Int. J. Food Properties.* 25, 936–947. doi: 10.1080/10942912.2022.2069806

Zhang, X. F., Zhu, Q., Hua, Y. Y., Jia, L. H. Y., Qiu, S. Y., Chen, Y. J., et al. (2024). Screening and validation of OsCYP22 interacting proteins in rice. *Acta Agronomica Sin.* 50, 1628–1634. doi: 10.3724/SP.J.1006.2024.32041

Zhao, H., Li, Z. X., Wang, Y. Y., Wang, J. Y., Xiao, M. G., Liu, H., et al. (2022b). Cellulose synthase-like protein OsCSLD4 plays an important role in the response of rice to salt stress by mediating abscisic acid biosynthesis to regulate osmotic stress tolerance. *Plant Biotechnol. J.* 20, 468–484. doi: 10.1111/pbi.13729

Zhao, M., Lin, Y., and Chen, H. (2020). Improving nutritional quality of rice for human health. *Theor. Appl. Genet.* 133, 1397–1413. doi: 10.1007/s00122-019-03530-x

Zhao, D. S., Zhang, C. Q., Li, Q. F., and Liu, Q. Q. (2022a). Genetic control of grain appearance quality in rice. *Biotechnol. Adv.* 60, 108014. doi: 10.1016/j.biotechadv.2022.108014

Zhou, H., Xia, D., and He, Y. Q. (2019). Rice grain quality—traditional traits for high quality rice and health-plus substances. *Mol. Breeding.* 40, 201–204. doi: 10.1007/s11032-019-1080-6

Zhou, T. Y., Zhou, Q., Li, E. P., Yuan, L. M., Wang, W. L., Zhang, H., et al. (2020). Effects of nitrogen fertilizer on structure and physicochemical properties of 'super' rice starch. *Carbohydr. Polymers.* 239, 116237. doi: 10.1016/j.carbpol.2020.116237

Zhu, X. P., Sun, M. H., An, J. P., Lu, J., Qi, C. H., You, C. X., et al. (2019). Functional identification of 6-phosphogluconate dehydrogenase *Md6PGDH6* gene in apple. *J. Fruit Sci.* 36, 1453–1462. doi: 10.13925/j.cnki.gsxb.20190148

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 August 2024
DOI 10.3389/fpls.2024.1451897

OPEN ACCESS

EDITED BY
Hao Zhou,
Sichuan Agricultural University, China

REVIEWED BY
Zhiqiang Li,
Chinese Academy of Agricultural Sciences, China
Xiaofeng Su,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Xiaobo Zhu
myself19861025@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 20 June 2024
ACCEPTED 22 July 2024
PUBLISHED 06 August 2024

CITATION
Yin J, Zhang C, Zhang Q, Long F, Hu W, Zhou Y, Mou F, Zhong Y, Wu B, Zhu M, Zou L and Zhu X (2024) A single amino acid substitution in the AAA-type ATPase LRD6-6 activates immune responses but decreases grain quality in rice.
Front. Plant Sci. 15:1451897.
doi: 10.3389/fpls.2024.1451897

A single amino acid substitution in the AAA-type ATPase LRD6-6 activates immune responses but decreases grain quality in rice

Junjie Yin[1†], Cheng Zhang[1†], Qianyu Zhang[1], Feiyan Long[1], Wen Hu[2], Yi Zhou[2], Fengying Mou[2], Yufeng Zhong[1], Bingxiu Wu[1], Min Zhu[1], Lijuan Zou[1,3] and Xiaobo Zhu[1*]

[1]State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University at Wenjiang, Chengdu, Sichuan, China, [2]Agricultural and Rural Bureau, Hanyuan County Government, Yaan, Sichuan, China, [3]Ecological Security and Protection Key Laboratory of Sichuan Province, Mianyang Teachers' College, Mianyang, China

Plant *spotted leaf* (*spl*) mutants are useful to reveal the regulatory mechanisms of immune responses. Thus, in crop plants, their agronomic traits, especially the grain quality are usually ignored. Here, we characterized a rice *spl* mutant named *spl-A* (*spotted leaf mutant from A814*) that shows autoimmunity, broad-spectrum disease resistance and growth deterioration including decreased rice quality. A single nucleotide mutation of C1144T, which leads to change of the 382nd proline to serine, in the gene encoding the ATPases associated with diverse cellular activities (AAA)-type ATPase LRD6-6 is responsible for the phenotype of the *spl-A* mutant. Mechanistically, this mutation impairs LRD6-6 ATPase activity and disrupts its interaction with endosomal sorting complex required for transport (ESCRT)-III subunits OsSNF7.1/7.2/7.3. And thus, leading to compromise of multivesicular bodies (MVBs)-mediated vesicle trafficking and accumulation of ubiquitinated proteins in both leaves and seeds of *spl-A*. Therefore, the immune response of *spl-A* is activated, and the growth and grain quality are deteriorated. Our study identifies a new amino acid residue that important for LRD6-6 and provides new insight into our understanding of how MVBs-mediated vesicle trafficking regulates plant immunity and growth, including grain quality in rice.

KEYWORDS

AAA-type ATPase, multivesicular bodies, vesicle trafficking, immune response, grain quality, rice

Introduction

During the long-term of plant-pathogen interaction, plants have evolved with a layered immune system to detect and fight with various of pathogens. Pattern recognition receptors (PRRs) localized at the plasm membrane function as the first layer to detect pathogen derived molecular patterns and activate pattern-triggered immunity (PTI) (Jones and Dangl, 2006; Zhao et al., 2022). Intracellular receptors such as nucleotide-binding, leucine-rich repeat receptors (NLRs) function as the second layer to monitor pathogen secreted effectors to activate the more potent immune responses, effector-triggered immunity (ETI) (Jones and Dangl, 2006; Zhao et al., 2022). Once the immune responses are activated, plant will increase the expression of *pathogenesis-related* (*PR*) genes, induce the production of reactive oxygen species (ROS) and promote the synthesis of antimicrobial metabolites etc (Tsuda and Katagiri, 2010; Yuan et al., 2021). Finally, the immunity system may launch hypersensitive response (HR) to trigger cell death in and around the infection sites to restrict pathogen invasion and prevent disease development (Mur et al., 2008; Balint-Kurti, 2019).

It's interesting that in the absence of pathogen infection, insect feeding, stress or mechanical damages, many plant mutants have been found to exhibit spontaneous, HR-like cell death spots on their bodies (Zhu et al., 2020). These mutants are named based on their phenotype as *spotted leaf* (*spl*), *lesion mimic mutant* (*lmm*), *accelerated cell death* (*acd*) and others (Bruggeman et al., 2015). Since these *spl* mutants usually carry the characteristics of activated immune responses, they are useful to uncover the regulatory mechanisms of plant immune responses. More than 35 genes that responsible for *spl* phenotype have been identified in rice at present (Zhu et al., 2020; Jiang et al., 2022; Ren et al., 2022; Zheng et al., 2022; Wang et al., 2023). Studies on these genes have revealed many processes or pathways are involved in plant immune responses regulation. These include the gene transcription and protein translation process, post-translational protein modification, the vesicle trafficking pathways, and several cellular metabolic pathways (Zhu et al., 2020). During the study of these genes or mutants, it has also found that these genes are usually involved in the regulation of plant growth as well (Zhu et al., 2020). However, little attention has been paid for their roles in plant growth.

The vesicle trafficking apparatus are critical for the growth and development of eukaryotes. Up to now, at least five vesicle-mediated trafficking pathways have been identified in plant. The coat protein I (COPI)-coated vesicles and the COPII-coated vesicles are involved in endoplasmic reticulum (ER)-Golgi trafficking, the clathrin-coated vesicles (CCVs) mediates endocytosis and post-Golgi trafficking, the exocyst positive organelles (EXPO) regulates exocytosis, and the multivesicular bodies (MVBs) modulates endocytosis, exocytosis, recycling and degradation of cargoes (Aniento et al., 2021; Zhu et al., 2023).

By studying of the *spl* mutants, several vesicle trafficking pathways have been found to be essential for rice immunity. SPL35 is a CUE domain-containing protein that interacts with the coatomer subunits δ-COP1 and δ-COP2 of the COPI vesicle (Ma et al., 2019). Suppression any of the three genes will lead to *spl* phenotype (Ma et al., 2019), providing evidence that the COPI trafficking pathway is involved in the regulation of immunity and HR-like cell death in rice. Loss of function of the *trans*-Golgi network (TGN)-derived CCVs components, SPL28, OsSCYL2 or OsCHC1 results in autoimmunity and cell death of rice (Qiao et al., 2010; Yao et al., 2022). OsSEC3A is a core subunit of the exocyst complex. Knockout of *OsSEC3A* causes enhanced immune responses and HR-like cell death likely due to the disruption of the EXPO-mediated exocytosis (Ma et al., 2017). The AAA-type ATPase LRD6-6 is a rice homology of VPS4/SKD1 which crucial for MVBs-mediated trafficking of ubiquitinated plasm membrane proteins (Zhu et al., 2016). During the formation of the MVBs, the ESCRT-I, -II, and -III complexes are sequentially recruited to coordinately drive vesicle maturation. At the final step, the VPS4/SKD1 assembles with the ESCRT-III at the neck region of the limiting membrane to hydrolyzes ATP and thus provide energy to facilitate neck scission and disassembly of the ESCRT-III complex (Piper and Katzmann, 2007; Hurley, 2008; Schmidt and Teis, 2012). Mutation of *Lrd6-6* in rice leads to autoimmunity and HR-like cell death (Zhu et al., 2016), indicating the important roles of LRD6-6 and MVBs-mediated trafficking in rice immune responses.

Besides their roles in plant immunity, the vesicle trafficking pathways have also been found to regulate rice chalkiness through regulating the transportation of storage proteins in endosperm, thus contributing to grain quality formation (Zhu et al., 2023). Rice endosperm cells contain two types of protein bodies, protein body I (PBI) and PBII, which are responsible for protein storage in seed. PBIs storage prolamins while PBIIs mainly accumulate glutelins and globulins (Li et al., 2022). The delivery of the storage proteins into PBIs and PBIIs are tightly regulated by vesicle trafficking. This is mainly achieved by studying of *glutelin precursor accumulation* (*gpa*) mutants with abnormal endosperm and decreased grain quality in rice (Li et al., 2022; Zhu et al., 2023). Interesting, loss of function of an ESCRT-II complex subunit OsVPS22 also leads to chalky endosperm likely for impaired grain filling (Zhang et al., 2013), suggesting that the MVBs-mediated vesicle trafficking pathway may also likely involved in storage protein transportation in rice endosperm cells. However, our knowledges about whether and how MVBs-mediated vesicle trafficking regulate rice endosperm development and grain quality formation are still remain largely unknown.

In this study, we isolated and characterized the rice spotted leaf mutant *spl-A*, which shows autoimmunity, broad-spectrum disease resistance and growth deterioration including decreased rice quality. By combining map-based cloning strategy and bulked segregant analysis (BSA) sequencing, we found a single nucleotide mutation of C1144T in the gene encoding the AAA-type ATPase LRD6-6 is responsible for the phenotype of the *spl-A* mutant. This mutation leads to change of the 382nd proline to serine, impairs LRD6-6 ATPase activity and disrupts its interaction with ESCRT-III subunits OsSNF7.1/7.2/7.3. And finally, result in compromise of MVBs-mediated vesicle trafficking and accumulation of ubiquitinated proteins in *spl-A*. Collectively, our results identify a new amino acid residue that important for LRD6-6 and point to a novel function for MVBs-mediated vesicle trafficking regulates grain quality in rice.

Materials and methods

Plant materials and growth conditions

The *spl-A* mutant was isolated from the methanesulfonate (EMS) treated *japonica* rice A814 (*XA21-OsSERK2Ri* in Kitaake) (Caddell et al., 2017). Two F_2 populations derived from the cross Jodan×*spl-A* and *spl-A*×Jodan were used for genetic analyses and mapping of the *spl-A* locus. All plants for mapping were grown in the controlled fields at Sichuan Agricultural University in Wenjiang, Chengdu, China.

For major agronomic traits analysis, plant materials were cultivated in the controlled fields with 6 rows (12 plants per row) at Wenjiang district, Chengdu, China. At maturity, forty plants in the middle row were selected for plant height, tiller number and panicle length determination, and then they were harvested respectively and air-dried. For seed setting rate and 1,000-grain weight, five A814 and *spl-A* plants were randomly and respectively selected. Scanning electron microscopy observation was performed as described before (Ren et al., 2020). For grain quality determination, the seeds from all the forty plants of A814 and *spl-A* were mixed respectively, and chalky grain percentage and chalkiness degree were detected using the WSeen SC-E type rice appearance quality detection analyzer (Wanshen Detection, Hangzhou, China) with three technical repetitions. Student's *t*-test was used to compare the differences between A814 and *spl-A* and the data were analyzed by using Microsoft Excel 2019.

Histochemical analysis

Trypan blue staining and DAB (3, 3'-diamiobenzidine) staining were performed on fresh leaves following the method as described previously (Zhu et al., 2016). Briefly, for trypan blue staining, samples were submerged in lactic acid-phenol-trypan blue solution (2.5 mg/ml trypan blue, 25% (w/v) lactic acid, 23% water-saturated phenol and 25% glycerol in H_2O) and were boiled in water for 2 min, then de-stained with solution containing 30% (w/v) chloral hydrate for 3 days with multiple exchanges of the solution. After distained completely, the samples were then equilibrated with 50% glycerol for five hours followed by photo-picture taken. For DAB staining, leaf samples were immersed in 1 mg/ml DAB containing10 mM MES (pH 6.5) for 12 h in the dark at 30°C. Then the leaf samples were transferred to solution containing 90% ethanol and 10% glycerol at 90°C until chlorophyll was completely removed. Images were acquired by a microscope (Zeiss imager A2).

RNA isolation and RT-qPCR

For determination of gene expression levels, the desired samples were harvested and total mRNA were respectively extracted using TRIzol agent (15596018CN, Invitrogen, Carlsbad, USA) following the procedures described by the manufacturer. The mRNA was then subject to reverse transcription to synthesize first-strand cDNA by using the PrimeScript™ RT reagent Kit with gDNA Eraser (RR047A, Takara, Beijing, China) for genomic DNA removal.

The reverse transcription-quantitative PCR (RT-qPCR) was conducted using a Bio-Rad CFX96 Real-Time System coupled to a C1000 Thermal Cycler (Bio-Rad, Hercules, USA). The reference gene *Ubiquitin* (*Ubq*) was used as reference (Zhu et al., 2016). Primer sequences are listed in Supplementary Table 1.

Disease resistance evaluation

For blast resistance, rice leaves from three-week-old seedlings of A814 and *spl-A* plants were used. The inoculation was performed as described by Li et al (Li et al., 2020). Briefly, the spore concentration of the blast was adjusted to 5×10^5 conidia mL^{-1}. 4 μL of the spore suspension was dipped onto the leaves with three spots per leaf. Then the leaves were kept in a culture dish contain 0.1% 6-benzylaminopurin (6-BA) for 5 days. The lesion length was measured and the relative fungi biomass was detected and calculated according to the method reported before (Li et al., 2020). The fungal isolates Zhong10-8-14, ZE-1, 0755-1-1 and 99-20-2 that are compatible with A814 were respectively used for blast resistance evaluation.

For bacterial blight disease, the *Xoo* strains P2, P6 and Xoo-4, which are compatible with A814 were used for inoculation. *Xoo* bacterial suspensions of OD600 = 0.5 were used to inoculate by using the scissors-dip method as described previously (Zhu et al., 2016). The lesion lengths and bacterial populations were determined respectively at day 0, 7, 14 post inoculation.

Genetic analysis and cloning of the locus *Spl-A*

Two F_2 populations derived from the cross Jodan×*spl-A* and *spl-A*×Jodan were used for genetic analyses and mapping of the *spl-A* locus as described before (Zhu et al., 2016). The SSR primers were synthesized according to the information from Gramene database (http://www.gramene.org/microsat).

For BSA sequencing, leaf samples of 30 F_2 individuals with the *spl-A* phenotype and the wild-type phenotype were respectively mixed at an equal ratio and subjected to whole-genome resequencing and analysis at Biomarker Technologies Corporation (Beijing, China).

For candidate gene validation, the DNA segments containing the mutation sites were amplified using gene-specific primers and sequenced at Tsingke Biotech (Chengdu, China). The genotypes of the *lrd6-6*/*spl-A* F_1 plants were also confirmed by sequencing of *spl-A* and *lrd6-6*-specific PCR-based agarose gel electrophoresis respectively. All the primers used for gene mapping and genotyping are listed in Supplementary Table 1.

Plasmid construction

The desired DNA segments were amplified by using the Phanta Super-Fidelity DNA Polymerase (P505-d1, Vazyme Biotech, Nanjing, China). The segments were then inserted into the vectors to produce the final plasmids by using the ClonExpress II One Step Cloning Kit (C112-02, Vazyme Biotech, Nanjing, China). All the plasmids were then confirmed by sequencing at Tsingke Biotech (Chengdu, China) subsequently. The primers and vectors used were listed in Supplementary Table 1.

ATPase activity assay

ATPase activity determination assay was performed as described previously (Zhu et al., 2016). The truncated protein His-LRD6-6(125-487)^{P382S} (covering the ATPase domain) was purified and subjected to this analysis. His-LRD6-6(125-487) and His-LRD6-6(125-487)^{E315Q} were included as positive and negative control respectively. ATPase activity of the recombinant proteins were measured by the malachite green-based colorimetric method (Zhu et al., 2016). The relative ATPase activity was calculated by setting the activity of His-LRD6-6(125-487) as 100%.

Protein interaction detection and subcellular localization

Yeast two-hybrid (Y2H) assay and bimolecular fluorescence complementation (BiFC) were performed as described previously (Zhu et al., 2016).

For subcellular localization, plasmids were introduced into *Nicotiana benthamiana* leaf by agroinfiltration. Fluorescence was examined under a confocal microscopy (Leica STELLARIS STED, Germany) 48 h post transformation. The fluorescence intensity were measured by using ImageJ (National Institutes of Health, USA).

Protein extraction and immunoblot analyses

For protein extraction from plant leaf samples, the denaturing buffer (50 mM Tris-HCl pH=7.5, 150 mM NaCl, 0.1% NP-40, 4 M urea and 1 mM PMSF) was used (Liu et al., 2010). For the dry seeds, the extraction buffer (125 mM Tris-HCl pH=6.8, 4% SDS, 4 M urea and β-mercaptoethanol) described before was used (Pan et al., 2021). The total protein content was determined by the Protein Content Assay Kit (BC3185, Solarbio, Beijing, China).

The proteins were resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis and followed by Coomassie brilliant blue staining or electro transfer to nitrocellulose membranes to perform immunoblot analyses using anti-His antibody (ab18184, Abcam, Shanghai, China), anti-Ub (P4D1) antibody (ab303664, Abcam, Shanghai, China) and anti-HSP82 antibody (AbM51099-31-PU, Beijing Protein Innovation, Beijing, China).

Results

Characterization of the *spl-A* mutant in rice

The spotted leaf mutant which we named *spl-A* (*spotted leaf mutant from A814*) was isolated from ethyl methanesulfonate (EMS) treated *japonica* rice A814 for its visible lesion mimic phenotype on the leaves when growing in the field. The *spl-A* mutant exhibited reddish-brown lesion spots on the leaves about 20 days after sowing and the lesion spots expand through the entire plants along with development (Figures 1A, B). To examine whether the agronomic traits including grain quality were affected in *spl-A*, we compared the chalky grain percentage, chalky grain degree, plant height, tiller number, panicle length, seed setting rate, 1,000-grain weight, and grain size between A814 and *spl-A*. It was found that the chalky grain percentage and chalky grain degree were significantly increased (Figures 1C–E), indicating that the grain quality of *spl-A* was seriously deteriorated. The plant height, seed setting rate, 1,000-grain weight and grain length of *spl-A* were also decreased prominently when compared to those of the wild-type A814, while the tiller number, panicle length and grain width of *spl-A* were not affected (Supplementary Figure 1).

Grain chalkiness is the opaque part of translucent endosperm in grains that usually caused by altered package of the starch granules (Yao et al., 2020). We then analyzed the starch granules organization of *spl-A* seed. Scanning electron microscopy revealed that the *spl-A* mutant endosperm was composed of loosely arranged and round-shaped compound starch granules, while that of the wild-type A814 was tightly packed and polyhedral shaped (Figure 1F). Grain chalkiness often affects processing quality (Li et al., 2022). Indeed, the head rice rate of *spl-A* was prominently decreased when compared to A814 (Figures 1G, H). Taken together, these results indicate that *Spl-A* regulates multiple aspects of rice development including grain quality.

The *spl-A* mutant presents autoimmunity and broad-spectrum disease resistance

The *spl* mutants usually exhibit excessive accumulation of hydrogen peroxide (H_2O_2) and autoimmunity which then lead to programmed cell death (Li et al., 2017; Yao et al., 2022). We thus employed the DAB (3, 3'-diamiobenzidine) staining and trypan blue staining methods to examine the accumulation of H_2O_2 and the existence of cell death in *spl-A* leaves respectively at 15 days after sowing, before visible lesions appeared. DAB staining analysis detected extensive stains in leaves of *spl-A* when compared to A814 (Figure 2A). Trypan blue staining also detected many blue staining spots on the leaves of *spl-A* (Figure 2B). To determinate whether the immune responses of the *spl-A* mutant was activated, we then compared the expression levels of immunity related genes, such as the *pathogenesis-related* (*PR*) genes, *OsNAC4*, *OsPBZ1*, *OsPR1a* and *OsPR10* between *spl-A* and A814 at 15 days after sowing (Li et al., 2017; Tao et al., 2021). We found that the

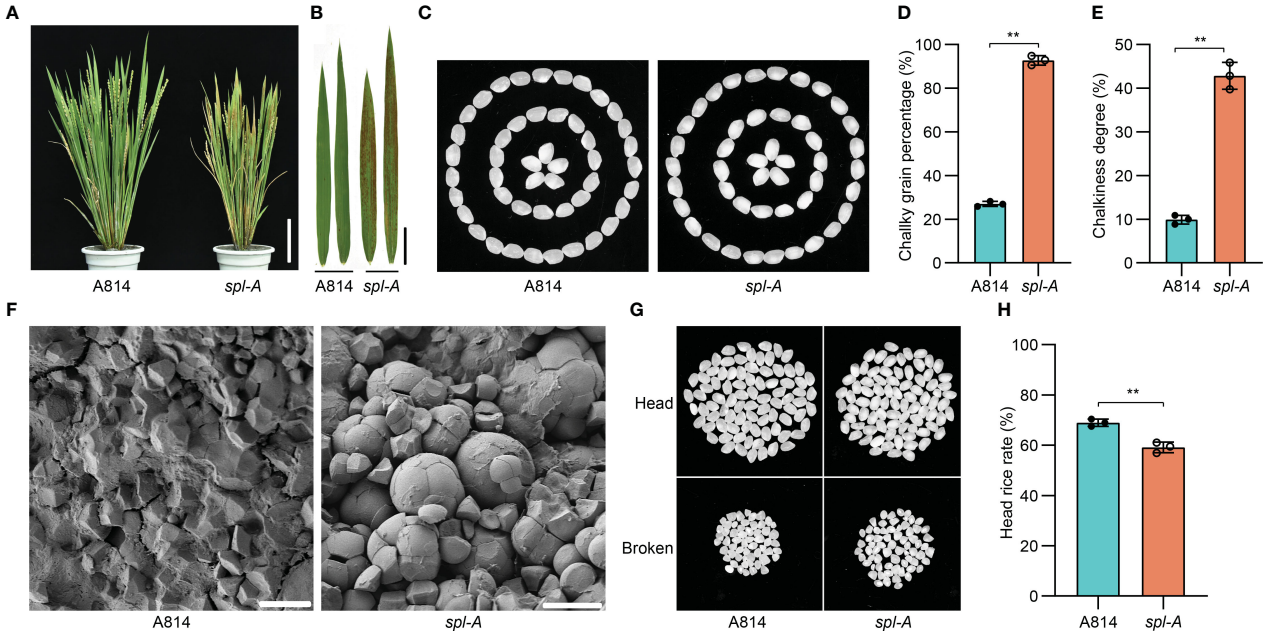


FIGURE 1
Phenotypic characterization of the *spl-A* mutant. **(A, B)** Photographs of the wild-type A814 and the *spl-A* mutant plants **(A)**, and their leaves **(B)** at the filling stage. Bar = 10 cm in **(A)**, Bar = 2 cm **(B)**. **(C)** The grain appearance of A814 and *spl-A*. **(D, E)** Comparison on the chalky grain percentage **(D)** and chalky grain degree **(E)** between A814 and *spl-A* (mean ± s.d.). **(F)** Scanning electron microscopy images of transverse sections on the wild-type A814 and the *spl-A* mutant dry seeds. Bars = 10 μm. **(G, H)** The head rice rate of the *spl-A* mutant. Representative images of 100 dehulled and milled rice grains per genotype were separated into head and broken grains **(G)**. **(H)** Statistical comparison of head rice rate between A814 and *spl-A* seeds (mean ± s.d.). All statistics were analyzed by Student's *t*-test for *P* values (**, $P <= 0.01$; *, $P <= 0.05$; ns, no significant differences).

expression levels of all the *PR* genes analyzed were largely increased in *spl-A* when compared to A814 (Figure 2C). These results suggest that the *spl-A* mutant accumulates an excess of H_2O_2 and high levels of *PR* genes which may cause cell death and enhance plant basal disease resistance.

We then challenged the *spl-A* mutant with the fungal pathogen *Magnaporthe oryzae* (*M. oryzae*) and the bacterial pathogen *Xanthomonas oryzae* pv *oryzae* (*Xoo*) which cause rice blast and bacterial blight diseases respectively. As expected, the lesion length and the relative fungal growth of leaves inoculated with *M. oryzae* strains (Zhong10-8-14, ZE-1, 0755-1-1 and 99-20-2) which compatible with A814 were all dramatically reduced in the *spl-A* mutant when compared with those of A814 (Figures 2D–F and Supplementary Figure 2). When inoculated with compatible *Xoo* strains (P2, P6 and Xoo-4), the *spl-A* mutant also exhibited enhanced bacterial resistance, showing much shorter lesion length and less bacteria than A814 (Figures 2G–I and Supplementary Figure 3). Collectively, these results indicate that the *spl-A* mutant possesses enhanced immunity and broad-spectrum disease resistance.

Cloning of the *Spl-A* gene by combining map-based cloning strategy and BSA sequencing

In order to isolate the *Spl-A* gene, we reciprocally crossed the *indica* rice Jodan with the *spl-A* mutant and developed two F_2 populations. Then, genetic analysis on the *spl-A* locus was carried out. We found the phenotype of *spl-A* was controlled by a single recessive nuclear locus through genetic analysis (Supplementary Table 2). Next, map-based cloning strategy and the bulked segregant analysis (BSA) sequencing method were simultaneously adopted to accelerate the cloning of *Spl-A* gene. The candidate gene was mapped in the interval with a physical distance of 3 Mb between the SSR markers RM19274 and RM19472 on chromosome 6 through map-based cloning strategy (Figure 3A). Combining with BSA sequencing analysis, we then identified the gene *LOC_Os06g03940* which harbored two single nucleotide mutations in the *spl-A* mutant was likely the candidate of *Spl-A* gene (Figures 3B, C). The C147T mutation in the first exon of *LOC_Os06g03940* is a nonsense mutation, while the C1144T mutation in the 10th exon leads to change of 382nd proline (P) to Serine (S) (Figures 3B, C).

The gene *LOC_Os06g03940* encoding an AAA-type ATPase has been previously reported by us as *Lesion resembling disease 6-6* (*Lrd6-6*). The *lrd6-6* mutant harbors a 1446-nt tandem repeat in *LOC_Os06g03940* gene, resulting in autoimmunity and broad-spectrum disease resistance (Zhu et al., 2016). We then found that the amino acid mutation P382S caused by the C1144T mutation in *spl-A* was conserved among the AAA-type ATPase family (Figure 3D). To further test whether *spl-A* and *lrd6-6* were allelic mutants caused by mutation of *LOC_Os06g03940*, we crossed *lrd6-6* with *spl-A* and investigated the phenotype of the F_1 progenies. We found that all the F_1 progenies from *lrd6-6*×*spl-A* exhibited lesion spots on the leaves as that of the *lrd6-6* and *spl-A*

10.3389/fpls.2024.1451897

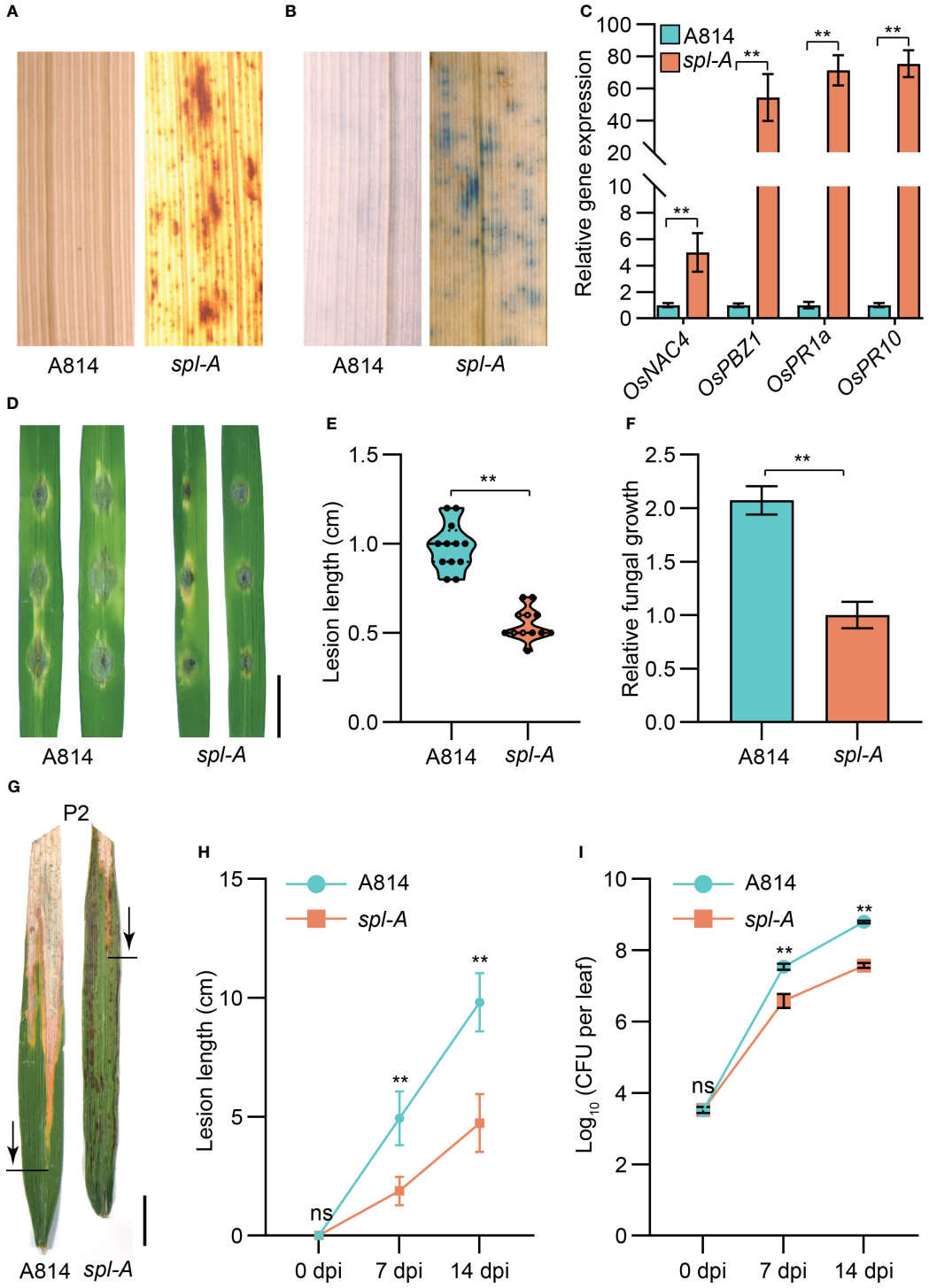


FIGURE 2
Determination on the immunity and disease resistance of the *spl-A* mutant. **(A, B)** The DAB staining **(A)** and trypan blue staining **(B)** analyses on A814 and *spl-A* at 15 days after sowing. **(C)** Comparison the expression levels of the *pathogenesis-related* (*PR*) genes between A814 and *spl-A* at 15 days after sowing. The expression levels of the *PR* genes were normalized to the reference gene *Ubiquitin* (mean ± s.d., n = 3 technical repetitions). **(D–F)** Evaluation on the blast resistance of the *spl-A* mutant. Punch inoculation method was employed. The blast fungal isolate Zhong10-8-14 that is compatible with A814 was used. Photograph of representative lesions **(D)** were shown. Bar = 1 cm. Lesion length **(E)** and the relative fungal growth **(F)** were measured at 5 day-post-inoculation (dpi). The relative fungal growth was determinated as fungi MoPot2 DNA to rice OsUbq DNA by qPCR (mean ± s.d., n = 3 technical repetitions). **(G–I)** Evaluation on the resistance of the *spl-A* mutant to bacterial blight disease. Photographs of representative leaves were taken at 14 dpi with *Xanthomonas oryzae pv oryzae* (*Xoo*) strain P2, which is compatible with A814 **(G)**. Bar = 1 cm. Disease lesion lengths **(H)** and bacterial populations **(I)** of the *spl-A* mutant and A814 were measured at 0, 7 and 14 dpi respectively (mean ± s.d., n = 12 for lesion lengths and n = 3 for bacterial populations). All statistics were analyzed by Student's t-test for P values (**, P <= 0.01; ns, no significant differences).

10.3389/fpls.2024.1451897

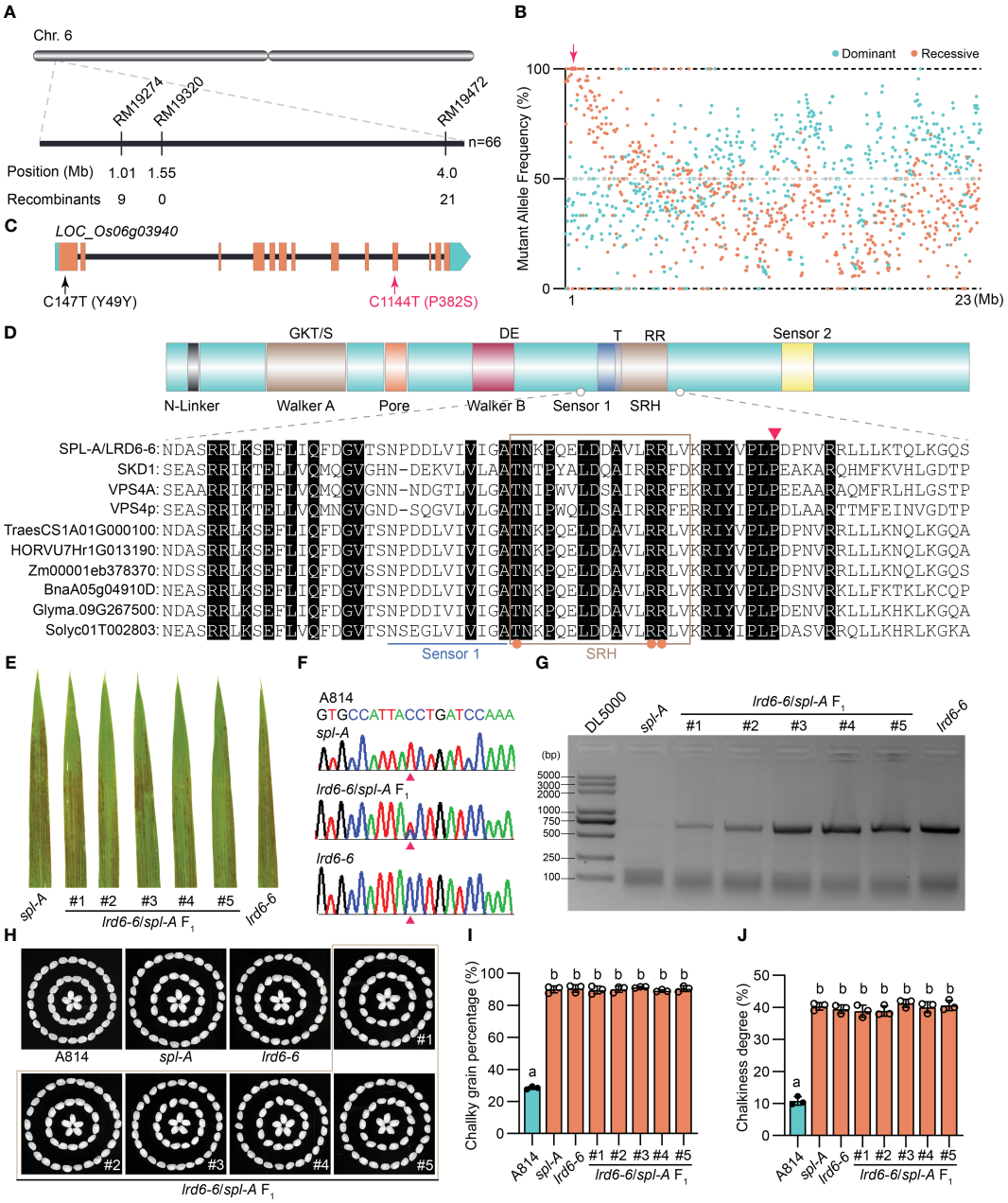


FIGURE 3
Molecular cloning and functional identification of *Spl-A* gene. **(A)** Preliminary mapping of the *Spl-A* locus. The *Spl-A* locus was delimited to a 3 Mb interval between SSR markers RM19274 and RM19472 on chromosome 6. The molecular markers and the number of recombinants is respectively shown. **(B)** Bulked segregant analysis (BSA) of the *Spl-A* locus. The candidate gene in the preliminary mapping interval is boxed and indicated by red arrow. **(C)** The schematic diagram of the *Spl-A* candidate gene identified by our combined strategy. The gene *LOC_Os06g03940* harbors two mutation sites in the *spl-A* mutant. One mutation is a nonsense mutation (C147T) in the first exon and another one is missense mutation (C1144T) in the 10th exon that leads to change of Proline (P) to Serine (S). **(D)** Conserved structure of the LRD6-6 protein. The characteristics of typical AAA ATPase (upper panel) and partial sequence alignment analysis on LRD6-6 with some known AAA ATPases (lower panel) are respectively shown. The key elements of Walker A and B motifs, the Pore, Sensors 1 and 2, and the second region of homology (SRH) are respectively marked. The conserved residues in the SRH motif are indicated by orange dots. The 382th amino acid P that mutated to S in *spl-A* is indicated by red triangle. **(E–J)** Allelic analysis between the *spl-A* and *lrd6-6* mutants. The leaves of *spl-A*, *lrd6-6* and five F_1 progenies from the crossing of *lrd6-6* and *spl-A* were respectively shown **(A)**. The genotypes of the five F_1 progenies were detected **(F, G)**. PCR-based sequencing was employed to determine the C1144T mutant in *spl-A***(F)**. The representative DNA sequencing chromatograms for the wild-type A814, the mutants *spl-A* and *lrd6-6*, and the F_1 progenies are respectively shown. PCR-based agarose gel electrophoresis was used to distinguish the genotype of *lrd6-6*. The grain appearance of *spl-A*, *lrd6-6* and five F_1 progenies from the crossing of *lrd6-6* and *spl-A* were also determined **(H–J)**. The photographs **(H)** and the chalky grain percentage **(I)**, chalky grain degree **(J)** are respectively shown. Statistical significance was determined by one-way ANOVA, where different letters indicate significant differences ($P <= 0.01$), while the same letter indicates no significant differences ($P > 0.05$).

mutants (Figure 3E). The heterozygous genotypes (*spl-A*/*lrd6-6*) of the F_1 progenies were confirmed by PCR-based sequencing and PCR-based agarose gel electrophoresis analysis respectively using specific primers (Figures 3F, G and Supplementary Table 1). Moreover, like the *lrd6-6* and *spl-A* mutants, the grain quality of all the F_1 progenies were found to deteriorate as indicated by increased chalky grain percentage and chalky grain degree (Figures 3H–J). These results suggest that *spl-A* and *lrd6-6* are allelic mutants and the C1144T mutation in *LOC_Os06g03940* is responsible for the autoimmunity and decreased grain quality of the *spl-A* mutant. In order to maintain the consistency of the gene name, *Lrd6-6* was adopted to represent the *Spl-A* gene hereafter.

The P382S mutation of LRD6-6 in *spl-A* impairs its ATPase activity and disrupts its interaction with ESCRT-III subunits OsSNF7.1/7.2/7.3

LRD6-6 is an active ATPase that can interact with itself and the ESCRT-III subunits OsSNF7.1/7.2/7.3 and OsVPS2 to regulate the MVBs-mediated vesicular trafficking pathway (Zhu et al., 2016). To determinate the effects of the P382S mutation in LRD6-6, we examined the ATPase activity of LRD6-6^{P382S}, the interaction of LRD6-6^{P382S} with itself and the ESCRT-III subunits OsSNF7.1/7.2/7.3 and OsVPS2, and the MVBs localization of LRD6-6^{P382S}.

We found that the P382S mutation in LRD6-6 impaired its ATPase activity because the His–LRD6-6(125–487)^{P382S} recombinant protein purified from *E. coli* showed significantly reduced ATPase activity when compared to the wild-type protein His–LRD6-6(125–487) (Figure 4). The self-interaction of LRD6-6 was not affected by the P382S mutation as detected by using the yeast two hybrid (Y2H) and bimolecular fluorescence complementation (BiFC) assays (Supplementary Figure 4). The interactions between LRD6-6^{P382S} and OsSNF7.1/7.2/7.3, LRD6-6^{P382S} and OsVPS2 were also examined using these two methods. And it was found that the interactions between LRD6-6^{P382S} and OsSNF7.1/7.2/7.3 were abolished while the interaction between LRD6-6^{P382S} and OsVPS2 was not affected (Figure 5). We also checked the MVBs localization of the protein LRD6-6^{P382S} by using RabF1/ARA6-RFP as the MVBs marker (Šamaj et al., 2005; Ebine et al., 2011). The co-localization of LRD6-6^{P382S}-GFP punctate green fluorescence with RabF1/ARA6-RFP punctate red fluorescence was observed in *Nicotiana benthamiana* (*N. benthamiana*) cells, similar to that of the wild-type protein LRD6-6-GFP (Supplementary Figure 5), suggesting the P382S mutation in LRD6-6 not affects its MVBs localization.

Taken together, these results indicating that the conserved amino acid residue proline at 382nd of LRD6-6 is essential for its ATPase activity and is also pivotal for interacting with the ESCRT-III subunits OsSNF7.1/7.2/7.3. These results further confirming that the C1144T mutation in *spl-A* which leads to P382S mutation of LRD6-6 is responsible for the autoimmunity, broad-spectrum disease resistance and decreased grain quality of the *spl-A* mutant.

The MVBs-mediated vesicle trafficking is compromised in the *spl-A* mutant

To dissect the mechanism on how LRD6-6^{P382S} affects rice growth and immune responses, we then collected leaf samples from A814 and *spl-A* respectively before the emergence of the spotted lesions on *spl-A* and analyzed the mRNA transcriptional level of genes associated with MVBs-mediated vesicle trafficking pathway firstly (Zhu et al., 2016). We found genes encoding possible MVBs-pathway components and MVBs-trafficking cargoes were largely dysregulated in the *spl-A* mutant when compared to the wild-type A814 (Figure 6A and Supplementary Table 3), suggesting that the MVBs-mediated vesicle trafficking pathway is defective in *spl-A*. Deficient in MVBs-mediated vesicle trafficking will lead to ROS

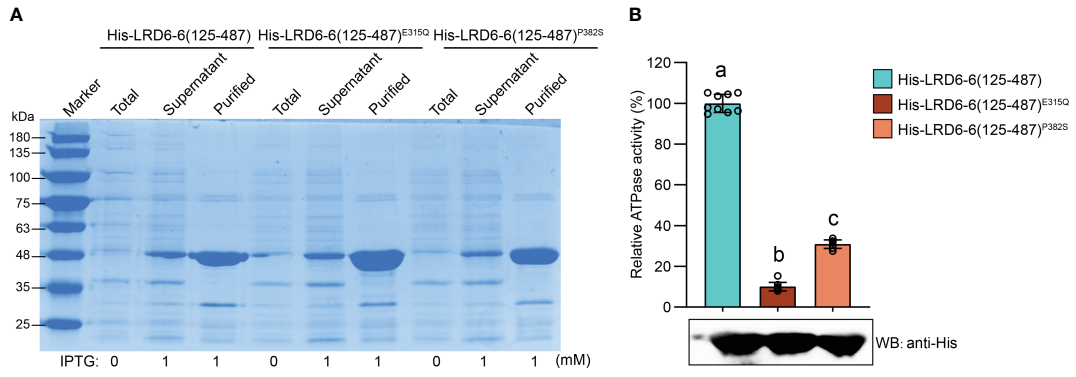


FIGURE 4
Determination of the ATPase activity on LRD6-6^{P382S} protein *in vitro*. **(A)** Expression and purification of the N-terminal truncated recombinant proteins His–LRD6-6(125–487), His–LRD6-6(125–487)^{E315Q} and His–LRD6-6(125–487)^{P382S} from *E. coli*. Coomassie brilliant blue staining was used to detect the total proteins from *E. coli* before IPTG addition, and proteins in supernatant and the purified proteins after IPTG induction. **(B)** *In vitro* ATPase assay on recombinant proteins His–LRD6-6(125–487), His–LRD6-6(125–487)^{E315Q} and His–LRD6-6(125–487)^{P382S}. The proteins His–LRD6-6(125–487) and His–LRD6-6(125–487)^{E315Q} were used as positive and negative control respectively. ATPase activities were measured using a malachite green-based colorimetric approach (mean ± s.d., n = 9 technical repetitions). The purified proteins used for ATPase activity determination were detected by anti-His antibody. Statistical significance was determined by one-way ANOVA, where different letters indicate significant differences ($P <= 0.01$).

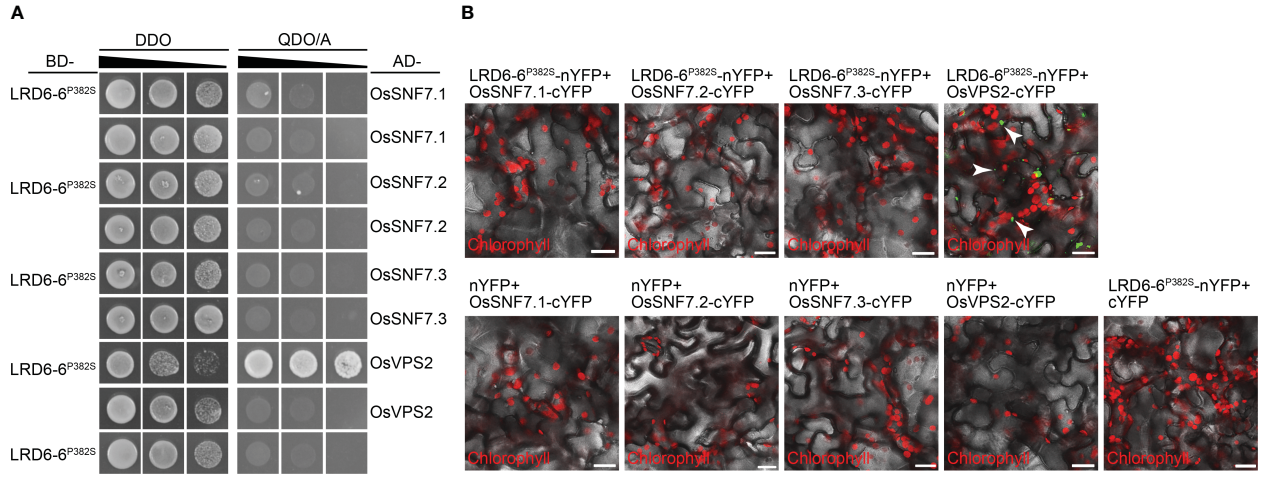


FIGURE 5
Detection the interaction between LRD6-6^{P382S} with OsSNF7 and OsVPS2. The interaction between LRD6-6^{P382S} with OsSNF7.1/7.2/7.3 and OsVPS2 were detected by using both yeast two hybrid (Y2H) **(A)** and bimolecular fluorescence complementation (BiFC) **(B)** approaches. BD, pGBKT7; AD, pGADT7; DDO, double dropout medium ((SD/−Leu/−Trp); QDO/A, quadruple dropout medium supplemented with Aureobasidin A (SD/−Ade/−His/−Leu/−Trp/AbA). BiFC assay was performed in *N. benthamiana*. The green fluorescence signals present the interaction between the tested proteins, the red signals represent the auto-fluorescence of chlorophyll. Some of the green fluorescence signals are indicated by white arrows to clearly shown the interaction. Bar = 20 μm.

burst and antimicrobial metabolites accumulation (Zhu et al., 2016). We thus also detected the mRNA expression levels of the genes associated with these processes (Zhu et al., 2016). In accordance with previous report, the expression of genes involved in ROS metabolism and biosynthesis of antimicrobial metabolites serotonin and diterpenoid phytoalexin were found to increase significantly in *spl-A* (Figure 6A and Supplementary Table 3), suggesting *spl-A* may accumulate excess of ROS and antimicrobial metabolites which can activate immune responses and increase disease resistance.

The MVBs are well-known for their function in sorting ubiquitinated plasma membrane proteins (Hurley, 2008; Gao et al., 2014). To further characterize the effects of LRD6-6 P382S mutation in *spl-A*, we then detected the ubiquitination level of total protein from the *spl-A* mutant leaf. More ubiquitinated proteins were detected in *spl-A* by using the anti-Ub (P4D1) antibody when compared to A814 (Figure 6B), indicating P382S mutation of LRD6-6 leading to accumulate of more ubiquitinated proteins in plant. Together, our current results suggest that the enhanced immune responses and deteriorated growth of the *spl-A* mutant is likely caused by the compromise of MVBs-mediated vesicle trafficking which leading to ROS burst and antimicrobial metabolites accumulation.

Since *spl-A* exhibited increased chalky grain percentage and chalky grain degree (Figures 1C–E), we speculated that the MVBs-mediated vesicle trafficking of ubiquitinated proteins in seeds were also inhibited. Thus, we also explored the ubiquitination level of total protein from *spl-A* seeds. It was found that the ubiquitinated proteins accumulated in *spl-A* seeds were obviously more than that of A814 (Figure 6C), suggesting the MVBs-mediated vesicle trafficking is also inhibited in seed. The contents of storage proteins in rice seeds are important for establishment of grain quality including chalkiness (Li et al., 2022). We then compared the contents of major types of rice seeds storage proteins, glutelins, α-globulin and prolamins between *spl-A* and A814. The SDS-PAGE based Coomassie blue staining showed that there were no obvious differences in the content of storage proteins between *spl-A* and A814 seeds (Figure 6D). These results suggest that the compromised MVBs-mediated vesicle trafficking in *spl-A* seed deteriorate grain chalkiness likely through a pathway independent of storage proteins contents.

Discussion

The *spl-A* mutant exhibits activated immune responses and decreased grain quality in rice

The mutation of the *Spl-A* locus has been previously reported as *lesion mimic resembling* (*lmr*), *lesion resembling disease 6-6* (*lrd6-6*) and *spotted leaf 4* (*spl4*) in rice (Fekih et al., 2015; Zhu et al., 2016; Song et al., 2019). Like most of *spl* mutants, the *lmr* and *lrd6-6* mutants accumulated an excess of H_2O_2 and death cells, exhibited enhanced disease resistance to both fungal and bacterial diseases (Lorrain et al., 2003; Fekih et al., 2015; Zhu et al., 2016, 2020). Consistently with previous reports, the *spl-A* mutant also presented constitutive activated immune responses and enhanced broad-spectrum disease resistance (Figure 2, Supplementary Figure 2 and Supplementary Figure 3). Besides immune responses, the *Spl-A* locus also regulates other aspects of plant phenotypes, such as leaf senescence and agronomic traits including grain weight, plant height and seed setting rate etc. It's interesting that the *lrd6-6* and *spl-A* mutants exhibited decreased grain weight, while the *spl4*

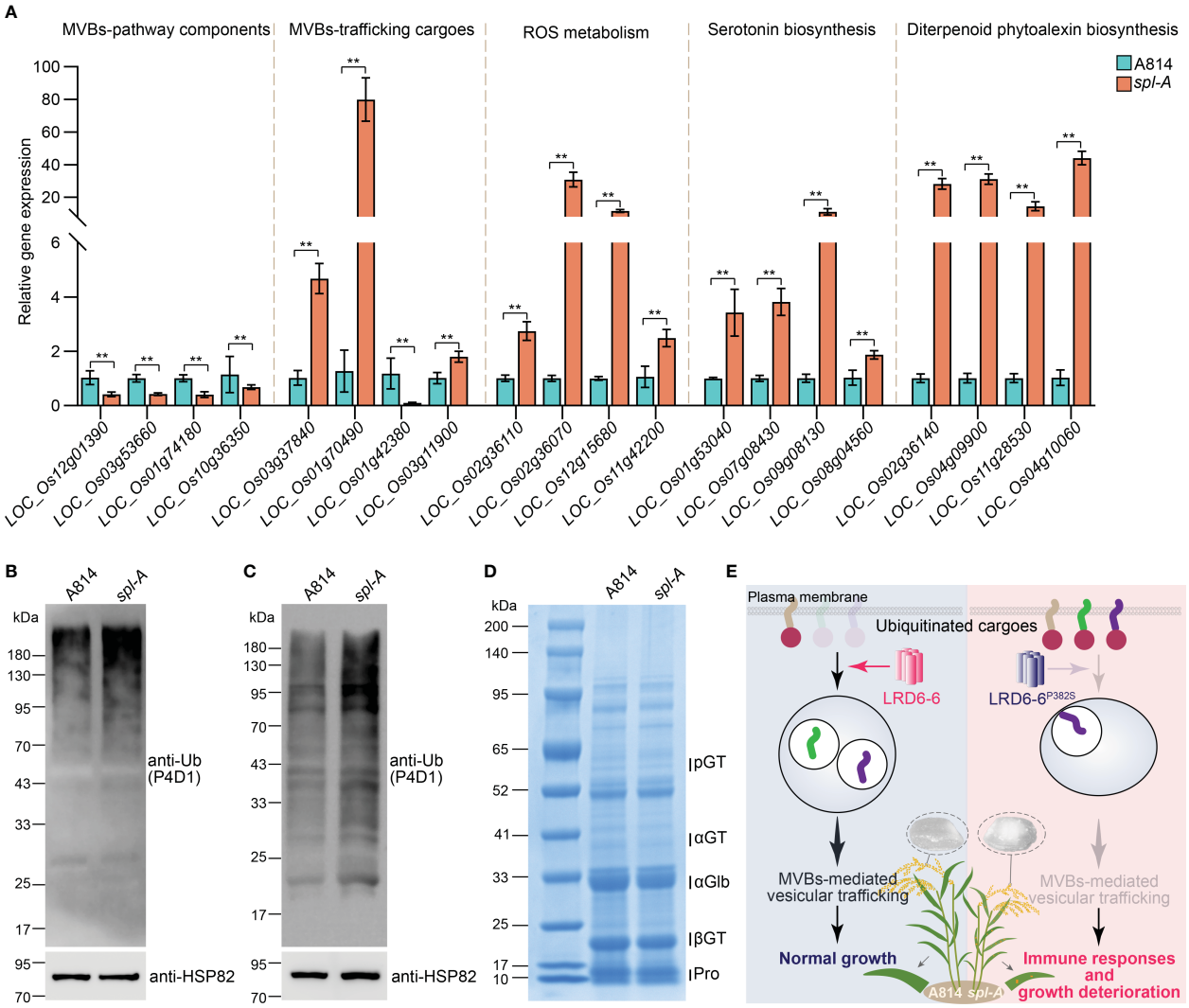


FIGURE 6
Characterization of the MVBs-mediated vesicle trafficking pathway in *spl-A*. **(A)** The mRNA transcriptional level of genes associated with MVBs-mediated vesicle trafficking pathway. The expression of genes that encoding possible MVBs-pathway components, possible MVBs-mediated trafficking cargoes, regulators involved in ROS metabolism, and key transcriptional factor and enzymes which regulate the biosynthesis of antimicrobial metabolites serotonin and diterpenoid phytoalexin was analyzed by RT-qPCR. The expression levels of the genes were normalized to the reference gene *Ubiquitin* and then normalized to the expression level of the genes in A814 (mean ± s.d., n = 3 technical repetitions). The detailed information of these genes was listed in Supplementary Table S2. All statistics were analyzed by Student's *t*-test for *P* values (**, *P* <= 0.01). **(B)** Detection of the ubiquitinated conjugates in A814 and *spl-A* leaves. **(C)** Detection of the ubiquitinated conjugates in A814 and *spl-A* dry seeds. Total protein from A814 and *spl-A* leaves **(C)** or dry seeds **(C)** were respectively extracted and subjected to immunoblot analysis using the anti-Ub antibody (P4D1). The protein levels of HSP82 were detected to indicate the loading amount of total protein in each lane. **(D)** Comparison the protein profiles between A814 and *spl-A* dry seeds on SDS–PAGE gels stained by Coomassie blue. 20 μg of total protein for each sample was loaded. blue. pGT, 57-kDa proglutelins; αGT, 40-kDa glutelin acidic subunits; αGlb, 26-kDa a-globulin; βGT, 20-kDa glutelin basic subunits; Pro, prolamins. **(E)** An illustration summarizes how *spl-A* mutant activates immune responses and deteriorates plant growth in rice. The AAA ATPase LRD6-6 regulates the maturation of MVBs and promotes MVBs-mediated transportation of ubiquitinated proteins to guarantee normal growth of rice plant (left panel). The P382S mutation of LRD6-6 in *spl-A* impairs its ATPase activity and disrupts its interaction with the ESCRT-III subunits OsSNF7.1/7.2/7.3, thus compromise MVBs-mediated transportation process and leading to immune responses activation and growth deterioration including decreased rice quality (right panel).

mutant showed increased grain weight (Fekih et al., 2015; Zhu et al., 2016; Song et al., 2019) (Supplementary Figure 1E). These results suggest that the *Spl-A* locus may regulate grain weight or even other traits depends on the genetic background of rice.

Although the *spl* mutants are broadly studied for their roles in immune responses, senescence, and plant growth or development (Zhu et al., 2020). It still remains unknown whether these mutants/genes regulate grain quality. In our study, we examined the chalkiness of the *spl-A* mutant and found that the *spl-A* mutant showed increased chalky grain percentage and chalky grain degree which deteriorates grain quality (Figures 1C–E and Figure 3H–J). Thus, our current study shed light on the new role of the *spl* mutants in grain quality regulation.

The conserved 382nd amino acid residue proline is essential for the full function of the AAA-type ATPase LRD6-6

The AAA-type ATPase family proteins contain conserved ATPase domains spanning 200–250 residues which cover the Walker A, Walker B and the SRH (Second Region of Homology) motifs that distinguish them from classic p-loop NTPases (Ogura and Wilkinson, 2001; Hanson and Whiteheart, 2005). VPS4 is a member of the AAA-type ATPase family which has been found to form hexamer and assemble with the ESCRT-III complex subunits to drive the maturation of MVBs (Babst et al., 1997; Schmidt and Teis, 2012). SKD1 and LRD6-6 are homologs of VPS4 from Arabidopsis and rice respectively, and they may function in MVBs maturation like that of VPS4 (Zhu et al., 2016). Several amino acid residues that essential for the function of VPS4/SKD1/LRD6-6 have been identified. For example, the 261st (LRD6-6 as reference) lysine in Walker A motif is essential for nucleotide binding, the 315th glutamic acid in Walker B is responsible for ATP hydrolysis, and the 372nd arginine in SRH is vital for both ATP hydrolysis and oligomerization (Babst et al., 1998; Ogura et al., 2004; Zhu et al., 2016). The *spl-A* mutant in this study was caused by a single nucleotide mutation (C1144T) which led to a single amino acid substitution (P382S) in the AAA-type ATPase LRD6-6 (Figures 3A–C). Unlike the mutations in *lmr*, *lrd6-6* and *spl4* whose genes were destructed (Fekih et al., 2015; Zhu et al., 2016; Song et al., 2019), the mutation of *spl-A* provides us a unique allele to study the function of the 382nd amino acid residue proline for LRD6-6.

The 382nd amino acid residue proline locates in neither of the above-mentioned motifs of LRD6-6, but it's conserved among this AAA-type ATPase family (Figure 3D). We then assayed the ATPase activity, the interaction with the ESCRT-III subunits OsSNF7.1/7.2/7.3 and OsVPS2, and the MVBs localization of the mutation protein LRD6-6^{P382S} from the *spl-A* mutant. Our results indicate that the P382S mutation of LRD6-6 impairs its ATPase activity and disrupts the interaction with the ESCRT-III subunits OsSNF7.1/7.2/7.3, while not affects its MVBs localization (Figures 4, 5 and Supplementary Figures 4, 5). Thus, in this study, we have characterized a new conserved amino acid residue that is important for the full function of LRD6-6 and likely, for all the AAA-type ATPase family proteins since this site is conserved among them.

LRD6-6 regulates MVBs-mediated vesicle trafficking to modulate rice immunity and grain quality

MVBs are single membrane bound organelles that serve as core converging station of the secretory and endocytic pathways for cargo trafficking in eukaryotic cells (Piper and Katzmann, 2007; Cui et al., 2016). Generally, ubiquitinated cargo proteins on the plasm membrane will be recognized and packaged into the intralumenal vesicles (ILVs) of MVBs by the ESCRT complexes. Then, these cargo proteins are transported by the MVBs to different destinations for proper functions (Piper and Katzmann, 2007; Cui et al., 2016; Zhu et al., 2023). Impairment of the MVBs-mediated vesicle trafficking pathway will lead to accumulate of ubiquitinated proteins in cells as reported in Yeast and Arabidopsis (Mizuno et al., 2003; Stringer and Piper, 2011; Gao et al., 2014; Katsiarimpa et al., 2014). However, whether rice cells will accumulate ubiquitinated proteins when this pathway is inhibited still remain unknown. In this study, we found ubiquitinated proteins accumulated in both the leaves and seeds of the *spl-A* mutant in which the MVBs-mediated vesicle trafficking was inhibited by mutation of the AAA-type ATPase LRD6-6 (Figures 6B, C). Thus, our discovery not only reveals the conserved roles of MVBs-mediated vesicle trafficking on the transportation of ubiquitinated proteins, but also expands its function in grain quality formation.

In summary, our study reveals that the conserved amino acid residue proline at the 382th of the AAA-type ATPase LRD6-6 is essential for its ATPase activity and its interaction with the ESCRT-III subunits OsSNF7.1/7.2/7.3. Once this site is mutated, the MVBs-mediated vesicle trafficking pathway governed by LRD6-6 is compromised and rice plant harboring this mutation will accumulate excessive ubiquitinated cargo proteins in both leaves and seeds. Then, the immune responses of this mutant plant will be activated, and the growth and grain quality of this plant will be deteriorated (Figure 6E).

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

JY: Funding acquisition, Methodology, Resources, Writing – original draft, Data curation, Investigation, Validation. CZ: Investigation, Methodology, Validation, Writing – review & editing. QZ: Investigation, Methodology, Validation, Writing – review & editing. FL: Investigation, Methodology, Validation, Writing – review & editing. WH: Investigation, Formal analysis, Writing – review & editing. YiZ: Formal analysis, Investigation, Writing – review & editing. FM: Formal analysis, Investigation, Writing – review & editing. YuZ: Investigation, Writing – review & editing. BW: Investigation, Writing – review & editing. MZ: Investigation, Writing – review & editing. LZ: Formal analysis, Funding acquisition, Writing – review & editing. XZ: Formal analysis, Funding acquisition, Conceptualization, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the National Key Research and Development Program of China for young scientists (2022YFD1401400), the National Natural Science Foundation of China (32072407, 32072041 and 32272560), and the Sichuan Science and Technology Program (2024NSFSC0326, 2023NSFSC1937 and 2022NSFSC0156).

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1451897/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Comparison on the main agronomic traits between and *spl-A*. The plant height **(A)**, tiller number **(B)**, panicle length **(C)**, seed setting rate **(D)** and 1,000-grain weight **(E)** of the *spl-A* mutant were respectively investigated and analyzed (mean ± s.d.). **(F–H)** Comparison on grain length **(G)** and grain width **(H)** between A814 and *spl-A* (mean ± s.d.). Representative photograph of ten grains were respectively shown. Bar = 1 cm. All statistics were analyzed by Student's *t*-test for *P* values (**, *P* <= 0.01; ns, no significant differences).

SUPPLEMENTARY FIGURE 2
Evaluation the blast resistance of *spl-A* to different blast fungal isolates. Punch inoculation method was employed. The blast fungal isolates ZE-1 **(A–C)**, 0755-1-1 **(D–F)** and 99-20-2 **(G–I)** that are compatible with A814 was respectively used. Photograph of representative lesions **(A, D, G)** were shown. Lesion length **(B, E, H)** and the relative fungal growth **(C, F, I)** were measured at 5 day-post-inoculation (dpi). The relative fungal growth was determinated as fungi *MoPot2* DNA to rice *OsUbq* DNA by qPCR (mean ± s.d., n = 3 technical repetitions). Bars = 1 cm. All statistics were analyzed by Student's *t*-test for *P* values (**, *P* <= 0.01).

SUPPLEMENTARY FIGURE 3
Evaluation the bacterial blight disease resistance of *spl-A* to different *Xoo* isolates. The resistance of *spl-A* to *Xoo* isolates P6 **(A–C)** and Xoo-4 **(D–F)** were tested. Photographs of representative leaves were taken at 14 dpi **(A, D)**. Disease lesion lengths **(B, E)** and bacterial populations **(C, F)** of A814 and the *spl-A* mutant were measured at 0, 7 and 14 dpi respectively (mean ± s.d., n = 12 for lesion lengths and n = 3 for bacterial populations). Bars = 1 cm. All statistics were analyzed by Student's *t*-test for *P* values (**, *P* <= 0.01).

SUPPLEMENTARY FIGURE 4
Detection the self-interaction of LRD6-6^{P382S} protein. The self-interaction of LRD6-6^{P382S} was detected by using both Y2H **(A)** and BiFC **(B)** approaches. BD, pGBKT7; AD, pGADT7; DDO, double dropout medium ((SD/−Leu/−Trp); QDO/A, quadruple dropout medium supplemented with Aureobasidin A (SD/−Ade/−His/−Leu/−Trp/AbA). BiFC assay was performed in *N. benthamiana*. The green fluorescence signals present the interaction between the tested proteins, the red signals represent the auto-fluorescence of chlorophyll. Some of the green fluorescence signals are indicated by white arrows to clearly shown the interaction. Bar = 20 μm.

SUPPLEMENTARY FIGURE 5
Subcellular localization of the protein LRD6-6^{P382S}. **(A)** Determination the subcellular localization of protein LRD6-6^{P382S} fused with green fluorescence protein (LRD6-6^{P382S}-GFP) in *N. benthamiana*. The RabF1/ARA6 protein that fused with red fluorescence protein (RabF1/ARA6-RFP) was used as the multivesicular bodies (MVBs) marker. Some of the overlapped punctate fluorescence signals were indicated by white arrows to clearly shown the co-localization. The fluorescence intensity along the white line was also measured and shown to indicate the co-localization **(B)**. The co-localization of the wild-type LRD6-6-GFP with RabF1/ARA6-RFP was included as control **(C, D)**. Bar = 20 μm.

References

Aniento, F., Sánchez De Medina Hernández, V., Dagdas, Y., Rojas-Pierce, M., and Russinova, E. (2021). Molecular mechanisms of endomembrane trafficking in plants. *Plant Cell* 34, 146–173. doi: 10.1093/plcell/koab235

Babst, M., Sato, T. K., Banta, L. M., and Emr, S. D. (1997). Endosomal transport function in yeast requires a novel AAA-type ATPase, Vps4p. *EMBO J.* 16, 1820–1831. doi: 10.1093/emboj/16.8.1820

Babst, M., Wendland, B., Estepa, E. J., and Emr, S. D. (1998). The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J.* 17, 2982–2993. doi: 10.1093/emboj/17.11.2982

Balint-Kurti, P. (2019). The plant hypersensitive response: concepts, control and consequences. *Mol. Plant Pathol.* 20, 1163–1178. doi: 10.1111/mpp.12821

Bruggeman, Q., Raynaud, C., Benhamed, M., and Delarue, M. (2015). To die or not to die? Lessons from lesion mimic mutants. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00024

Caddell, D. F., Park, C.-J., Thomas, N. C., Canlas, P. E., and Ronald, P. C. (2017). Silencing of the rice gene LRR1 compromises rice Xa21 transcript accumulation and XA21-mediated immunity. *Rice* 10, 23. doi: 10.1186/s12284-017-0162-5

Cui, Y., Shen, J., Gao, C., Zhuang, X., Wang, J., and Jiang, L. (2016). Biogenesis of plant prevacuolar multivesicular bodies. *Mol. Plant* 9, 774–786. doi: 10.1016/j.molp.2016.01.011

Ebine, K., Fujimoto, M., Okatani, Y., Nishiyama, T., Goh, T., Ito, E., et al. (2011). A membrane trafficking pathway regulated by the plant-specific RAB GTPase ARA6. *Nat. Cell Biol.* 13, 853–859. doi: 10.1038/ncb2270

Fekih, R., Tamiru, M., Kanzaki, H., Abe, A., Yoshida, K., Kanzaki, E., et al. (2015). The rice (Oryza sativa L.) LESION MIMIC RESEMBLING, which encodes an AAA-type ATPase, is implicated in defense response. *Mol. Genet. Genomics* 290, 611–622. doi: 10.1007/s00438-014-0944-z

Gao, C., Luo, M., Zhao, Q., Yang, R., Cui, Y., Zeng, Y., et al. (2014). A unique plant ESCRT component, FREE1, regulates multivesicular body protein sorting and plant growth. *Curr. Biol.* 24, 2556–2563. doi: 10.1016/j.cub.2014.09.014

Hanson, P. I., and Whiteheart, S. W. (2005). AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.* 6, 519–529. doi: 10.1038/nrm1684

Hurley, J. H. (2008). ESCRT complexes and the biogenesis of multivesicular bodies. *Curr. Opin. Cell Biol.* 20, 4–11. doi: 10.1016/j.ceb.2007.12.002

Jiang, M., Yu, N., Zhang, Y., Liu, L., Li, Z., Wang, C., et al. (2022). Deletion of diterpenoid biosynthetic genes CYP76M7 and CYP76M8 induces cell death and enhances bacterial blight resistance in indica rice '9311'. *Int. J. Mol. Sci.* 23, 7234. doi: 10.3390/ijms23137234

Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

Katsiarimpa, A., Muñoz, A., Kalinowska, K., Uemura, T., Rojo, E., and Isono, E. (2014). The ESCRT-III-interacting deubiquitinating enzyme AMSH3 is essential for degradation of ubiquitinated membrane proteins in *Arabidopsis thaliana*. *Plant Cell Physiol.* 55, 727–736. doi: 10.1093/pcp/pcu019

Li, P., Chen, Y.-H., Lu, J., Zhang, C.-Q., Liu, Q.-Q., and Li, Q.-F. (2022). Genes and their molecular functions determining seed structure, components, and quality of rice. *Rice* 15, 18. doi: 10.1186/s12284-022-00562-8

Li, W., Wang, K., Chern, M., Liu, Y., Zhu, Z., Liu, J., et al. (2020). Sclerenchyma cell thickening through enhanced lignification induced by OsMYB30 prevents fungal penetration of rice leaves. *New Phytol.* 226, 1850–1863. doi: 10.1111/nph.16505

Li, Z., Ding, B., Zhou, X., and Wang, G.-L. (2017). The rice dynamin-related protein OsDRP1E negatively regulates programmed cell death by controlling the release of cytochrome c from mitochondria. *PloS Pathog.* 13, e1006157. doi: 10.1371/journal.ppat.1006157

Liu, L., Zhang, Y., Tang, S., Zhao, Q., Zhang, Z., Zhang, H., et al. (2010). An efficient system to detect protein ubiquitination by agroinfiltration in Nicotiana benthamiana. *Plant J.* 61, 893–903. doi: 10.1111/tpj.2010.61.issue-5

Lorrain, S., Vailleau, F., Balagué, C., and Roby, D. (2003). Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? *Trends Plant Sci.* 8, 263–271. doi: 10.1016/S1360-1385(03)00108-0

Ma, J., Chen, J., Wang, M., Ren, Y., Wang, S., Lei, C., et al. (2017). Disruption of OsSEC3A increases the content of salicylic acid and induces plant defense responses in rice. *J. Exp. Bot.* 69, 1051–1064. doi: 10.1093/jxb/erx458

Ma, J., Wang, Y., Ma, X., Meng, L., Jing, R., Wang, F., et al. (2019). Disruption of gene SPL35, encoding a novel CUE domain-containing protein, leads to cell death and enhanced disease response in rice. *Plant Biotechnol. J.* 17, 1679–1693. doi: 10.1111/pbi.13093

Mizuno, E., Kawahata, K., Kato, M., Kitamura, N., and Komada, M. (2003). STAM proteins bind ubiquitinated proteins on the early endosome via the VHS domain and ubiquitin-interacting motif. *Mol. Biol. Cell* 14, 3675–3689. doi: 10.1091/mbc.e02-12-0823

Mur, L., Kenton, P., Lloyd, A. J., Ougham, H., and Prats, E. (2008). The hypersensitive response; the centenary is upon us but how much do we know? *J. Exp. Bot.* 59, 501–520. doi: 10.1093/jxb/erm239

Ogura, T., Whiteheart, S. W., and Wilkinson, A. J. (2004). Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J. Struct. Biol.* 146, 106–112. doi: 10.1016/j.jsb.2003.11.008

Ogura, T., and Wilkinson, A. J. (2001). AAA+ superfamily ATPases: common structure–diverse function. *Genes to Cells* 6, 575–597. doi: 10.1046/j.1365-2443.2001.00447.x

Pan, T., Wang, Y., Jing, R., Wang, Y., Wei, Z., Zhang, B., et al. (2021). Post-Golgi trafficking of rice storage proteins requires the small GTPase Rab7 activation complex MON1–CCZ1. *Plant Physiol.* 187, 2174–2191. doi: 10.1093/plphys/kiab175

Piper, R. C., and Katzmann, D. J. (2007). Biogenesis and function of multivesicular bodies. *Annu. Rev. Cell Dev. Biol.* 23, 519–547. doi: 10.1146/annurev.cellbio.23.090506.123319

Qiao, Y., Jiang, W., Lee, J., Park, B., Choi, M.-S., Piao, R., et al. (2010). SPL28 encodes a clathrin-associated adaptor protein complex 1, medium subunit μ1 (AP1M1) and is responsible for spotted leaf and early senescence in rice (Oryza sativa). *New Phytol.* 185, 258–274. doi: 10.1111/j.1469-8137.2009.03047.x

Ren, Y., Wang, Y., Pan, T., Wang, Y., Wang, Y., Gan, L., et al. (2020). GPA5 encodes a Rab5a effector required for post-Golgi trafficking of rice storage proteins. *Plant Cell* 32, 758–777. doi: 10.1105/tpc.19.00863

Ren, D., Xie, W., Xu, Q., Hu, J., Zhu, L., Zhang, G., et al. (2022). LSL1 controls cell death and grain production by stabilizing chloroplast in rice. *Sci. China Life Sci.* 65, 2148–2161. doi: 10.1007/s11427-022-2152-6

Šamaj, J., Read, N. D., Volkmann, D., Menzel, D., and Baluška, F. (2005). The endocytic network in plants. *Trends Cell Biol.* 15, 425–433. doi: 10.1016/j.tcb.2005.06.006

Schmidt, O., and Teis, D. (2012). The ESCRT machinery. *Curr. Biol.* 22, R116–R120. doi: 10.1016/j.cub.2012.01.028

Song, G., Kwon, C.-T., Kim, S.-H., Shim, Y., Lim, C., Koh, H.-J., et al. (2019). The Rice SPOTTED LEAF4 (SPL4) encodes a plant spastin that inhibits ROS accumulation in leaf development and functions in leaf senescence. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01925

Stringer, D. K., and Piper, R. C. (2011). A single ubiquitin is sufficient for cargo protein entry into MVBs in the absence of ESCRT ubiquitination. *J. Cell Biol.* 192, 229–242. doi: 10.1083/jcb.201008121

Tao, H., Shi, X., He, F., Wang, D., Xiao, N., Fang, H., et al. (2021). Engineering broad-spectrum disease-resistant rice by editing multiple susceptibility genes. *J. Integr. Plant Biol.* 63, 1639–1648. doi: 10.1111/jipb.13145

Tsuda, K., and Katagiri, F. (2010). Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. *Curr. Opin. Plant Biol.* 13, 459–465. doi: 10.1016/j.pbi.2010.04.006

Wang, Y., Teng, Z., Li, H., Wang, W., Xu, F., Sun, K., et al. (2023). An activated form of NB-ARC protein RLS1 functions with cysteine-rich receptor-like protein RMC to trigger cell death in rice. *Plant Commun.* 4, 100459. doi: 10.1016/j.xplc.2022.100459

Yao, D., Wu, J., Luo, Q., Li, J., Zhuang, W., Xiao, G., et al. (2020). Influence of high natural field temperature during grain filling stage on the morphological structure and physicochemical properties of rice (Oryza sativa L.) starch. *Food Chem.* 310, 125817. doi: 10.1016/j.foodchem.2019.125817

Yao, Y., Zhou, J., Cheng, C., Niu, F., Zhang, A., Sun, B., et al. (2022). A conserved clathrin-coated vesicle component, OsSCYL2, regulates plant innate immunity in rice. *Plant Cell Environ.* 45, 542–555. doi: 10.1111/pce.14240

Yuan, M., Ngou, B. P. M., Ding, P., and Xin, X.-F. (2021). PTI-ETI crosstalk: an integrative view of plant immunity. *Curr. Opin. Plant Biol.* 62, 102030. doi: 10.1016/j.pbi.2021.102030

Zhang, X.-Q., Hou, P., Zhu, H.-T., Li, G.-D., Liu, X.-G., and Xie, X.-M. (2013). Knockout of the VPS22 component of the ESCRT-II complex in rice (*Oryza sativa L.*) causes chalky endosperm and early seedling lethality. *Mol. Biol. Rep.* 40, 3475–3481. doi: 10.1007/s11033-012-2422-1

Zhao, Y., Zhu, X., Chen, X., and Zhou, J.-M. (2022). From plant immunity to crop disease resistance. *J. Genet. Genomics* 49, 693–703. doi: 10.1016/j.jgg.2022.06.003

Zheng, Y., Zhu, Y., Mao, X., Jiang, M., Wei, Y., Lian, L., et al. (2022). SDR7-6, a short-chain alcohol dehydrogenase/reductase family protein, regulates light-dependent cell death and defence responses in rice. *Mol. Plant Pathol.* 23, 78–91. doi: 10.1111/mpp.13144

Zhu, X., Yin, J., Guo, H., Wang, Y., and Ma, B. (2023). Vesicle trafficking in rice: too little is known. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1263966

Zhu, X., Yin, J., Liang, S., Liang, R., Zhou, X., Chen, Z., et al. (2016). The multivesicular bodies (MVBs)-localized AAA ATPase LRD6-6 inhibits immunity and cell death likely through regulating MVBs-mediated vesicular trafficking in rice. *PloS Genet.* 12, e1006311. doi: 10.1371/journal.pgen.1006311

Zhu, X., Ze, M., Chern, M., Chen, X., and Wang, J. (2020). Deciphering rice lesion mimic mutants to understand molecular network governing plant immunity and growth. *Rice Sci.* 27, 278–288. doi: 10.1016/j.rsci.2020.05.004

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 August 2024
DOI 10.3389/fpls.2024.1452520

OPEN ACCESS

EDITED BY
Yuqing He,
Huazhong Agricultural University, China

REVIEWED BY
Xiangjin Wei,
China National Rice Research Institute (CAAS), China
Yong Liang,
Chongqing Normal University, China

*CORRESPONDENCE
Xianjun Wu
wuxjsau@126.com

†These authors have contributed equally to this work

RECEIVED 21 June 2024
ACCEPTED 22 July 2024
PUBLISHED 14 August 2024

CITATION
Chen X, Guo Q, Yang X, Yuan M, Song J, Fu H, Zhang H, Xu P, Liao Y, Ali A, Du K and Wu X (2024) Triple gene mutations boost amylose and resistant starch content in rice: insights from *sbe2b*/*sbe1*/OE-*Wxa* mutants. *Front. Plant Sci.* 15:1452520. doi: 10.3389/fpls.2024.1452520

Triple gene mutations boost amylose and resistant starch content in rice: insights from *sbe2b*/*sbe1*/OE-*Wxa* mutants

Xiaoqiong Chen†, Qiaoling Guo†, Xiaoli Yang†, Meng Yuan, Jianguo Song, Hongyan Fu, Hongyu Zhang, Peizhou Xu, Yongxiang Liao, Asif Ali, Kangxi Du and Xianjun Wu*

State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Rice Research Institute, Sichuan Agricultural University, Chengdu, China

Previous studies have modified rice's resistant starch (RS) content by mutating single and double genes. These mutations include knocking out or reducing the expression of *sbe1* or *sbe2b* genes, as well as overexpressing Wx^a. However, the impact of triple mutant *sbe2b*/*sbe1*/OE-Wx^a on RS contents remained unknown. Here, we constructed a double mutant with *sbe2b*/RNAi-*sbe1*, based on *IR36ae* with *sbe2b*, and a triple mutant with *sbe2b*/RNAi-*sbe1*/OE-Wx^a, based on the double mutant. The results showed that the amylose and RS contents gradually increased with an increase in the number of mutated genes. The triple mutant exhibited the highest amylose and RS contents, with 41.92% and 4.63%, respectively, which were 2- and 5-fold higher than those of the wild type, which had 22.19% and 0.86%, respectively. All three mutants altered chain length and starch composition compared to the wild type. However, there was minimal difference observed among the mutants. The Wx^a gene contributed to the improvement of 1000-grain weight and seed-setting rate, in addition to the highest amylose and RS contents. Thus, our study offers valuable insight for breeding rice cultivars with a higher RS content and yields.

KEYWORDS

resistant starch, amylose, OE-Wx a, triple mutant, agronomic trait

Introduction

Rice (*Oryza sativa* L.) is a staple food for a large proportion of the global population. Starch, the main component of rice grains, provides up to 40% of the caloric intake for those who rely on rice as a staple food (Awika et al., 2011). Most starch in cereal products is digested rapidly in the upper gastrointestinal tract; however, resistant starch (RS) is unaffected by enzymatic hydrolysis and hence escapes degradation in the stomach and small intestine (Englyst et al., 1982; Yang et al., 2016).

RS diets have numerous health benefits, such as reducing pathogenic infections, colon cancer rates, postprandial glycemia, and low-density lipoprotein (LDL) serum cholesterol levels (Zhang et al., 2015; Marlatt et al., 2018; Medina-Remón et al., 2018). A high-RS diet is particularly effective in reducing the risk of developing type-2 diabetes (Topping et al., 2003), a condition that is increasingly prevalent worldwide and affects many young individuals. Consequently, several studies have focused on enhancing RS content through genetic breeding, biotechnology techniques, and physical and chemical methods (Bird et al., 2008; Jiang et al., 2010; Ambigaipalan et al., 2014).

The RS content in rice is influenced by various factors, including amylose content, the ratio of amylose to amylopectin, and molecular characteristics of starch such as chain structure, granule size, and amorphous content. These characteristics are determined by genetic factors, such as genes involved in the starch biosynthesis pathway and transcription factors (Guzman et al., 2017; Bemiller, 2019). Starch biosynthesis in rice is a complex process involving different enzymes namely, Granule-bound starch synthase I (GBSSI), starch synthase (SS), disbranching enzyme (DBE), and starch branching enzyme (SBE). These enzymes play a role in the production of RS by affecting amylose content, chain length, granule size, gel consistency, and gelatinization temperature (Hennen-Bierwagen et al., 2008). GBSSI plays a crucial role in the amylose formation in the rice endosperm and is encoded by the *waxy* allele (*Wx*) gene on chromosome 6 (Liu et al., 2009). To date, nine *Wx* alleles have been identified: Wx^a, Wx^b, Wx^{in}, Wx^{lv}, Wx^{op}, Wx^{mq}, Wx^{mp}, Wx^{la}, and *Wx* (Zhou et al., 2021). Numerous studies have demonstrated that Wx^a leads to high amylose content in *indica* rice varieties (Ayres et al., 1997; Wanchana et al., 2003; Mikami et al., 2008; Liu et al., 2009; Yang et al., 2013; Zhang et al., 2019). SSs catalyze the formation of linear α-1,4-linked glucan chains while DBEs and pullulanase are associated with semicrystalline amylopectin (Wattebled et al., 2005; Fujita et al., 2009). SBEs facilitate chain transfer by cleaving an α-1,4 linkage following the α-1,6 linkage condensation (Tanaka et al., 2004). Rice possesses three SBE isoforms: SBEI, SBEIIa, and SBEIIb (Zhu et al., 2012; Yang et al., 2016).

Considerable studies have shown that RS content is tightly related to amylose content, both of which are determined by genes involved in the starch biosynthesis pathway. It has been found that the downregulation of the *SBEII* gene can significantly increase amylose and RS content (Wei et al., 2010; Butardo et al., 2011; Zhao et al., 2015). Additionally, in *indica* rice, higher expression of Wx^a has been shown to significantly increase amylose and RS contents (Zhou et al., 2016). Thus, these studies have focused on improving amylose and RS content by manipulating the expression of genes related to the starch biosynthesis pathway, for example, single mutants have been created to increase RS content by either knocking out or reducing *ss3a* or *sbe2b*, or by overexpressing *GBSSI*. Double mutants have also been utilized such as those with *ss3a* and *sbe2b*, *sbe1* and *sbe2b*, as well as overexpression of Wx^a combined with knockout of *OsSBEIIb* (Shim et al., 2022). However, the effects of triple mutants involving *sbe2b*/RNAi-*sbe1*/OE-Wx^a on amylose and RS contents remain unknown.

Our study aimed to determine the impact of mutations in three specific genes on RS formation to provide useful information for breeding cultivars with higher RS content. To achieve this, we created double and triple mutants, by knockdown of *SBEI* (Os06g0726400) expression in *IR36ae* with the *sbe2b* mutant, and overexpression of Wx^a (Os06g0133000) in *IR36ae* and RNAi *SBEI*, respectively. Subsequently, we accessed the agronomic traits, measured amylose and RS contents, analyzed starch molecular characteristics, and examined the expression of genes involved in the starch biosynthesis pathway.

Materials and methods

Wild-type IR36 was selected from the resources available at Sichuan Agricultural University, China. *IR36ae* (single mutant) was obtained from a mutant library of Sichuan Agricultural University. This mutant was induced using Ethyl methanesulfonate (EMS) as a mutagen, located at the splice site between exon and intron 11. This resulted in a change from G to A, causing a premature stop codon and producing a truncated form of BEIIb (Os02g0528200).

Plasmid construction and transformation

To prepare the double mutant (*sbe2b*/RNAi-*sbe1*), we followed the same method as described by (Sallaud et al., 2003) to knock down the expression of *OsBEI* (Os06g0726400) based on *IR36ae* with the mutation of *sbe2b*. A 320-bp cDNA fragment belonging to a region encoding sequence in rice cultivar IR36 cDNA was amplified. The fragments were inserted on both sides of the intron in opposite orientations into the binary vector pTCK303. Afterward, it was cloned into the binary vector pCambia1300. To prepare the triple mutant (*sbe2b*/RNAi-*sbe1*/*OE-*Wx^a), we overexpressed Wx^a based on the double mutant according to the procedure described by (Luo et al., 2022). To summarize, we first amplified a full-length 1833-bp fragment from the cDNA sequence of rice cultivar IR36, next we cloned it into the binary vector pCambia1300. The transgenic plants were generated using Agrobacterium-mediated transformation and confirmed using gene-specific primers (Supplementary Table 1). These transgenic plants were observed for three consecutive years.

DNA extraction

Fresh leaves were collected from each plant to extract genomic DNA using the CTAB method. PCR was performed by following these steps: initial denaturation at 95°C for 5 minutes, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 57°C for 30 seconds, extension at 72°C for 30 seconds, and a final extension at 72°C for 5 minutes.

Real-time qPCR analysis

The spikelets were harvested on the 5th and 10th day of flowering. All spikelets samples were immediately frozen in liquid

nitrogen and stored at -80°C until use. Total RNA was extracted from dehulled seeds using an RNA extraction reagent (Vazyme Co., Ltd, Nanjing, China) following the manufacturer's protocol. The extracts were treated with DNase I (Ambion). For cDNA synthesis, 2 μg of total RNA from caryopses was used with SuperScript III (Invitrogen). qPCR experiments were performed using gene-specific primers (Supplementary Table 2) in the SsoFast EvaGreen Supermix reaction system (Bio-Rad) on a CFX96 Real-Time System (Bio-Rad Co., Ltd, Singapore), according to the manufacturer's instructions. The rice actin gene (LOC_Os03g50885) was used as an internal reference.

Sample preparation for amylose and RS contents and starch molecular characteristics

Wild type, single mutants (*IR36ae* with *sbe2b*), double mutants (*sbe2b*/RNAi-*sbe1*), and triple mutants (*sbe2b*/RNAi-*sbe1*/*OE-Wx*^{a}) were planted in Wenjiang, Chengdu, China. Mature seeds were randomly selected from the wild-type and single, double, and triple mutants, with each group, replicated three times. All samples underwent milling refinement using SINOGRAIN, China Grain and Oil Reserve Co., Ltd. in Suzhou, China. The milled rice powder was then mixed and sifted through a 100-mesh sieve. These milled flours were used for further experiments.

Determination of amylose and RS contents

The amylose contents of the milled rice flour samples were measured according to the method described by Tao et al. (2019). The RS content was examined using a Resistant Starch Assay Kit (Megazyme Co., Ltd. Bray, Ireland) and following the manufacturer's protocol which is based on the method described by Parween et al. (2020).

Rapid visco analysis

The pasting properties of the milled rice were investigated using a rapid viscosity analyzer (Brabender Micro Visco-Amylo-Graph Co., Ltd. Brabender, Germany) and analyzed using Thermal Cycle for Windows software. Details of this procedure have been described (Zhang et al., 2013).

Starch isolation

Starch samples from each treatment group were isolated following the method described by Lu and Lu (2012), with minor modifications. Briefly, samples (20 g) were soaked in 100 mL of ultrapure water containing sodium metabisulfite and 10 mg/g alkaline protease at 42°C for 24 h. The samples were mixed in a blender and sifted through a 200-mesh sieve. The resulting slurry was collected and allowed to stand for 12 hours. This step was repeated 5–8 times until the settled starch layer was purified. Starch samples were collected and air-dried. Samples were stored at 20°C until further use. Starch isolation was employed to analyze various properties including swelling power, water solubility, chain length, molecular weight, branching degree, granularity distribution, and granule morphology.

Swelling power and water solubility determination of starch

The swelling power and water solubility of starch were measured at 95°C for 20 min following the method described by De la hera et al. (2013).

Determination of starch granule morphology

To examine the endosperm cross-section, rice seeds were fractured along the short axis and fully dried under low pressure. The surfaces were sputter-coated with gold and observed under a scanning electron microscope (Zeiss, Oberkochen, Germany).

The starch granule morphology was observed following the protocol described by Zou et al. (2020). Samples (100 mg) were mounted on a metal stub, covered with gold, and then observed and photographed using a Zeiss Merlin Compact scanning electron microscope (SEM, Zeiss, Oberkochen, Germany).

Determination of granule size

To a 100 mg sample in EP tubes, 1 mL of 75% alcohol was added and mixed using ultrasound. Samples were analyzed using a Mastersizer 3000 laser diffraction particle size analyzer (Malvern Instruments Ltd., Worcestershire, UK). All samples were measured in triplicate.

Starch molecular weight determination

Starch (5 mg) was thoroughly mixed with 5 mL of 0.5% LiBr (w/w) in DMSO solution and heated in a thermomixer at 80°C for 3 h, as described by Zou et al. (2020). The molecular weight of the starch was determined using the gel permeation chromatography-refractive index-multiangle laser light scattering (GPC–RI–MALLS) method with an Optilab T-rEX differential refractive index detector (Wyatt Technology) and a DAWN HELEOSII (Wyatt Technology). Data were analyzed using ASTRA v6.1 (Wyatt, Santa Barbara, CA, USA). Three biological replicates were used for each sample.

Determination of branch chain length distribution

The branch chain length distribution of amylopectin was measured using an ICS-5000 high-performance anion-exchange

chromatograph (Thermo Fisher Scientific, Waltham, MA, USA) with a DionexTM CarbopacTM PA10 anion-exchange column, as described by Li et al. (2018). Three biological replicates were used for each sample.

Analysis of X-ray diffraction patterns and branching degrees

To obtain XRD patterns, starch samples (100 mg each) were scanned using an X'Pert Pro X-ray diffractometer (PANalytical, Almelo, Netherlands) and was performed by Sanshu Biotechnology Co., LTD (Shanghai, China). The X-ray source used was Cu K_α with 0.154-nm filtered radiation. Each 100-mg sample was scanned at scattering angles of 5–60° (2θ) using a scanning rate of 4°/min. The degree of crystallinity was calculated using MDI-Jade v5.0 (Material Data, Inc., Livermore, CA, USA). A BioSpin NMR spectrometer (Bruker) was used to determine the branching degree. Three biological replicates were used for each sample.

Statistical analysis

Mean values and standard deviations were calculated for each sample in each measurement. One-way analysis of variance (ANOVA) and Pearson correlation analysis was conducted using SPSS version 18.0 (SPSS, Inc., Chicago, IL, USA). Significant differences were assessed with Student's *t*-test and considered statistically significant at $p < 0.05$.

Results

Mutant construction and examination of expression levels of genes involved in the starch biosynthesis pathway

The *IR36ae* mutant with *sbe2b* was identified by sequencing (Supplementary Figure 1A). Double mutants of *sbe2b*/RNAi-*sbe1* and triple mutants of *sbe2b*/RNAi-*sbe1*/OE-Wx^a were successfully generated. RNAi-*sbe1* and OE-Wx^a were identified through gel band analysis (Supplementary Figures 1B, C).

First, we assessed important genes involved in starch biosynthesis to determine the impact of *SBEIIB*, *SBEI*, and Wx^a on expression levels of other genes related to the starch biosynthesis pathway, subsequently affecting amylose and RS levels (Figure 1). The expression levels of *ISA3* (after 5 days of flowering) in the double and triple mutants were 5 and 11 times higher, respectively, compared to the wild type. The study further revealed that the lack of activity of *SBEIIb* and suppressed expression of *SBEI* had a significant impact on the expression of *SSIIa* and *SSIIIa*. However, they have a significant increase in the expression of *isoamylase 3* (*ISA3*), *ADP-glucose pyrophosphorylase 2b* (*AGPSIIb*), and Wx^a. The expression levels of Wx^a and *pullulanase* after 10 days of flowering in single mutants were more than 5 and 9 times higher,

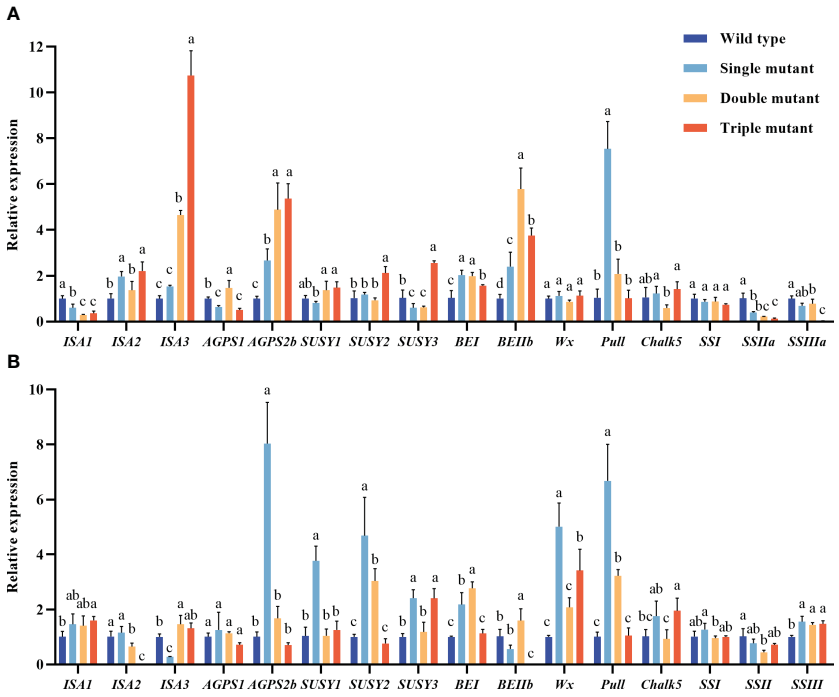


FIGURE 1
Genes expression levels in caryopsis from starch biosynthesis pathway. **(A)** Genes expression from 5 days after flowering; **(B)** Genes expression from 10 days after flowering. Values are expressed as mean ± SD (n=3). Single mutant represents mutation of gene *sbe2b*; Double mutant represents mutation of genes *sbe2b* and RNAi-*sbe1*; Triple mutant represents mutation of genes *sbe2b*, RNAi-*sbe1* and over expressed gene Wx^a. Different superscript letters represent significant differences at $P < 0.05$, and significant was compared in same index.

respectively, as compared to the wild type. This indicates that mutated *sbe2b* and RNAi*sbe1* influence the expression of the other genes associated with the starch biosynthesis pathway.

Amylose and RS contents were significantly increased with an increase in the number of mutated genes

We recorded agronomic traits for three consecutive years because amylose content is influenced by various factors, including agronomic traits. The results exhibited that the wild type had the highest seed-setting rate at 79.51%, followed by single and triple mutants at 73.45% and 62.67% respectively. The lowest seed setting was observed for a double mutant with a 46.96% value. The 1000-grain weight was similar to the seed setting rate, with the highest weight observed in wild type, followed by single mutant, and triple mutant, and the lowest was in double mutant, with weights of 24.18, 19.80, 17.61, and 15.18 g, respectively. In brief, the double mutant had a notable impact on seed setting rate and 1000-grain weight, resulting in a significant decrease in the yield (Figures 2A–D and Supplementary Table 3). In contrast, the triple mutant significantly restored the 1000-grain weight and seed-setting rate similar to those of the single mutants. This indicates that OE-Wx^a contributes to the restoration of yield.

The amylose and RS contents were analyzed for three consecutive years. Their range was highest in the triple mutant over three years, followed by the double and single mutants, with the lowest contents in the wild type (Figures 2E, F and Supplementary Table 4). In 2022, the triple mutant had amylose and RS contents of 41.92% and 4.63%, respectively. The double mutants had lower contents with 35.87% for amylose and 3.49% for RS. The single mutant has even lower contents of amylose (30.57%) and RS (1.79%). The lowest contents were found in the wild-type, at 22.19% and 0.86% for amylose and RS, respectively. This indicates that amylose and RS content significantly increased as the number of mutated genes increased.

RVA properties gradually decreased with an increasing number of mutated genes

The study then examined RVA properties to identify whether Wx^a, *sbe2b*, and *sbe1* have an impact on these properties. Analysis of RVA spectra showed that most of the indices showed a decreasing tendency, except for peak temperature (PT). For instance, the breakdown viscosity in the wild type was 4, 7, and 41 times higher than that of single, double, and triple mutants, respectively (Figure 3). This indicates that *sbe1*, *sbe2b*, and Wx^a all significantly influenced the RVA properties, with their effect gradually diminishing with an increasing number of mutated genes.

Genes *SBEIIb* and Wx^a influenced swelling power

The gel consistency was the highest in the triple mutant (3.03 cm), followed by the wild type (2.77 cm), double mutant (2.73 cm), and single mutant (2.60 cm) (Table 1). Water solubility was highest in the double mutant (8.04%), followed by the single

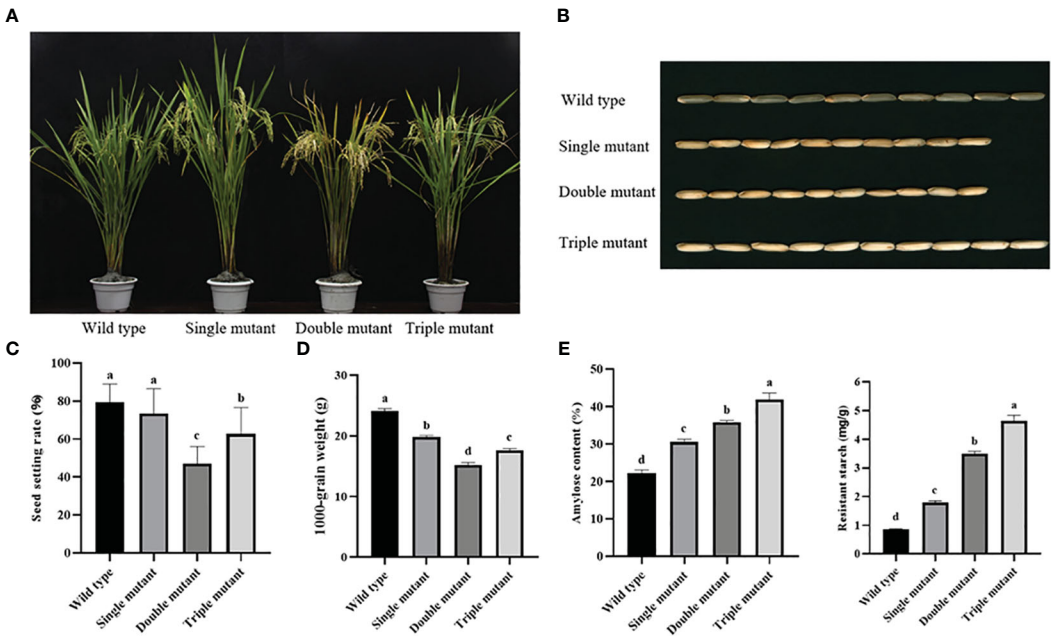
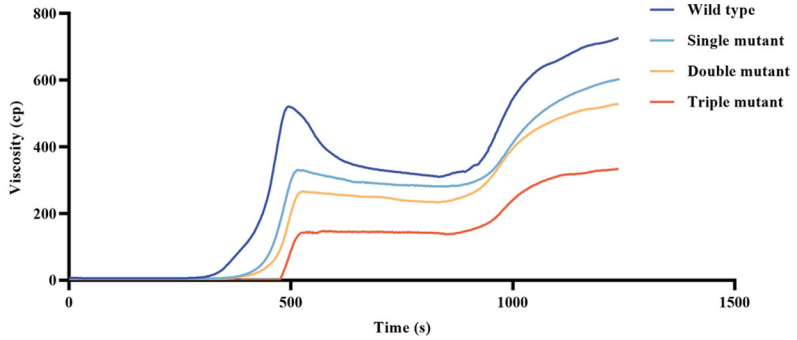


FIGURE 2
Phenotype, seed retting rate, 1000-grain weight, amylose and resistant starch content in wild type and three mutants. **(A)** Phenotype of plant; **(B)** Appearance of mature dehulled grains; **(C)** Seed setting rate; **(D)** 1000-grain weight; **(E)** Amylose content; **(F)** Resistant starch content. All agronomic traits and data were obtained from 2022. Single mutant represents mutation of gene *sbe2b*; Double mutant represents mutation of genes *sbe2b* and RNAi-*sbe1*; Triple mutant represents mutation of genes *sbe2b*, RNAi-*sbe1* and over expressed gene Wx^a. Values are means ± SD (n=3). Different superscript letters represent significant differences at $P < 0.05$.

A



B

Samples	PT (°C)	PKV (cP)	HPV (cP)	CPV (cP)	SBV(cP)	BDV (cP)	CSV (cP)
Wild type	72.2±1.20^{c}	521±1.32^{a}	313±2.42^{a}	707±1.49^{a}	186±0.97^{c}	208±0.63^{a}	394±1.12^{a}
Single mutant	79.1±0.94^{b}	331±2.11^{b}	281±1.72^{b}	582±2.15^{b}	251±1.01^{a}	50±0.09^{b}	301±0.78^{b}
Double mutant	79.9±0.87^{b}	266±1.25^{c}	235±1.35^{c}	515±1.57^{c}	249±0.47^{a}	31±0.03^{b}	280±0.72^{c}
Triple mutant	89.4±0.91^{a}	147±1.01^{d}	142±0.41^{d}	326±1.04^{d}	179±0.34^{c}	5±0.02^{c}	184±0.51^{d}

FIGURE 3
Rapid viscosity analyzer (RVA) spectra of samples from wild and three mutants. **(A)** RVA spectral; **(B)** RVA spectral indexes. PT, peak temperature; PKV, peak viscosity; HPV, hot paste viscosity; CPV, cool paste viscosity; SBV, setback viscosity; BDV, breakdown viscosity; CSV, consistence viscosity. Single mutant represents mutation of gene *sbe2b*; Double mutant represents mutation of genes *sbe2b* and RNAi-*sbe1*; Triple mutant represents mutation of genes *sbe2b*, RNAi-*sbe1* and over expressed gene Wx^{a}. Values are expressed as the mean ± standard deviation (n = 3). Superscript letters represent significant differences at $p < 0.05$, and significant was compared in same index.

mutant (6.99%), triple mutant (6.66%), and wild type (5.92%). However, the swelling power showed a different pattern compared to gel consistency and water solubility. It was highest in the wild type (12.56 g/g), followed by the double mutant (8.75 g/g), single mutant (8.67 g/g), and triple mutant (7.69 g/g). This indicates that the *SBEIIb* and Wx^{a} *genes* significantly influence swelling power.

Genes *sbe1*, *sbe2b*, and OE-Wx^{a} altered the starch granule morphology

First, we observed a gradual decrease in transparency with an increasing number of mutated genes. The triple mutant indicated the lowest transparency, while the wild type displayed high transparency (Figures 4A–D). The starch granule morphology was determined by examining the cross-section, and the results revealed that the granule morphology of the wild type was crystalline and closely spaced, while the single, double, and triple mutants had rounder and looser granules (Figures 4F–I). Granule morphology was further observed using isolated starch granules, which confirmed that the crystalline structure was only present in the wild type, while all three mutants displayed rounder and looser granules (Figures 4J–M). We also examined the particle size of starch, and the results were in the following order: single mutant > triple mutant > wild type > double mutant. It is indicated that the single mutant had the largest starch particles. The triple mutant had a significantly larger starch particle size than the wild-type and smaller particles compared to the single mutant. In addition, there was no significant difference in starch particle size between the wild type and double mutants (Figure 5). These findings indicate that genes *SBE1*, *SBEIIb*, and Wx^{a} all play a role in regulating starch granule morphology.

sbe1, sbe2b and Wx^{a} genes altered polydispersity, but no cumulative effect was observed with an increasing number of genes

Polydispersity is a measure of the broadness of molecular weight distribution, and can be calculated via X-ray-based analysis (Zhang et al., 2020). The single mutant significantly decreased both the number average molecular weight (Mn) and mean molecular weight (Mz). These values were more than twice as low as that of the wild type. The Mn and Mz values of the double and triple mutants were identical to those of the single mutant. Next, we analyzed the two most important polydispersity indices: Mw/Mn and Mz/Mn. These indices were significantly lower in the wild type compared to the single, double, and triple mutants. However, no significant differences were observed among the three mutants (Table 2). This indicates that genes *SBE1*, *SBEIIb*, and Wx^{a} all influence polydispersity, but their effects are not cumulative.

All three genes altered chain length and starch composition, but no cumulative effect was observed with an increasing number of genes

SBEIIb plays a specific role in the formation of branches in the crystalline lamellae of amylopectin clusters in the rice endosperm and affects the distribution of chain length in amylopectin. The current study revealed, that the single mutant *sbe2b* significantly decreased A (DP ≤ 12) and B1 (13 ≤ DP ≤ 24) chains, these chains were 19% and 6% lower, respectively, compared to the wild type. On

TABLE 1 GC, water solubility and swelling power of wild type and three mutants.

Samples	GC (cm)	Water solubility (%)	Swelling power (g/g)
Wild type	2.77 ± 0.06^{b}	5.92 ± 0.20^{b}	12.56 ± 0.31^{a}
Single mutant	2.60 ± 0.10^{c}	6.99 ± 0.33^{ab}	8.67 ± 0.33^{b}
Double mutant	2.73 ± 0.25^{b}	8.04 ± 1.36^{a}	8.75 ± 0.18^{b}
Triple mutant	3.03 ± 0.06^{a}	6.66 ± 0.71^{ab}	7.69 ± 0.63^{c}

GC: gel consistency. Single mutant represents mutation of gene sbe2b; Double mutant represents mutation of genes sbe2b and RNAi-sbe1; Triple mutant represents mutation of genes sbe2b, RNAi-sbe1 and over expressed gene Wx^{a}. Values are expressed as mean ± SD (n=3). Different superscript letters represent significant differences at $P < 0.05$, and significant was compared in same index.

the other hand, *Sbe2b* significantly increased B2 (25 ≤ DP ≤ 36), B3 (37 ≤ DP ≤ 76), peak 2, and peak 3 (DP ≥ 100) by 25%, 37%, 28%, and 26%, respectively, compared to the wild type (Figures 6A, B, D). Thus, all three genes altered chain length and starch composition, but there was no cumulative effect when more genes were involved.

The *sbe1, sbe2b*, and OE-*Wx*^{a} genes did not alter the diffraction patterns of starch

The highly ordered crystalline structure of starch is attributed to intra- and intermolecular hydrogen bonding. XRD is a widely used method to assess this structure (Garg and Jana, 2011). Our study showed that single, double, and triple mutants displayed A-type diffraction patterns similar to the wild type (Figure 6C). However, we found that peak degrees significantly differed between the wild type and mutants as well as among the mutants. We further found that the 17° peak was highest in the wild type, followed by the single and triple mutants. In contrast, the single mutant had the highest peak (18°), followed by the double and triple mutants, while the wild type had the lowest intensity (Figure 6E). In brief, mutated *sbe1*, *sbe2b*, and OE-*Wx*^{a} genes did not alter the diffraction patterns of starch.

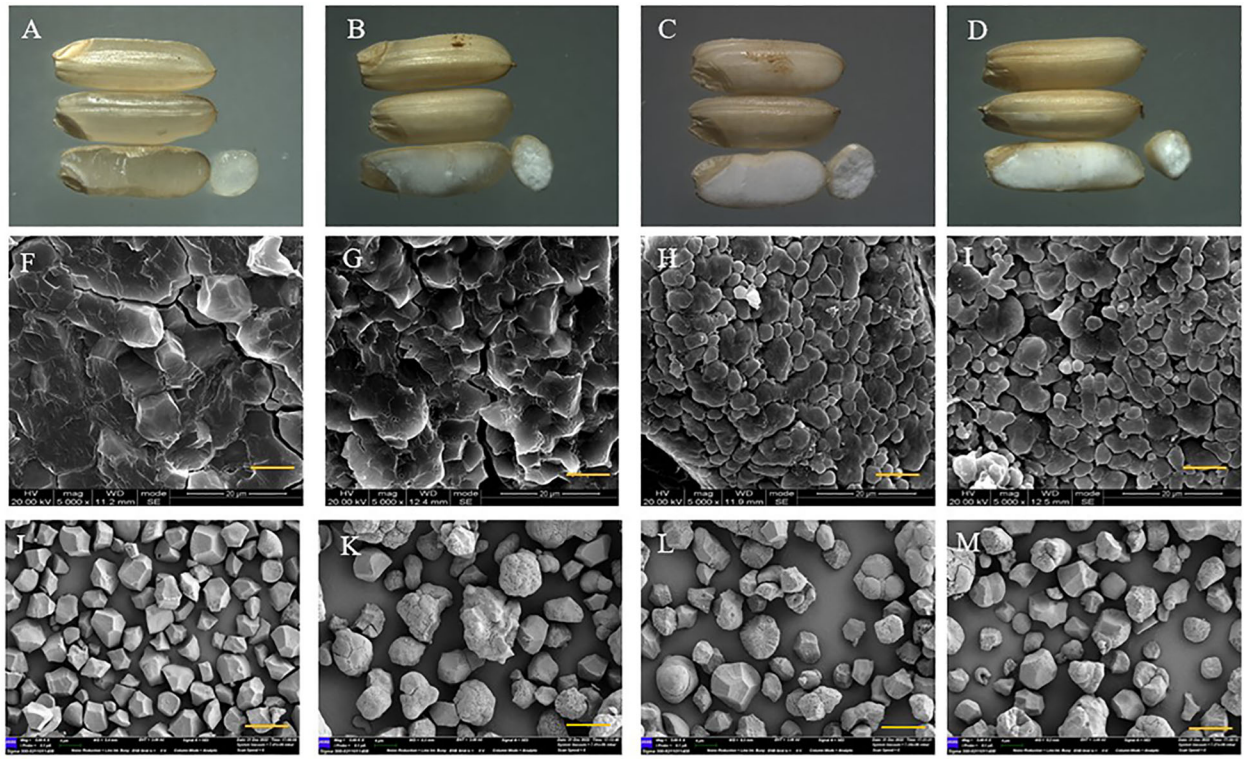
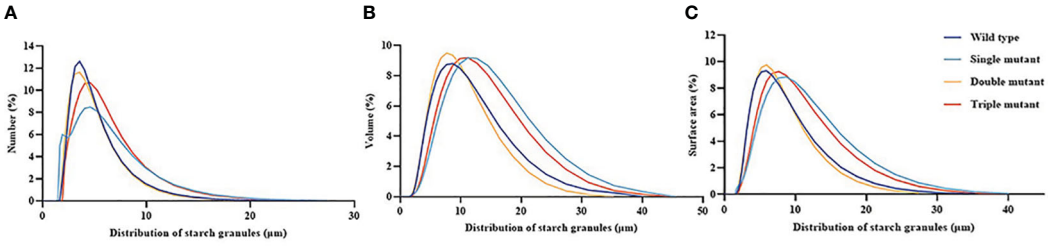


FIGURE 4
Morphology of seeds and endosperm from wild type and three mutants. **(A–D)** Morphology of seeds for wild type, single mutant, double mutant, and triple mutant, respectively; **(F–I)** Scanning electron micrographs of cross section of grains at 2000 times for wild type, single mutant, double mutant, and triple mutant, respectively, bar represses 20 μm; **(J–M)** Scanning electron micrographs of starch granules at 7000 times for wild type, single mutant, double mutant, and triple mutant, respectively, bar represses 3 μm. Single mutant represents mutation of gene *sbe2b*; Double mutant represents mutation of genes *sbe2b* and RNAi-*sbe1*; Triple mutant represents mutation of genes *sbe2b*, RNAi-*sbe1* and over expressed gene *Wx*^{a}.

10.3389/fpls.2024.1452520



D

Samples	Mean diameter (μm)			d(0.1)	d(0.5)	d(0.9)
	Number	Volume	Surface area			
Wild type	4.74±0.07	10.25±0.32	7.71±0.21	3.49±0.06	6.53±0.16	13.53±0.45
Single mutant	5.29±0.03	13.53±0.15	10.17±0.15	4.18±0.04	8.81±0.12	18.03±0.25
Double mutant	4.60±0.02	9.47±0.15	7.33±0.09	3.37±0.02	6.36±0.07	12.53±0.15
Triple mutant	5.66±0.06	12.50±0.20	9.54±0.07	4.21±0.04	8.23±0.05	16.63±0.25

FIGURE 5
Particle size of starch extracted from the grains of wild type, single, double and triple mutant. **(A–C)**, Distributions of **(A)** size, **(B)** volume, and **(C)** surface area among starch granules. **(D)** Distribution of the granule diameter of wild type, single mutant, double mutant and triple mutant. Single mutant represents mutation of gene *sbe2b*; Double mutant represents mutation of genes *sbe2b* and RNAi-*sbe1*; Triple mutant represents mutation of genes *sbe2b*, RNAi-*sbe1* and over expressed gene Wx^a. Values are expressed as the mean ± standard deviation (n = 3).

The *sbe1, sbe2b*, and Wx^a genes significantly influenced crystallinity and branching degree

The single mutant displayed the highest degree of crystallinity (25.05%), while the wild type had the lowest value (18.26%). In addition, the degree of crystallinity did not significantly differ between the double (20.16%) and triple mutants (20.50%). Our study further revealed that no notable variation was observed among the wild-type, double-mutant, and triple-mutant strains. Regarding the branching degree, the lowest value was observed in the single mutant (1.66%), but there were minimal differences among the wild-type, double, and triple mutants, similar to the degree of crystallinity (Figure 6E). This indicates that *sbe2b* greatly influenced the branching degree, while *sbe1* contributed in restoring the reduced branching degree caused by the mutation of *sbe1*.

Discussion

A daily diet rich in RS has been proven beneficial to human health, particularly in diabetic people (Topping et al., 2003). However, the previous study indicated that increasing RS content led to decreased yield due to lower seed weight. Thus, it is interesting to provide information to breed rice cultivars with greater RS content and higher yield.

The *sbe1, sbe2b*, and Wx^a genes significantly influenced the expression of genes related to starch biosynthesis pathway

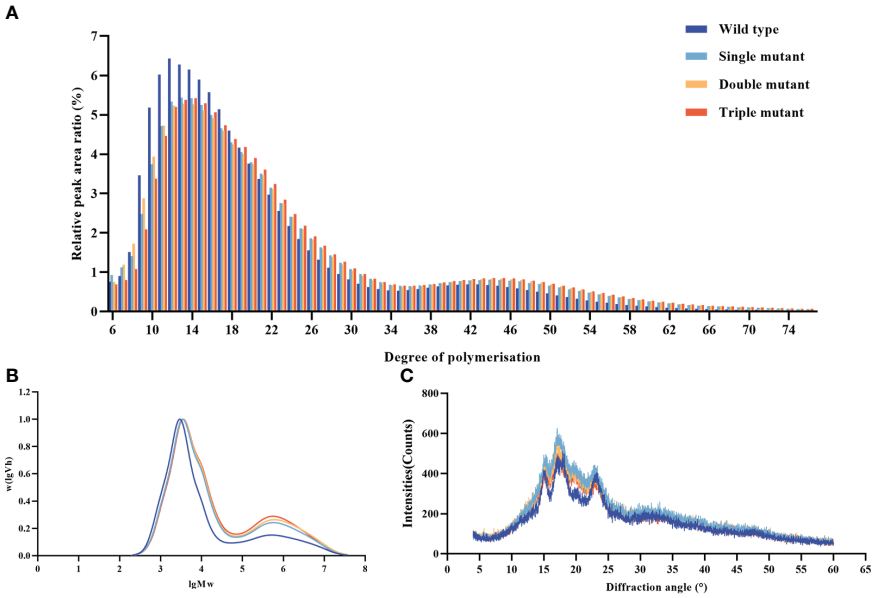
We first, evaluated the key genes related to starch biosynthesis to determine the impact of *SBEIIB*, *SBEI*, and Wx^a on the

TABLE 2 Starch molecular weights of wild and three mutants.

Samples	Mn (KDa)	Mp (KDa)	Mw (KDa)	Mz (KDa)	Mw/Mn	Mz/Mn
Wild type	39299.97 ± 69.35^a	106545.21 ± 132.74^a	80453.95 ± 57.41^a	190951.48 ± 398.55^a	2.05 ± 0.03^b	4.86 ± 0.24^c
Single mutant	17908.95 ± 21.72^c	90005.19 ± 78.93^b	47417.56 ± 36.84^c	145402.12 ± 187.51^b	2.65 ± 0.04^a	8.12 ± 0.43^a
Double mutant	23355.18 ± 25.20^b	95515.94 ± 91.85^b	59412.73 ± 42.68^b	159032.01 ± 254.99^b	2.54 ± 0.03^a	6.81 ± 0.30^b
Triple mutant	20245.19 ± 19.87^b	88479.58 ± 66.42^c	50485.92 ± 64.27^b	137864.39 ± 226.37^c	2.49 ± 0.02^a	6.81 ± 0.35^b

Single mutant represents mutation of gene sbe2b; Double mutant represents mutation of genes sbe2b and RNAi-sbe1; Triple mutant represents mutation of genes sbe2b, RNAi-sbe1 and over expressed gene Wx^a. Mn: number average molecular weight; MW: weight average molecular weight; MZ: mean molecular weight; MP: peak molecular weight. Mw/Mn and MZ/Mn: polydispersity Index. Values are expressed as mean ± SD (n=3), and significant was compared in same index.

10.3389/fpls.2024.1452520



D

Samples	A (DP≤12)	B_1 (13≤DP≤24)	B_2 (25≤DP≤36)	B_3 (37≤DP≤76)	Peak1 (%)	Peak2 (%)	Peak3 (%)
Wild type	24.28±0.14^a	52.62±0.26^a	11.10±0.09^b	12.00±0.10^c	54.61±2.11^a	24.33±0.91^b	21.06±0.59^c
Single mutant	19.73±0.09^c	49.75±0.32^c	13.87±0.13^a	16.65±0.08^a	42.40±1.78^b	31.16±0.78^a	26.44±0.61^b
Double mutant	20.46±0.11^b	48.92±0.28^c	13.51±0.12^a	17.12±0.21^a	40.8±1.45^b	31.11±0.94^a	28.09±0.76^a
Triple mutant	17.69±0.15^d	50.53±0.39^b	14.11±0.08^a	17.67±0.16^a	39.35±1.55^b	31.84±0.86^a	28.81±0.81^a

E

Samples	15°	17°	18°	20°	23°	Crystallinity degree (%)	Crystal pattern	Branching degree (%)
Wild type	5.2±0.23^c	9.3±0.10^a	0.7±0.01^c	1.9±0.09^b	5.3±0.13^b	18.26±0.32^b	A	9.89^a
Single mutant	12.3±0.46^a	5.9±0.34^b	6±0.35^a	3.2±0.06^a	6±0.11^a	25.05±0.27^a	A	1.66^b
Double mutant	9.3±0.27^b	4±0.21^c	4.7±0.15^b	4.3±0.13^a	3±0.08^d	20.16±0.54^b	A	10.48^a
Triple mutant	9.8±0.57^{ab}	5±0.16^{bc}	3.7±0.12^b	3±0.10^a	4.4±0.7^c	20.50±0.48^b	A	9.89^a

FIGURE 6
Starch chain length and X-ray diffraction (XRD) patterns of starch granules in grains collected from wild type and three mutants. **(A)** Detection of chain lengths from six to 76 degrees of polymerization (DP) via ion chromatography. **(B)** Detection of chain lengths with ≥ 100 DP in via gel permeation chromatography. **(C)** XRD patterns. Peak 1, A-chains and short B-chains of amylopectin; peak 2, intermediate and long chains of amylopectin; peak 3, amylose with chain length and DP ≥ 100. **(D)** Different chain length distribution; **(E)** Crystallinity degree, crystal pattern and branching degree. Single mutant represents mutation of gene *sbe2b*; Double mutant represents mutation of genes *sbe2b* and RNAi-*sbe1*; Triple mutant represents mutation of genes *sbe2b*, RNAi-*sbe1* and over expressed gene Wx^a. Values are expressed as the mean ± standard deviation (n = 3). Superscript letters represent significant differences at $p < 0.05$.

expression levels of other genes, as well as on amylose and RS levels (Figure 1). The results indicated that the absence of *SBEIIb* activity and suppressed expression of *SBEI* significantly inhibited the expression of *SSIIa* and *SSIIIa*, while significantly increased the expression of *isoamylase 3* (*ISA3*), *ADP-glucose pyrophosphorylase 2 b* (*AGPSIIb*), and Wx^a. However, our findings contradict those of Baysal et al. (2020) who reported that lack of SBEIIb activity triggered the activity of SSI, SSIIa, and AGPSIIb, while leading to downregulation of the genes encoding GBSS, SDBE, pullulanase, and starch phosphorylase. Because of these conflicting results, future studies should examine amylose and RS contents combined with the gene expression levels, to identify genes contributing to the formation of higher RS content. This would aid in providing information for breeding rice cultivars with a higher RS content.

OE-Wx^a significantly increased RS content and restored agronomic traits

Our study focused on observing agronomic traits, which are directly related to starch contents. The results indicated that the double mutant had a significant decrease in the seed setting rate and 1000-grain weight compared to the single mutant, ultimately leading to a significant decrease in yield (Figures 2A–D and Supplementary Table 3). In contrast, the triple mutant significantly restored the 1000-grain weight and seed-setting rate to levels similar to those of the single mutants. Miura et al. (2021) reported that mutated *sbe1* did not affect seed weight, while the mutated *sbe2b* gene did have a significant impact on seed weight. They found that *sbe1* in the double mutant (*sbe1/sbe2b)* was able to restore the lower weight of seeds caused by

lack of SBEIIb activity. However, our results were inconsistent with those of Miura et al., because the 1000-grain weight in the double mutant was 4.62 g lower than that in the single mutant. Therefore, our study not only achieved the highest RS and amylose contents but also found better agronomic traits. However, further studies are required to understand how *sbe1* and *sbe2b* regulate these traits.

The amylose and RS contents were measured, and the results showed that the triple mutant had the highest levels. We further found that induction of OE-Wx^a based on *sbe2b*/RNAi*sbe1* increased 16.9% and 32.7%, 37.1% and 158.7%, 88.9% and 438.4%, respectively, when compared to the double mutant *sbe2b*/RNAi*sbe1*, single mutant with *sbe2b*, and wild type. Similar to our study, previous studies have reported that downregulation of *SBEII* can result in high amylose content in mutant rice, maize, wheat, and barley (Wei et al., 2010; Butardo et al., 2011; Zhao et al., 2015). Double mutant *sbe1*/*sbe2b* exhibited significantly increased amylose and RS contents compared to the single mutant *sbe2b* (Miura et al., 2021). Higher expression of Wx^a significantly increased amylose and RS contents in *indica* rice compared to the wild-type (Zhou et al., 2016). Therefore, the triple mutant presented the highest RS content, by introducing OE-Wx^a, and further enhancing amylose biosynthesis through repression of *sbe1* and *sbe2b* gene expression. In addition, Zhou et al. (2016) reported that defective *SSIIIa* leads to increased RS levels, which depend on the higher expression of Wx^a. This indicates that mutations in the genes related to starch biosynthesis can further elevate RS content.

The elevated RS content was related to altering physiochemical properties and starch molecular characteristics by mutated genes *sbe1, sbe2b,* and OE-Wx^a

Our results showed that RS content gradually increased as the number of mutated genes increased. Previous studies showed that starch molecular characteristics play a significant role in RS formation (Zhou et al., 2018; Tao et al., 2022; Sangwongchai et al., 2023). Thus, our study needed to explore whether genes *sbe1*, *sbe2b*, and *Wx* play a role in RVA properties, swelling power, starch granule size, and molecular characteristics, ultimately affecting RS formation.

The *sbe1, sbe2b*, and OE-Wx^a all decreased RVA properties

The RVA properties which are vital taste values were checked, and the results showed that most of the indices exhibited decreasing tendencies, except for peak temperature (PT). For instance, the breakdown viscosity in the wild type was 4, 7, and 41 times higher compared to the single, double, and triple mutants, respectively (Figure 3). This indicates that *sbe1*, *sbe2b*, and Wx^a all had a significant impact on the RVA properties. A study by Shim et al. (2022) reported that GBSSI combined with a *sbe2b* mutant that had a mutation in one T/C SNP in the 16th exon, resulting in a change from leucine to proline, led to alteration in RVA and other physiochemical properties. Miura et al. (2022) reported that the high expression of Wx^a based on *sbe2b* mutant had a significant affected on the thermal properties of starch, which aligns with the findings of our study. This indicates that *sbe1*, *sbe2b*, and Wx^a impact RS and amylose content, along with altered physiochemical properties.

The elevated RS content was related to decreasing swelling power by mutated genes *sbe2b* and OE-Wx^a

Teng et al. (2016) reported that *Wx* alleles have an impact on the swelling property of starch. In our study, we observed that both *sbe2b* and *Wx* may decrease swelling power. It was evident from the triple mutant having the lowest swelling power, and the single mutant showing significantly lower swelling power than that of the wild type. However, there was no significant difference between the single and double mutants. Given the RS content, the results suggest that high RS content is associated with decreased swelling power due to the lack of SBEIIb and OE-*Wx* activity.

The elevated RS content was related to altering starch molecular characteristics by mutated *sbe1*, *sbe2b*, and OE-Wx^a

Butardo et al. (2011) discovered that *sbe2b* mutants, through the use of artificial microRNA- and hairpin RNA-mediated RNA silencing, altered the granule morphology of starch, resulting in rounder, looser, and more pronounced granules with larger spaces than that of the wild type. Similar to these results, Yang et al. (2016) also reported altered granule morphology. We further validated that the mutation of genes involved in starch biosynthesis leads to alterations in the starch granule structure and affects transparency. It was evident from observed differences in the granule morphology between the wild type and the single, double and triple mutants. The wild type displayed a crystalline and closely spaced, whereas the mutant presented rounder and looser granules (Figures 4F–I). The particle size of the starch also was examined, and the results showed that the single mutant had the largest starch particles. Interestingly, the starch particle size of the triple mutant was significantly larger compared to the wild type, while it was smaller compared to the single mutant. In addition, there was no significant difference in starch particle size between the wild type and double mutants (Figure 5). This indicates that the starch granule morphology is related to *SBE1*, *SBEIIb* and Wx^a genes. By considering the RS contents, it can be deduced that the increased RS content is likely influenced by factors other than altering starch granule size caused by the mutated genes *sbe2b*, *sbe1*, and OE-Wx^a.

The molecular weight was checked, and the results showed that the single mutant significantly decreased the number average molecular weight (Mn) and mean molecular weight (Mz), compared to the wild type. The Mn and Mz of the double and triple mutants were identical to those of the single mutant. Next, Mw/Mn and Mz/Mn were also analyzed and it was demonstrated

that these indices were significantly lower in the wild type compared to the single, double, and triple mutants. No significant differences were observed among the three mutants (Table 2). This indicates that genes *SBE1*, *SBEIIb*, and Wx^a all have an impact on polydispersity, but do not have cumulative effects. Combining RS content indicates that the molecular weight of starch is closely related to increased RS content due to mutated genes *sbe2b*, *sbe1*, and OE-Wx^a.

Yang et al. (2016) found that the proportion of short chains (DP ≤ 13) in *sbe2b* with a single amino acid mutation was significantly lower compared to the wild type. On the other hand, there was a notable increase in long chains (DP ≥ 15). Unlike the report of Yang et al. (2016), our study showed that *sbe2b* significantly decreased A (DP ≤ 12) and B1 (13 ≤ DP ≤ 24) chains, which were 19% and 6% lower in the single mutant compared to the wild type, respectively. *Sbe2b* significantly increased B2 (25 ≤ DP ≤ 36), B3 (37 ≤ DP ≤ 76), peak 2, and peak 3 (DP ≥ 100) by 25%, 37%, 28%, and 26%, respectively, compared to those of the wild type (Figures 6A, B, D). Similar to our study, Shim et al. (2022) reported that A and B1 chains in *sbe2b* due to mutation in the 16th exon, resulting in a change from leucine to proline, were significantly decreased compared to wild type. Conversely, B_2, B3, and peak 3 (amylose) showed a considerable increase. Ying et al. (2022) reported that *IR36ae* which lacks the activity of *sbe2b* exhibited changes in the chain length, consistent with the study of Shim et al. We hypothesize that different mutant locations of *SBEIIb* result in distinct branched structures, impacting the formation of RS and the physicochemical properties of starch. Furthermore, by analyzing the expression levels of *SBEI* and Wx^a in mutant with *sbe2b*, we concluded that the absence of SBEIIb activity leads to the reduction of A and B1 chains but an increase in the B2 and B3 chains due to the significantly increased activity of GBSSI. Thus, our study re-confirms that SBEIIb is responsible for synthesizing B_1 chains rather than B2 or B3 chains. We observed that there was no cumulative effect on the ratio of B2, B3, peak 2, and peak 3, or the ratio of A, B1, and peak 1 as the number of mutated genes increased. This was because there were minimal differences among the various mutants (Figure 6D). Therefore, we discovered that the RS content gradually increased with the number of mutated genes, mainly due to the rise in amylose content.

The crystalline structure of starch was also examined. The results showed that the single, double, and triple mutants displayed A-type diffraction patterns similar to those of the wild type (Figure 6C). However, there were significant differences in the peak degrees between the wild type and mutants and also among the mutants as well. Particularly we observed that the 17° peak was highest in the wild type, followed by the single and triple mutants. On the other hand, at 18°, the peak was highest in the single mutant, followed by the double and triple mutants, and lowest in the wild type (Figure 6E). In a study by Yang et al. (2016), it was reported that the *sbe2b* mutant, with a single amino acid mutation, exhibited a C-type diffraction pattern, while the wild type displayed an A-type pattern. They also reported a B-type pattern due to a deficiency in BEIIB (Tappiban et al., 2022). Another study by Miura et al. (2021) reported that the *sbe1* single mutant displayed an A-type pattern, whereas the *sbe2b* single and *sbe2b*/*sbe1* double mutants displayed a B-type pattern. Consistent with our findings, Butardo et al. (2011) found that the *sbe2b* mutant, silenced using hairpin RNA-mediated RNA silencing, did not alter the diffraction patterns of starch in rice. We hypothesize that these different diffraction patterns are the results of mutations occurring in different locations of the genes, which impact the enzyme activity and subsequently lead to significant differences in starch structures.

Previous studies have reported that crystallinity and branching degree of starch influence RS formation (Jane et al., 1997). In the current study, we observed that the *sbe2b* significantly increased the degree of crystallinity, as the single mutant exhibited the highest degree of crystallinity. However, there was no significant difference in crystallinity, which did not significantly differ between the double (20.16%) and triple mutants (20.50%). Furthermore, our study found that there were no significant differences in crystallinity among the wild-type, double-mutant, and triple-mutant strains. In terms of branching degree, we observed that the single mutant had the lowest level, while the wild type, double, and triple mutants showed hardly any difference, similar to the degree of crystallinity (Figure 6E). This indicates that *sbe2b* greatly influence the branching degree, while *sbe1* helps to restore the reduced branching degree caused by the mutation of *sbe1*. Bringing into account the previous studies on the impact of crystallinity on starch digestibility (Chung et al., 2006; Sajilata et al., 2006; Zhou et al., 2018), our findings indicate that the increased RS content by mutated *sbe2b* is related to increasing the degree of crystallinity.

Conclusion

In brief, our study revealed that the triple mutant exhibited the highest amylose and RS content, followed by the double and single mutants, while the lowest was in the wild type. Additionally, it was observed that the over-expression of Wx^a in the triple mutant significantly improved the seed setting rate and 1000-grain weight. These findings suggest that an increase in the number of mutated genes could further enhance the RS content. This provides valuable insights into improving RS content through gene-editing in starch biosynthesis-related genes. Furthermore, it offers valuable information for breeding rice varieties with higher RS content and favorable agronomic traits as soon as possible.

Data availability statement

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

Author contributions

XC: Data curation, Project administration, Validation, Writing – original draft, Writing – review & editing. QG: Investigation, Methodology, Writing – review & editing. XY: Investigation, Writing – review & editing. MY: Conceptualization, Writing –

review & editing. JS: Conceptualization, Writing – review & editing. HF: Data curation, Writing – review & editing. HZ: Investigation, Writing – review & editing. PX: Conceptualization, Writing – review & editing. YL: Data curation, Writing – review & editing. AA: Data curation, Writing – review & editing. KD: Data curation, Writing – review & editing. XW: Conceptualization, Data curation, Project administration, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Science and Technology Department of Sichuan Province, China (Project No. 2023NSFSC1933, 2023NSFSC0215, and 2024NSFSC0333), the Sichuan National Science Foundation Innovation Research Group (Project No. 2023NSFSC 1996), and the Bureau of Science and Technology of Chengdu City, Sichuan China (Project No 2022-YF09-00036-SN and 2023-YF08-00005-SN).

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.2024.1452520/full#supplementary-material

References

Ambigaipalan, P., Hoover, R., Donner, E., and Liu, Q. (2014). Starch chain interactions within the amorphous and crystalline domains of pulse starches during heat-moisture treatment at different temperatures and their impact on physicochemical properties. *Food Chem.* 143, 175–184. doi: 10.1016/j.foodchem.2013.07.112

Awika, J. M., Piironen, V., and Bean, S. (2011). Advances in cereal science: implications to food processing and health promotion. *Sorghum Protein Structure Chemistry: Implications Nutr. Functionality* 1089, 131–147. doi: 10.1021/symposium

Ayres, N. M., Mcclung, A. M., Larkin, P. D., Bligh, H. F. J., Jones, C. A., Park, W. D. J. T., et al. (1997). Microsatellites and a single-nucleotide polymorphism differentiate apparentamylose classes in an extended pedigree of US rice germ plasm. *Theor. Appl. Genet.* 94, 773–781. doi: 10.1007/s001220050477

Baysal, C., He, W. S., Drapal, M., Villorbina, G., Medina, V., Capell, T., et al. (2020). Inactivation of rice starch branching enzyme IIb triggers broad and unexpected changes in metabolism by transcriptional reprogramming. *PNAS*. doi: 10.1073/pnas.2014860117

Bemiller, J. (2019). Carbohydrate nutrition, dietary fiber, bulking agents, and fat mimetics. *Carbohydr. Chem. Food Scientists*. doi: 10.1016/B978-0-12-812069-9.00017-0

Bird, A. R., Vuaran, M. S., King, R. A., Noakes, M., Keogh, J., Morell, M. K., et al. (2008). Wholegrain foods made from a novel high-amylose barley variety (Himalaya 292) improve indices of bowel health in human subjects. *Br. J. Nutr.* 99, 1032–1040. doi: 10.1017/S000711450783902X

Butardo, V. M., Fitzgerald, M. A., Bird, A. R., Gidley, M. J., Flanagan, B. M., Larroque, O., et al. (2011). Impact of down-regulation of starch branching enzyme IIb in rice by artificial microRNA- and hairpin RNA-mediated RNA silencing. *J. Exp. Bot.* 62, 4927–4941. doi: 10.1093/jxb/err188

Chung, H. J., Lim, H. S., and Lim, S. T. (2006). Effect of partial gelatinization and retrogradation on the enzymatic digestion of waxy rice starch. *J. Cereal Sci.* 43, 353–359. doi: 10.1016/j.jcs.2005.12.001

De la hera, E., Gomez, M., and Rosell, C. M. (2013). Particle size distribution of rice flour affecting the starch enzymatic hydrolysis and hydration properties. *Carbohydr. Polym.* 19, 421–427. doi: 10.1016/j.carbpol.2013.06.002

Englyst, H., Wiggins, H. S., and Cummings, J. H. (1982). Determination of the non-starch polysaccharides in plant foods by gas-liquid chromatography of constituent sugars as alditol acetates. *Analyst* 107, 307–318. doi: 10.1039/an9820700307

Fujita, N., Toyosawa, Y., Utsumi, Y., Higuchi, T., Hanashiro, I., Ikegami, A., et al. (2009). Characterization of pullulanase (PUL)-deficient mutants of rice (Oryza sativa L.) and the function of PUL on starch biosynthesis in the developing rice endosperm. *J. Exp. Bot.* 60, 1009–1023. doi: 10.1093/jxb/ern349

Garg, S., and Jana, A. K. (2011). Characterization and evaluation of acylated starch with different acyl groups and degrees of substitution. *Carbohydr. Polymers* 83, 1623–1630. doi: 10.1016/j.carbpol.2010.10.015

Guzman, M. K., Parween, S., Butardo, V. M., Alhambra, C. M., Anacleto, R., Seiler, C., et al. (2017). Investigating glycemic potential of rice by unraveling compositional variations in mature grain and starch mobilization patterns during seed germination. *Sci. Rep.* 7, 5854. doi: 10.1038/s41598-017-06026-0

Hennen-Bierwagen, T. A., Liu, F., Marsh, R. S., Kim, S., Gan, Q., Tetlow, I. J., et al. (2008). Starch biosynthetic enzymes from developing maize endosperm associate in multisubunit complexes. *Plant Physiol.* 146, 1892–1908. doi: 10.1104/pp.108.116285

Jane, J. L., Wong, K. S., and Mcpherson, A. E. (1997). Branch-structure difference in starches of A- and B-type X-ray patterns revealed by their Naegeli dextrins. *Carbohyd Res.* 300, 219–227. doi: 10.1016/S0008-6215(97)00056-6

Jiang, H., Srichuwong, S., Campbell, M., and Jane, J. L. (2010). Characterization of maize amylose-extender (ae) mutant starches: Part II. Structures and properties of starch residues remaining after enzymatic hydrolysis at boiling-water temperature. *Carbohydr Polym* 81, 885–891. doi: 10.1016/j.carbpol.2010.03.064

Li, H., Wen, Y., Wang, J., and Sun, B. (2018). Relations between chain-length distribution, molecular size, and amylose content of rice starches. *Int. J. Biol. Macromol.* 120, 2017–2025. doi: 10.1016/j.ijbiomac.2018.09.204

Liu, L., Ma, X., Liu, S., Zhu, C., Jiang, L., Wang, Y., et al. (2009). Identification and characterization of a novel Waxy allele from a Yunnan rice landrace. *Plant Mol. Biol.* 71, 609–626. doi: 10.1007/s11103-009-9544-4

Lu, D., and Lu, W. (2012). Effects of protein removal on the physicochemical properties of waxy maize flours. *Starch - Strke* 64, 874–881. doi: 10.1002/star.201200038

Luo, Q. Y., Chen, S., Zhu, J. Z., Ye, L. H., Hall, N. D., Basak, S., et al. (2022). Overexpression of *EiKCS* confers paraquat-resistance in rice [*Oryza sativa* L.] by promoting the polyamine pathway. *Pest Manag Science.* 78, 246–262. doi: 10.1002/ps.6628

Marlatt, K. L., White, U. A., Beyl, R. A., Peterson, C. M., Martin, C. K., Marco, M. L., et al. (2018). Role of resistant starch on diabetes risk factors in people with prediabetes: Design, conduct, and baseline results of the STARCH trial. *Contemp Clin. Trials* 65, 99–108. doi: 10.1016/j.cct.2017.12.005

Medina-Remón, A., Kirwan, R., Lamuela-Raventós, R. M., and Estruch, R. (2018). Dietary patterns and the risk of obesity, type 2 diabetes mellitus, cardiovascular diseases, asthma, and neurodegenerative diseases. *Crit. Rev. Food Sci. Nutr.* 58, 262–296. doi: 10.1080/10408398.2016.1158690

Mikami, I., Uwatoko, N., Ikeda, Y., Yamaguchi, J., Hirano, H. Y., Suzuki, Y., et al. (2008). Allelic diversification at the wx locus in landraces of Asian rice. *Theor. Appl. Genet.* 116, 979–989. doi: 10.1007/s00122-008-0729-z

Miura, S., Koyama, N., Crofts, N., Hosaka, Y., Abe, M., and Fujita, N. (2021). Generation and starch characterization of non-transgenic BEI and BEIIb double mutant rice (Oryza sativa) with ultra-high level of resistant starch. *Rice* 14, 1–6.

Miura, S., Narita, M., Crofts, N., Itoh, Y., Hosak, Y., Oitome, N. F., et al. (2022). Improving agricultural traits while maintaining high resistant starch content in rice. *Rice* doi: 10.1186/s12284-022-00573-5

Parween, S., Anonuevo, J. J., Butardo, V., Misra, G., and Sreenivasulu, N. (2020). Balancing the double-edged sword effect of increased resistant starch content and its impact on rice texture: its genetics and molecular physiological mechanisms. *Plant Biotechnol. J.* doi: 10.1111/pbi.13339

Sajilata, M. G., Singhal, R. S., and Kulkarni, P. R. (2006). Resistant starch-A review. *Compr. Rev. Food Sci. Food Saf.* 5, 1–17. doi: 10.1111/j.1541-4337.2006.tb00076.x

Sallaud, C., Meynard, D., Boxtel, J., Gay, C., Bes, M., Larmande, P., et al. (2003). Highly efficient production and characterization of T-DNA plants for rice (Oryza sativa L.) functional genomics. *Theor. Appl. Genet.* 106, 1396–1408. doi: 10.1007/s00122-002-1184-x

Sangwongchai, W., Tananuwong, K., Krusong, K., Natee, S., and Thitisaksakul, M. (2023). Starch chemical composition and molecular structure in relation to physicochemical characteristics and resistant starch content of four thai commercial rice cultivars differing in pasting properties. *Polymers (Basel)*. 15 (3), 574. doi: 10.3390/polym15030574

Shim, K. C., Adeva, C., Kang, J. W., Luong, N. H., Lee, H. S., Cho, J. H., et al. (2022). Interaction of starch branching enzyme 3 and granule-bound starch synthase 1 alleles increases amylose content and alters physico-chemical properties in japonica rice (Oryza sativa L.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.968795

Tanaka, N., Fujita, N., Nishi, A., Satoh, H., Hosaka, Y., Ugaki, M., et al. (2004). The structure of starch can be manipulated by changing the expression levels of starch branching enzyme IIb in rice endosperm. *Plant Biotechnol. J.* 2, 507–516. doi: 10.1111/j.1467-7652.2004.00097.x

Tao, K., Liu, X., Yu, W., Neoh, G., and Gilbert, R. G. (2022). Starch molecular structural differences between chalky and translucent parts of chalky rice grains. *Food Chem.* 394, 133471. doi: 10.1016/j.foodchem.2022.133471

Tao, K., Yu, W., Prakash, S., and Gilbert, R. G. (2019). High-amylose rice: starch molecular structural features controlling cooked rice texture and preference. *Carbohydrate Polymers*. 219, 251–260.

Tappiban, P., Hu, Y., Deng, J., Zhao, J., Ying, Y., Zhang, Z., et al. (2022). Relative importance of branching enzyme isoforms in determining starch fine structure and physicochemical properties of indica rice. *Plant Mol. Biol.* 108, 399–412. doi: 10.1007/s11103-021-01207-y

Teng, B., Zhang, Y., Du, S. Y., Wu, J. D., Li, Z. F., Luo, Z. X., et al. (2016). Crystalline, thermal and swelling properties of starches from single-segment substitution lines with different *Wx* alleles in rice (Oryza sativa L.). *J. Sci. Food Agric*. doi: 10.1002/jsfa.7693

Topping, D. L., Morell, M. K., King, R. A., Li, Z., Bird, A. R., and Noakes, M. (2003). Resistant starch and health — himalaya 292, a novel barley cultivar to deliver benefits to consumers. *Starch/Stärke*. 55, 539–545. doi: 10.1002/star.200300221

Wanchana, S., Toojinda, T., Tragoonrung, S., and Vanavichit, A. (2003). Duplicated coding sequence in the waxy allele of tropical glutinous rice (Oryza sativa L.). *Plant Sci.* 165, 1193–1199. doi: 10.1007/s001220050477

Wattebled, F., Dong, Y., Dumez, S., Delvallé, D., Planchot, V., Berbezy, P., et al. (2005). Mutants of Arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. *Plant Physiol.* 138, 184–195. doi: 10.1104/pp.105.059295

Wei, C., Qin, F., Zhu, L., Zhou, W., Chen, Y., Wang, Y., et al. (2010). Microstructure and ultrastructure of high-amylose rice resistant starch granules modified by antisense RNA inhibition of starch branching enzyme. *J. Agric. Food Chem.* 58, 1224–1232. doi: 10.1021/jf9031316

Yang, R., Bai, J., Fang, J., Wang, Y., Lee, G., and Piao, Z. (2016). A single amino acid mutation of OsSBEIIb contributes to resistant starch accumulation in rice. *Breed Sci.* 66, 481–489. doi: 10.1270/jsbbs.16037

Yang, J., Wang, J., Fan, F. J., Zhu, J. Y., Chen, T., Wang, C. L., et al. (2013). Development of AS-PCR marker based on a key mutation confirmed by resequencing of Wx-mp in Milky Princess and its application in japonica soft rice (Oryza sativa L.) breeding. *Plant Breed.* 132, 595–603. doi: 10.1111/pbr.12088

Ying, Y., Xu, F., Zhang, Z., Tappiban, P., and Bao, J. (2022). Dynamic change in starch biosynthetic enzymes complexes during grain-filling stages in BEIIb active and deficient rice. *Int. J. Mol. Sci.* 23 (18), 10714. doi: 10.3390/ijms231810714

Zhang, C. Q., Bing, H. U., Zhu, K. Z., Zhang, H., Leng, Y. L., Tang, S. Z., et al. (2013). QTL mapping for rice RVA properties using high-throughput re-sequenced chromosome segment substitution lines. *Rice Sci*. doi: 10.1016/S1672-6308(13)60131-6

Zhang, L., Li, X., Janaswamy, S., Chen, L., and Chi, C. (2020). Further insights into the evolution of starch assembly during retrogradation using SAXS. *Int. J. Biol. Macromol* 154, 521–527. doi: 10.1016/j.ijbiomac.2020.03.135

Zhang, L., Li, H. T., Shen, L., Fang, Q. C., Qian, L. L., and Jia, W. P. (2015). Effect of dietary resistant starch on prevention and treatment of obesity-related diseases and its possible mechanisms. *BioMed. Environ. Sci.* 28, 291–297.

Zhang, C., Zhu, J., Chen, S., Fan, X., Li, Q., Lu, Y., et al. (2019). Wx(lv), the ancestral allele of rice waxy gene. *Mol. Plant* 12, 1157–1166. doi: 10.3390/proceedings2019036140

Zhao, Y., Li, N., Li, B., Li, Z., Xie, G., and Zhang, J. (2015). Reduced expression of starch branching enzyme IIa and IIb in maize endosperm by RNAi constructs greatly increases the amylose content in kernel with nearly normal morphology. *Planta* 241, 449–461. doi: 10.1007/s00425-014-2192-1

Zhou, H., Wang, L., Liu, G., Meng, X., Jing, Y., Shu, X., et al. (2016). Critical roles of soluble starch synthase SSIIIa and granule-bound starch synthase Waxy in synthesizing resistant starch in rice. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12844–12849. doi: 10.1073/pnas.1615104113

Zhou, H., Xia, D., Zhao, D., Li, Y., Li, P., Wu, B., et al. (2021). The origin of Wx(la) provides new insights into the improvement of grain quality in rice. *J. Integr. Plant Biol.* 63, 878–888. doi: 10.1111/jipb.13011

Zhou, X., Ying, Y., Hu, B., Pang, Y., and Bao, J. (2018). Physicochemical properties and digestibility of endosperm starches in four indica rice mutants. *Carbohydr Polym* 195, 1–8. doi: 10.1016/j.carbpol.2018.04.070

Zhu, L., Gu, M., Meng, X., Cheung, S. C., Yu, H., Huang, J., et al. (2012). High-amylose rice improves indices of animal health in normal and diabetic rats. *Plant Biotechnol. J.* 10, 353–362. doi: 10.1111/j.1467-7652.2011.00667.x

Zou, J., Xu, M., Wen, L., and Yang, B. (2020). Structure and physicochemical properties of native starch and resistant starch in chinese yam (dioscorea opposita thunb.). *Carbohydr. Polymers* 237, 116188. doi: 10.1016/j.carbpol.2020.116188

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 21 August 2024
DOI 10.3389/fpls.2024.1454615

OPEN ACCESS

EDITED BY
Hao Zhou,
Sichuan Agricultural University, China

REVIEWED BY
Yu Ling,
Guangdong Ocean University, China
Guanfu Fu,
China National Rice Research Institute (CAAS), China

*CORRESPONDENCE
Xiaojian Deng
xjdeng@sicau.edu.cn
Changhui Sun
sunhui0307@163.com

†These authors have contributed equally to this work

RECEIVED 25 June 2024
ACCEPTED 05 August 2024
PUBLISHED 21 August 2024

CITATION
Li J, He C, Liu S, Guo Y, Zhang Y, Zhang L, Zhou X, Xu D, Luo X, Liu H, Yang X, Wang Y, Shi J, Yang B, Wang J, Wang P, Deng X and Sun C (2024) Research progress and application strategies of sugar transport mechanisms in rice. *Front. Plant Sci.* 15:1454615. doi: 10.3389/fpls.2024.1454615

Research progress and application strategies of sugar transport mechanisms in rice

Jun Li[1†], Changcai He[1†], Shihang Liu[1], Yuting Guo[1], Yuxiu Zhang[1], Lanjing Zhang[1], Xu Zhou[1], Dongyu Xu[1], Xu Luo[1], Hongying Liu[1], Xiaorong Yang[1], Yang Wang[2], Jun Shi[3], Bin Yang[1], Jing Wang[1], Pingrong Wang[1], Xiaojian Deng[1*] and Changhui Sun[1*]

[1]State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Rice Research Institute, Sichuan Agricultural University, Chengdu, China, [2]College of Agricultural Science, Panxi Crops Research and Utilization Key Laboratory of Sichuan Province, Xichang University, Liangshan, China, [3]Mianyang Academy of Agricultural Sciences, Crop Characteristic Resources Creation and Utilization Key Laboratory of Sichuan Province, Mianyang, China

In plants, carbohydrates are central products of photosynthesis. Rice is a staple that contributes to the daily calorie intake for over half of the world's population. Hence, the primary objective of rice cultivation is to maximize carbohydrate production. The "source-sink" theory is proposed as a valuable principle for guiding crop breeding. However, the "flow" research lag, especially in sugar transport, has hindered high-yield rice breeding progress. This review concentrates on the genetic and molecular foundations of sugar transport and its regulation, enhancing the fundamental understanding of sugar transport processes in plants. We illustrate that the apoplastic pathway is predominant over the symplastic pathway during phloem loading in rice. Sugar transport proteins, such as SUTs and SWEETs, are essential carriers for sugar transportation in the apoplastic pathway. Additionally, we have summarized a regulatory pathway for sugar transport genes in rice, highlighting the roles of transcription factors (OsDOF11, OsNF-YB1, OsNF-YC12, OsbZIP72, Nhd1), OsRRM (RNA Recognition Motif containing protein), and GFD1 (Grain Filling Duration 1). Recognizing that the research shortfall in this area stems from a lack of advanced research methods, we discuss cutting-edge analytical techniques such as Mass Spectrometry Imaging and single-cell RNA sequencing, which could provide profound insights into the dynamics of sugar distribution and the associated regulatory mechanisms. In summary, this comprehensive review serves as a valuable guide, directing researchers toward a deep understanding and future study of the intricate mechanisms governing sugar transport.

KEYWORDS

rice (*Oryza sativa* L.), carbohydrates, "source-sink" theory, sugar transporters, regulatory pathway, high-yield

1 Introduction

In plants, the primary photosynthetic products are carbohydrates, mainly monosaccharides (glucose and fructose), disaccharides (sucrose), and polysaccharides (starch), providing essential energy sources and matter for living organisms' survival and growth. In cereal production activities, harvesting more edible carbohydrates is the ultimate goal (Alexandratos and Bruinsma, 2012; Julius et al., 2017). For high-yield crop breeding, the "source-sink" theory proposed by Mason M.L. and Maskell E.J. in 1928 has become an essential guiding theory (Mason and Maskell, 1928). The transportation and distribution of carbohydrates are the most critical components in this theory. The "source" refers to organs producing and exporting carbohydrates, such as leaves. In contrast, the "sink" represents organs that receive these sugars and store them as starch, such as seeds. The "flow" corresponds to the organs involved in sugar transportation between the "source" and "sink," notably the phloem in the stem. According to this theory, high-yield breeding primarily focuses on optimizing the relationships of source-flow-sink, involving the enhancement of source organs to produce more photosynthetic products, increasing the storage capacity of sink organs, and improving the transport capacity of flow organs (Julius et al., 2017; Durand et al., 2018; Zhang and Turgeon, 2018).

Rice (*Oryza sativa* L.) plays a vital role as a primary provider of carbohydrates in the human diet and feeds more than half of the world's population. Enhancing yield per unit area remains the foremost objective in rice breeding. Since the first Green Revolution and the introduction of hybrid rice technology, there has been no significant breakthrough in rice yield for years. In agricultural production, some rice varieties have large sinks and potent sources. However, their grains are not filled fully after maturity, with a low seed-setting rate, which seriously affects yield due to poor "flow" capacity (Mohammad et al., 1993; Mathan et al., 2021a; Li et al., 2022e). Furthermore, a recent study has shown that a more robust flow function may be a critical factor in the higher yield of cultivated rice than wild rice. While wild rice exhibits a more vital photosynthetic capacity and outputs more sucrose in the leaves, a weaker vascular system and reduced sucrose transporter activity limit the efficiency of transporting sucrose to the grains and yields (Mathan et al., 2021a).

Over the past decade, research on "source" and "sink" has been more in-depth, with a large number of genes related to photosynthetic efficiency and panicle traits being cloned and studied (Chen et al., 2022). However, our understanding of "flow" regulation has lagged despite the progress in the sugar transport of rice. Therefore, it is critical to explore essential genes related to sugar transport, elucidate their functions, and uncover the regulatory mechanisms (Ma et al., 2023). In this review, we reviewed the research progress and outlined a research prospect for sugar transport to facilitate further advancements in guiding high-yield rice breeding.

2 Core carbohydrates for study: sucrose and starch

Sucrose and starch, which are fundamental components of the plant's sugar metabolism, are central indicators in flow study. By closely monitoring the levels and distribution of sucrose and starch, researchers could gain valuable insights into the dynamics of sugar transport and allocation in plants.

Sucrose, produced by photosynthesis, is the most prevalent disaccharide in plants. It is the primary transport form of sugar over long-distance in the phloem. Subsequently, a portion is allocated to support growth and development, with the remainder stored as starch in the sink (Patrick et al., 2013; Li et al., 2022c). Sucrose is a non-reducing sugar, rendering it more resistant to degradation, particularly during long-distance transport within plant vascular systems. Its low viscosity guarantees swift flow during transportation with higher speeds ranging from 0.5 to 3 $m{\cdot}h^{-1}$. Furthermore, sucrose is known for its remarkable solubility, with phloem concentrations typically falling within 200 to 1600 $mmol{\cdot}L^{-1}$, creating exceptionally high osmotic potential (van Bel, 1996; Kühn et al., 1999). These characteristics are crucial for enabling the efficient long-distance transport of carbon assimilates as sucrose within plants.

Starch, a polysaccharide composed of glucose molecules, is the primary storage form of sugar in plants. Its glucose molecule arrangement facilitates efficient hydrolysis, releasing energy when needed. This characteristic enables plants to store energy during periods of abundant photosynthesis and utilize it to support growth when necessary (Vandeputte and Delcour, 2004; Bao, 2019).

3 Process of sucrose transport

Sucrose transport in plants involves a highly coordinated process of three essential components: phloem loading, vascular transport, and unloading in the sink. These components ensure the successful delivery of sugars to where they are needed for growth, energy, and various metabolic processes (Figure 1). Furthermore, this extensive journey necessitates precise regulation and adaptability to meet plants' growth and survival demands (Williams et al., 2000; Wen et al., 2022; Li et al., 2022c; Singh et al., 2023).

3.1 Phloem loading

The first crucial step in sucrose transport is loading it into the phloem, which occurs in source tissues such as leaves. Generally, plants have three major loading pathways: symplastic, apoplastic, and polymer trap pathways (Zhang and Turgeon, 2018).

3.1.1 Symplastic pathway

The symplastic pathway could be the primary route initially taken for sucrose transport in the leaves. This pathway utilizes plasmodesmata, which are intercellular connections that provide channels for cytoplasmic continuity. By these channels, sucrose enters the phloem SE-CC (Sieve Element-Companion Cell) complex through passive diffusion along a concentration gradient. It does not require sucrose transport carrier proteins to cross membranes and does not consume energy, but it may face additional biological limitations, such as the distribution and connectivity of plasmodesmata (Miras et al., 2022). However,

10.3389/fpls.2024.1454615

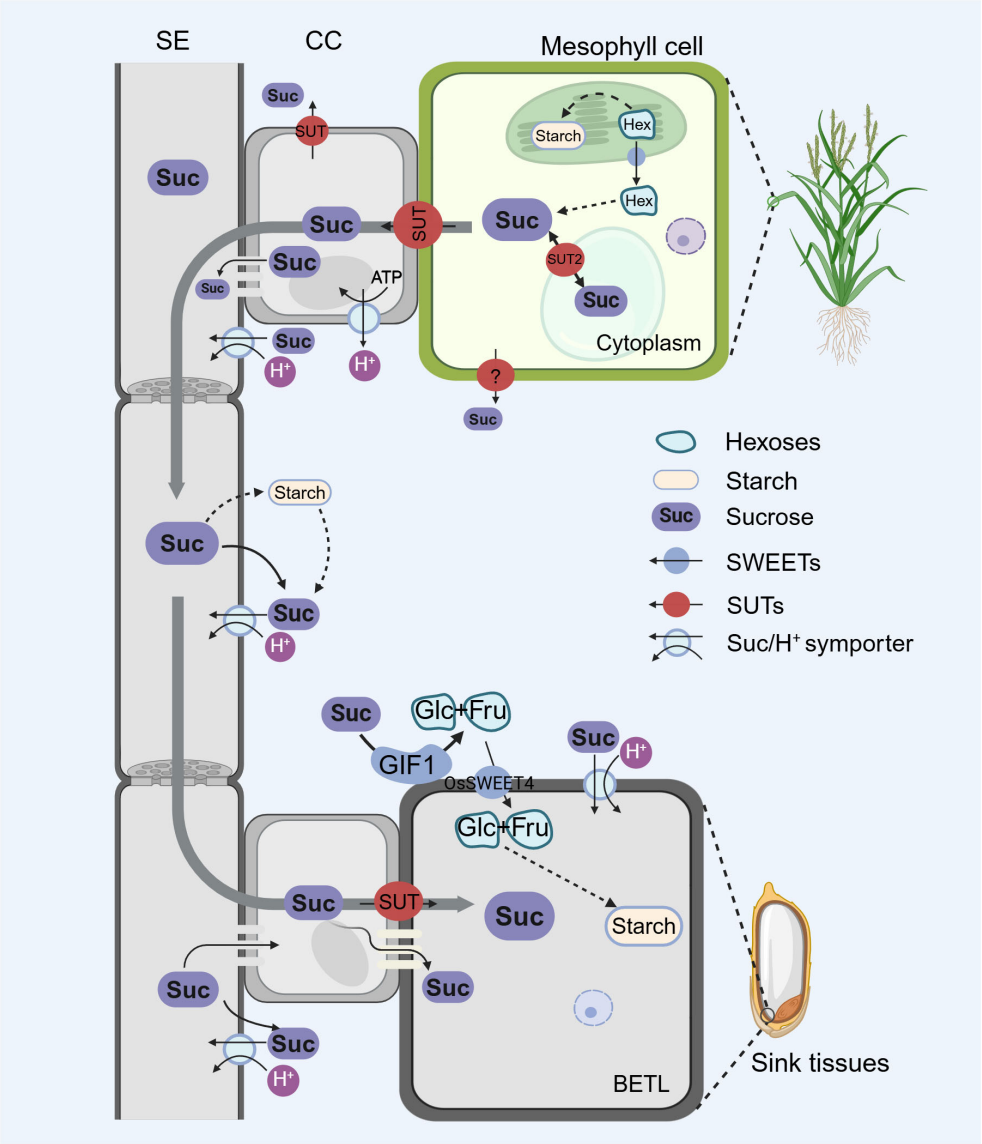


FIGURE 1
Long-distance transport of sucrose from source leaves to sink tissues, adapted from (Aoki et al., 2012). Sugars are produced in the mesophyll cells of source leaves, including hexoses, and are exported into the cytoplasm via sugar transport proteins. After undergoing a series of enzymatic catalysis reactions, hexoses are converted into sucrose. Sucrose enters the CC through symplastic and apoplastic pathways and is loaded into the SE of the phloem by these CC, then transported to sink tissues along the pressure potential gradient within the sieve elements. Sucrose is subsequently transported from the SE-CC complex to the surrounding parenchyma cells through plasmodesmata. In the apoplastic unloading pathway, sucrose is cleaved by cell wall invertase, generating hexoses such as Glc (Glucose) and Fru (Fructose). The produced Glc and Fru molecules are then transported into the BETL (Basal Endosperm Transfer Layer) for starch synthesis via the sugar transport protein OsSWEET4. Alternatively, sucrose could also be directly transported into storage cells. The figure was created with BioRender.com.

since sucrose could leak out of the cell during its transport in the phloem, sucrose transporters are still necessary to recover sucrose into the partition and maintain a higher chemical concentration gradient during symplastic transport (Braun et al., 2014).

3.1.2 Apoplastic pathway

In some plants, the SE-CC complex lacks plasmodesmata connections with surrounding cells, so the apoplastic pathway is typically used for phloem loading. Unlike the symplastic pathway, which relies on plasmodesmata, the apoplastic pathway transports sugars through the cell apoplast. This pathway may offer greater flexibility for sugar transport, particularly during long-distance movement and in response to environmental changes (Aoki et al., 2012). In this pathway, the sucrose/H^+ symporter is believed to mediate the transport of sucrose from the apoplast into SE-CC, and SWEET (Sugars Will Eventually be Exported Transporters) and SUT/SUC (Sucrose Transporters) are proved to be essential carriers. With the assistance of SWEET, sucrose is transported across membranes into the cell apoplast and then further transported to the SE-CC complex by SUT. This process entails active transport

against a concentration gradient, which necessitates energy consumption, thus the efficiency of the apoplastic pathway may be influenced by factors such as pH (Hu et al., 2021).

3.1.3 Polymer trap pathway

In the polymer trap pathway, sucrose is not directly loaded into the sieve tubes; instead, it is converted into oligosaccharides, such as raffinose and stachyose, in intermediary cells. These oligosaccharides then enter the SE-CC complex through larger plasmodesmata connections that rely on proton motive force for transport. Significantly, oligosaccharides like raffinose and stachyose could not diffuse back into mesophyll cells due to their larger molecular weight, thereby forming a kind of "trap" in the sieve tubes, which facilitates the loading of sugars (Comtet et al., 2017). This pathway plays a key role in ensuring that carbohydrates are efficiently conveyed to regions requiring them the most, thus facilitating growth and development.

3.2 Vascular transport

After phloem loading, sucrose enters the transport process within the vascular bundles. In general, vascular transport primarily involves long-distance translocation, during which metabolites are transported from the leaf vasculature to satisfy the diverse requirements of various plant tissues. The vascular bundles comprising the xylem and phloem tissues are the structural framework for metabolite transport. The phloem, primarily composed of sieve tubes and companion cells, forms an uninterrupted network throughout the plant, playing a crucial role in sucrose transport (White and Ding, 2023).

3.3 Unloading in sink

The final component is sucrose release from the phloem into sink tissues. Sink tissues can include developing grains, roots, and other organs that require a supply of sugars for growth and development. The phloem unloading process is also tightly regulated and can occur through symplastic, apoplastic pathways, or a combination of both. In the symplastic pathway, sucrose is transported from the SE-CC complex to the surrounding parenchyma cells through plasmodesmata. In the apoplastic unloading pathway, sucrose is subject to enzymatic cleavage by cell wall invertase, producing glucose and fructose. These resulting glucose and fructose molecules are then transported into the storage cells. Alternatively, sucrose can be directly transported into the storage cells (Braun, 2022; Sun et al., 2022; Wen et al., 2022; Li et al., 2022c).

4 Rice sucrose loading pathway

In grass crops, wheat, barley, and maize have been proven to use the apoplasmic loading pathway in sucrose transport, while it is still not entirely clear for rice (Julius et al., 2017). It has been reported that rice phloem sap collected by severing stylets of brown planthoppers contains primarily only one sugar, sucrose (Fukumorita and Chino, 1982). Therefore, rice does not employ the polymer trap pathway for sucrose phloem loading. The symplastic pathway or the apoplastic pathway could be used in rice.

4.1 Symplastic loading pathway in rice

Some studies suggest that rice has the fundamentals of cell structure for the symplastic loading of sucrose. Unlike maize, plasmodesmata could connect the parenchyma cells with companion cells in the leaf vascular bundles of rice (Botha et al., 2008; Eom et al., 2012). Using the characteristic that CFDA (Carboxyfluorescein Diacetate) can only be transported within the plastid after entering the plant body, the dye-feeding experiments showed that there is a continuous connection, known as symplastic continuity, between the phloem and the surrounding parenchyma cells (Scofield et al., 2007b). However, a later study reported that the frequency of such plasmodesmata connecting the cytoplasms of vascular bundles and companion cells accounted for only 1.4% of the total plasmodesmata measured across cell types (Braun et al., 2014). Though sucrose loading via a symplastic pathway may exist in rice, such loading capacity of sucrose may be relatively low. Therefore, a symplastic pathway may not be the primary pathway for sucrose transport in rice (Julius et al., 2017).

4.2 Apoplastic loading pathways in rice

Several recent studies support that sucrose phloem loading in rice predominantly takes the apoplastic pathway (Braun, 2022; Li et al., 2022b; Scofield et al., 2007a; Wang et al., 2021). Based on studies in Arabidopsis, it was hypothesized that if the membrane specificity of proton pyrophosphatases is localized within the SE-CC complex of leaves, there would be the potential existence of an apoplastic transport pathway in the plant (Gaxiola et al., 2012; Pizzio et al., 2015). In rice, immunogold labeling revealed that proton pyrophosphatases primarily localized to the plasma membrane of SE-CC complexes in the minor vein of rice leaves, reinforcing the notion of apoplastic loading. This interpretation was further supported by sucrose synthase-specific immunogold labeling at the SE-CC complex (Regmi et al., 2016).

Moreover, transgenic rice plants with overexpressed yeast invertase displayed reduced sucrose loading in the phloem, resulting in sucrose and starch accumulation in the leaves, akin to the growth-inhibited phenotype observed in *ossut1* mutants. The application of PCMBS (p-chloromercuribenzenesulfonic acid), an inhibitor of sugar transport protein function, notably hindered sucrose transport to rice phloem (Giaquinta, 1979; Wang et al., 2021). These results demonstrate the essential role of sucrose transport proteins in rice sucrose loading. Through electron microscopy, phloem tracer transportation, and ^{13}C labeling, an additional study observed that the SE-CC complex did not exhibit a symplastic link with adjacent parenchyma cells in both leaves and stems. Moreover, the reduction in ^{13}C isotope remobilization from leaves to stems and panicles upon PCMBS treatment indicated

sucrose's active transportation into the rice phloem (Li et al., 2022b). Therefore, rice is more inclined to use the apoplastic strategy for phloem loading.

5 Sucrose unloading pathway in rice

Till now, the sucrose unloading pathway in rice has been partially elucidated. Existing research indicates that, before unloading, the sucrose is initially enzymatically cleaved into glucose and fructose, with a crucial role played by a cell-wall invertase GIF1 (Grain Incomplete Filling 1) (Wang et al., 2008). Subsequently, glucose and fructose are transported into the endosperm via the sugar transporter protein OsSWEET4, which facilitates the transfer of glucose and fructose from the basal endosperm transfer layer to the endosperm for starch synthesis. Both *gif1* and *ossweet4* showed grain-filling defects, and *ossweet4* even resulted in defective seed filling, suggesting that the primary unloading pathway in rice grain might be the apoplastic pathway (Sosso et al., 2015).

6 Sugar transporters and their regulators in rice

Based on the above analysis, the primary pathway for sucrose transport in rice is the apoplastic pathway, underscoring the pivotal role of sugar transporters in this physiological process. In plants, sugar transport is facilitated by various sugar transporters, primarily including the SWEET family and MFS (Major Facilitator Superfamily). The MFS can be further categorized based on the substrate specificity into the DST (Disaccharide Transporter) family and the MST (Monosaccharide Transporter) family (Williams et al., 2000; Büttner, 2007; Deng et al., 2019; Zhou M. et al., 2023). In addition, there are still some other transporters specific to various forms of sugar, which are beyond the current scope of this review.

In rice, SWEETs and the disaccharide transporter SUTs, have been recognized as crucial facilitators in sugar transport (Eom et al., 2011; Julius et al., 2017; Mathan et al., 2021a). Although some other types of sugar-related transport proteins have been reported, such as OsMEX1 (Maltose Excess 1), OsTMTs (Tonoplast Monosaccharide Transporters), OsNope1 (No perception 1), OsBT1 (Brittle 1) and so on, further introduction is not provided here because these studies are currently neither in-depth nor systematic enough (Cho et al., 2010; Ryoo et al., 2013; Cakir et al., 2016; Nadal et al., 2017; Li et al., 2017; Song et al., 2020).

6.1 SUT family

SUT proteins have 12 transmembrane domains, forming a pore to transport sucrose across the plasma membrane (Riesmeier et al., 1992). The *SUT* family can be divided into three clades (Aoki et al., 2003). Type I *SUTs* are exclusive to eudicots, while type II and III *SUTs* are found in all plants. The rice genome encodes five *SUTs*; four are type II, and one (*OsSUT2*) is type III (Aoki et al., 2003). The five *SUTs* possess diverse roles in both sink and source tissues (Aoki et al., 2003). *OsSUT1*, *OsSUT3*, *OsSUT4*, and *OsSUT5* encode proteins in the plasma membrane, whereas *OsSUT2* encodes a protein in the vacuolar membrane (Aoki et al., 2003; Eom et al., 2011; Siao et al., 2011).

In Arabidopsis, AtSUC2 mediates efflux and retrieval via the apoplast pathway (Srivastava et al., 2008, 2009). Interestingly, *OsSUT1*, the type II SUT like *AtSUC2*, was able to complement the *atsuc2* mutant, indicating that *OsSUT1* is a functional homolog of *AtSUC2* for sucrose transporter (Matsukura et al., 2000; Furbank et al., 2001; Scofield et al., 2007a, 2007b; Hirose et al., 2010; Eom et al., 2016). *OsSUT1* is highly expressed in the endosperm, leaf sheaths, nodes, vascular parenchyma, and nucellar projection (Matsukura et al., 2000). However, it is curious that no visible phenotype was observed in *ossut1* knockdown plants at the beginning. Typical phenotypes of sugar transporter mutants, such as sugar accumulation within leaves and retarded growth, are not followed (Scofield et al., 2002; Hirose et al., 2010). Further study exhibited that it might be due to the low demand for sugar transport during the growth and development of rice in the greenhouse with low light. In the field-grown conditions, the *ossut1* mutant showed reduced growth and grain yield due to sucrose and starch accumulation in the leaves (Feng et al., 2018). Therefore, OsSUT1 could function to load sucrose from the apoplast into the phloem in rice leaves.

OsSUT2 is highly expressed in lateral roots, inflorescence, stems, seed coats, leaf mesophyll, and bundle sheath cells but not in the veins (Aoki et al., 2003; Eom et al., 2011; Siao et al., 2011; Eom et al., 2012). *OsSUT2* may be involved in vacuolar-to-cytosolic sucrose translocation, influencing sucrose movement into the phloem (Sun et al., 2008; Eom et al., 2011; Siao et al., 2011; Eom et al., 2012). The mutant *ossut2* demonstrated distinctive growth-inhibited characteristics, shedding light on the crucial role of OsSUT2 in rice sugar transport mechanisms. Sugar export from the leaf was also reported to be significantly reduced in *ossut2* mutants, associated with notable accumulations of sucrose, glucose, and fructose in mature leaves (Eom et al., 2011, 2012). *OsSUT3* is preferentially expressed in pollen, indicating its involvement in pollen development and maturity rather than phloem loading in source leaves (Aoki et al., 2003; Scofield et al., 2007b; Li et al., 2020). *OsSUT4* shows expression dominance in leaves and stems. The *ossut4* mutant deficiency exhibited an accumulation of sucrose and starch in leaves and a significant reduction in net photosynthetic rate, which are the typical phenotypes of phloem loading obstruction in the vascular tissue (Mengzhu et al., 2020). *OsSUT5* exhibits the highest expression level in storage leaves (Aoki et al., 2003). Interestingly, through biochemical activity studies in *Xenopus* laevis oocytes, OsSUT5 was found to have a higher substrate affinity for sucrose and less substrate specificity than OsSUT1. Thus, OsSUT5 could potentially replace the function of OsSUT1 in sucrose phloem loading (Sun et al., 2010).

6.2 SWEET family

SWEETs are identified as sugar transporters containing seven transmembrane domains, which are present in all organism kingdoms (Yuan and Wang, 2013; Anjali et al., 2020; Singh et al., 2023). SWEETs mediate the bidirectional transmembrane transport of sugar substances, playing crucial roles in various physiological processes throughout plant growth and development. Among the 21 paralogs in rice, OsSWEET4 and OsSWEET11 are believed to play a predominant role in grain filling and yield formation (Zhou et al., 2014; Morii et al., 2020; Li et al., 2022d). Similar to OsSWEET4, OsSWEET11 is crucial in facilitating sucrose transport from maternal tissue to the maternal-filial interface in the stages of caryopsis development. Knockout of *OsSWEET11* led to decreased sucrose concentration in embryo sacs. Still, it increased the sucrose content and reduced starch levels in mature caryopses (Ma et al., 2017; Li et al., 2022d).

OsSWEET15 was found to share a similar function with OsSWEET11 in regulating the efflux of sucrose across the nucellar epidermis/aleurone interface. Notably, in the double mutant *ossweet11 15*, starch accumulation was observed in the pericarp, while a reduction in starch content was evident in the endosperm. This dynamic led to the development of shriveled caryopses (Ma et al., 2017; Yang et al., 2018). The single knockout mutants of *ossweet14* did not exhibit discernible phenotypic changes. In contrast, the double mutant *ossweet11 14* displayed considerably exacerbated phenotypes compared to the single-knockout mutants of *ossweet11*. These severe manifestations included a significant grain weight and yield reduction, a slower grain-filling rate, and an elevated accumulation of starch in the pericarp (Ma et al., 2017; Eom et al., 2019; Fei et al., 2021). Paradoxically, other studies have shown that *ossweet14* had a shorter plant phenotype and produced smaller seeds (Antony et al., 2010; Li et al., 2022d).

OsSWEET5, which acts as a galactose transporter, is believed to regulate plant growth and senescence by interacting with auxin signaling pathways (Zhou et al., 2014). Overexpressed *OsSWEET5* plants exhibited growth retardation and early senescence during the seedling stage. However, no discernible phenotype differences were observed in *OsSWEET5*-knockdown lines, suggesting the functional redundancy of galactose transporters existed in rice (Zhou et al., 2014). OsSWEET3a likely serves as a key player in facilitating the loading of glucose and GA (Gibberellin Acid) from the endosperm into the phloem for subsequent transport, thereby influencing early shoot development in rice. *OsSWEET3a* is primarily expressed in the vascular bundles of basal rice seedlings, while mutants lacking *OsSWEET3a* have significant deficiencies during germination and the initial stages of shoot development. Intriguingly, these impairments showed improvement when exogenous GA was introduced (Morii et al., 2020).

It is worth noting that overexpressing sugar transport proteins in a rough manner did not increase yield in rice. Some studies overexpressed *OsSWEET11* or *OsSWEET14*. However, the transgenic lines showed unfavorable phenotypes (Yuan et al., 2009; Gao et al., 2018; Singh et al., 2021; Kim et al., 2021b). The overexpression of *OsSWEET11* plants was markedly shorter and had fewer tillers than the wild type (Yuan et al., 2009), and *OsSWEET14* overexpressed plants resulted in a dwarf phenotype, reduced 1,000-grain weight and grain number per panicle, and tillers (Kim et al., 2021b). Another study even co-expressed three sugar transporters, *OsSWEET11*, *OsSWEET14*, and *OsSUT1*, which got similar results to the plants that overexpress either *OsSWEET11* or *OsSWEET14* (Singh et al., 2021). These results showed that simply overexpressing sugar transporters does not improve the transport capacity of the flow but may have the opposite effect.

6.3 Transcriptional regulation of sugar transporters

Sugar transport through the phloem can be affected by many environmental factors that alter source-sink relationships, including abiotic and biotic, such as drought, saline-alkali, heat, mineral deficiency, pathogenic microbes, viruses, and so on. However, most of the exact mechanisms are far from being completely known at the molecular level (Krasensky and Jonak, 2012; Lemoine et al., 2013; Mathan et al., 2021a; Salvi et al., 2022).

Recent reports indicate that sugar transporters are intricately connected with various regulatory pathways involving ABA (Abscisic Acid), GA, stress responses, nitrogen uptake, circadian rhythm, etc. Several transcription factors regulate the expression of rice sugar transporter genes, thereby controlling the sugar transport rate into various tissues (Yanagisawa and Sheen, 1998; Laloum et al., 2013; Li et al., 2013; Bai et al., 2016; Wu et al., 2018; Xiong et al., 2019; Liu et al., 2020; Sun et al., 2021; Zhang et al., 2021a; Kim et al., 2021b; Mathan et al., 2021b; Lee et al., 2022; Li et al., 2022a, 2023) (Figure 2).

DOF (DNA binding with one finger) transcription factors can recognize promoter elements containing (A/T)AAAG and subsequently regulate various aspects of plant development (Yanagisawa and Sheen, 1998). OsDOF11 is identified as a regulator of *SUT* and *SWEET* genes. It exhibits expression in vascular cells within photosynthetic organs and various sink tissues. The *osdof11* mutants displayed a semi-dwarf phenotype with reduced tiller numbers and smaller panicles, ultimately leading to a decreased sucrose transport rate compared to the wild type. Furthermore, the expression profiles of transporter genes, including *OsSUT1*, *OsSUT3*, *OsSUT4*, *OsSUT5*, *OsSWEET11*, and *OsSWEET14*, are significantly altered in the *osdof11* mutant. Chromatin immunoprecipitation assays have provided conclusive evidence for direct binding of OsDOF11 to the promoter regions of *OsSUT1*, *OsSWEET11*, and *OsSWEET14*. Interestingly, overexpression of *OsDOF11* enhanced resistance to sheath blight, while reducing yield production. Alternatively, tissue-specific activation of OsDOF11 via VP16 significantly improves yield production and enhances sheath blight, resistance, underscoring the pivotal role of tissue-specific in optimizing sugar transport-related genes (Li et al., 2013; Kim et al., 2021b).

NF-Y (Nuclear factor-Y) is a class of conserved transcription factors that binds to the promoter's CCAAT box (Laloum et al., 2013). In rice, the *osnf-yb1* mutant exhibited chalky endosperm and defective grain-filling. RNA *in situ* hybridization assays displayed

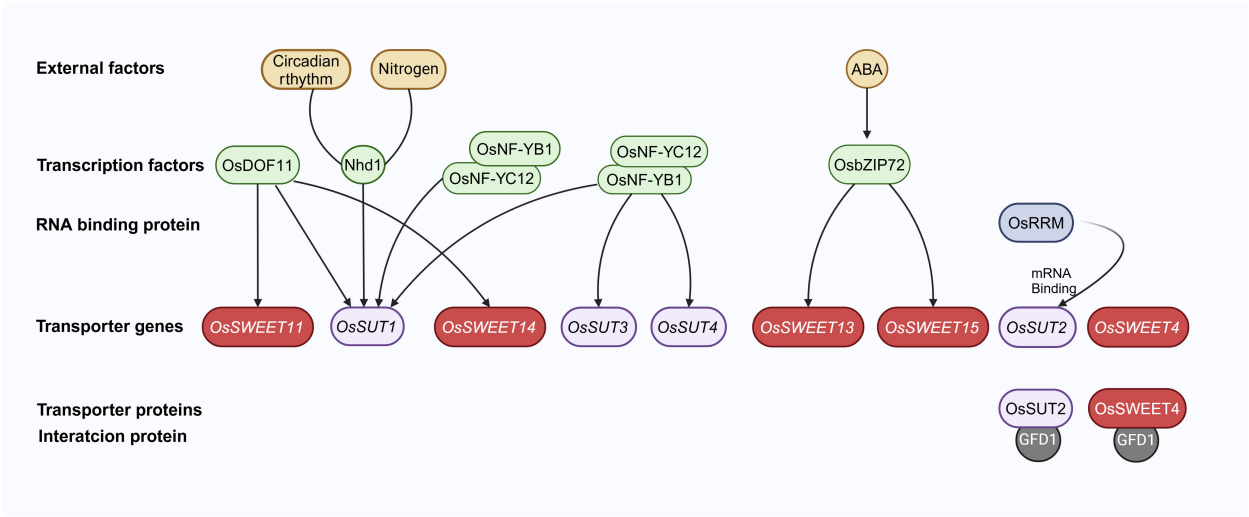


FIGURE 2
A simple regulation pathway of sugar transporters in rice. Transcription factors, such as OsDOF11, OsNF-YB1, OsNF-YC12, OsbZIP72, and Nhd1, directly regulate sugar transporter genes. These transcription factors are linked to the circadian clock, nitrogen absorption, and ABA stress signaling. The mRNA of *OsSUT2* could be directly bound by an RNA Recognition Motif containing protein, OsRRM. The GFD1 protein directly interacts with both OsSUT2 and OsSWEET4. Arrows represent positive regulation of sugar transporter genes. The figure was created with BioRender.com.

that *OsNF-YB1* was localized to the aleurone layer. EMSA (Electrophoretic Mobility Shift Assays) demonstrated that OsNF-YB1 could directly bind to the CCAAT box in the promoters of *OsSUT1*, *OsSUT3*, and *OsSUT4* (Bai et al., 2016). Another study showed that OsNF-YC12 was directly bound to the promoter of *OsSUT1* using various ChIP (Chromatin Immunoprecipitation) and Y1H (Yeast One-Hybrid assays). Furthermore, OsNF-YC12 could also physically interact with OsNF-YB1 (Xiong et al., 2019). Therefore, OsNF-YB1 and OsNF-YC12 dimers could regulate the expression of *OsSUTs* in the aleurone layer during the grain-filling stage.

OsbZIP72 encodes a transcription factor responsive to ABA. Transactivation experiments have demonstrated that OsbZIP72 could directly bind to the promoters of *OsSWEET13* and *OsSWEET15* and promote their expressions. This finding suggests a plausible mechanism for regulating sucrose transport and allocation in response to abiotic stresses, and such a mechanism can potentially maintain sugar homeostasis in rice when faced with challenges such as drought and salinity (Mathan et al., 2021b).

An MYB transcription factor Nhd1 (N-mediated heading date-1), also known as a core clock component factor OsLHY (Late Elongated Hypocotyl), has been reported to regulate nitrogen uptake and flowering time (Sun et al., 2021; Zhang et al., 2021a; Lee et al., 2022; Li et al., 2022a). A subsequent study uncovered that the deletion of *Nhd1* led to the inhibition of OsSUT1 expression. Notably, when compared to the wild type, both *nhd1* and *ossut1* mutants displayed similar characteristics, including reduced plant height, decreased shoot biomass, and lower sucrose concentration. Intriguingly, the overexpression of *OsSUT1* could restore the impaired sucrose transport, partially alleviating the compromised growth observed in *nhd1* mutants. EMSA, transient transactivation assays, and ChIP-qPCR in rice leaf protoplasts provided evidence that Nhd1 could directly activate the transcriptional expression of *OsSUT1*, suggesting a conventional regulatory module involving *Nhd1* and *OsSUT1* that maintains carbon and nitrogen balance in rice (Li et al., 2023).

A recent study identified a novel mechanism for the expression regulation of rice sugar transporter. An OsRRM (RNA Recognition Motif) containing protein could interact with mRNAs of sugar transporter genes and positively regulate their expression levels, such as *OsSUT2* (Liu et al., 2020). To summarize, sugar transporter genes' expression in rice is dynamically controlled by transcription factors and RNA-binding proteins to fine-tune sugar homeostasis (Dhungana and Braun, 2021).

6.4 Interacting proteins of sugar transporters

The transport activity of many transporters depends on the interactions of membrane proteins (Lalonde et al., 2010). For example, Arabidopsis nitrate transporter proteins NRT2.1 and NRT2.2 require interaction with the nitrate assimilation-related protein NAR2 to form a 150 kDa plasma membrane complex, which enables high-affinity nitrate transport (Quesada et al., 1994; Boone et al., 2007; Wirth et al., 2007; Lalonde et al., 2008; Jonikas et al., 2009; Yong et al., 2010; Jones et al., 2014; Sun et al., 2023). In Arabidopsis, the MIND (Membrane-based Interaction Network Database) revealed many interacting proteins of sugar transporter proteins (Cusick et al., 2005). However, their physiological and genetic functions are rarely reported in plants (Bitterlich et al., 2014; Jones et al., 2014; Abelenda et al., 2019; Xu et al., 2020; Wipf et al., 2021).

In rice, the interacting proteins of sugar transporters were unknown before a MATE (Multidrug and toxic compound

extrusion) transporter protein GFD1 (Grain Filling Duration 1) was studied in our recent study (Sun et al., 2023). MATE transporter is one of the most prominent families of transporter proteins for cation transportation in most organisms (Omote et al., 2006; Takanashi et al., 2014; Upadhyay et al., 2019; Qin et al., 2021b; Zhang et al., 2021b; Zhou C. et al., 2023). Initially, due to the 12 transmembrane domains shared with SUT proteins, it was classified as a member of MFS (Marger and Saier, 1993; Omote et al., 2006). However, due to differences in protein sequence, it was later defined as an independent MATE transporter class (Brown et al., 1999).

In our recent study, GFD1 was identified as a MATE transporter protein that may be associated with sugar transport and starch accumulation in rice (Sun et al., 2023). The mutant *gfd1* disrupted the balance of carbohydrate distribution in the stem and grain but not in the leaves. The mutant showed a prolonged grain filling period, increased grain size, and reduced number of spikelets. Through Y2H (Yeast Two-Hybrid), LCI (Firefly luciferase Complementation Imaging assay), and BiFC (Bimolecular Fluorescence Complementation), it was proven that GFD1 could interact with two sugar transporters, OsSWEET4 and OsSUT2. Genetic analysis using double mutants indicated that GFD1 might control the grain filling period through OsSWEET4 and regulate grain size through OsSUT2, while working with OsSUT2 and OsSWEET4 to handle the number of grains per spike. However, the interaction effects on the functions of OsSUT2 and OsSWEET4 remain unclear (Figure 2).

On the other hand, in addition to the photosynthetic products from the leaves, the formation of rice grain yield also relies significantly on the starch stored in the stem before flowering. Stem starch can contribute up to 30% of rice yield, and this contribution rate increases with increased storage capacity, particularly under adverse environmental conditions. When rice photosynthesis is impaired, the stem becomes a primary source of assimilates for grain filling, significantly mitigating the negative effects of stress on yield (CocK and Yoshida, 1972). Furthermore, stored starch in the stem could also be a crucial factor contributing to the increased yield of cultivated rice compared to wild rice (Mathan et al., 2021a). Therefore, increasing the accumulation of starch in rice stems and improving the transport efficiency are essential to enhance rice yield. However, the molecular mechanism of starch accumulation in the stem remains unclear. Our study showed that GFD1 is an important regulatory factor for starch accumulation in the stem, and the deficient mutant *gfd1* significantly accumulates starch particles in the stem before grain filling (Sun et al., 2023). Therefore, GFD1 plays a significant role in the long-distance sugar transport within the stem.

7 Prospects for sugar transport study and application strategies in rice

As biotechnology advances, focusing solely on studying sugar transport in organs such as roots, stems, leaves, and panicles is no longer sufficient. The development of single-cell sequencing and MSI (Mass Spectrometry Imaging) opens up new possibilities for a more precise understanding of the sugar transport mechanism. In order to facilitate a better understanding of the molecular mechanisms underlying sugar transport, we propose a study prospect here (Figure 3).

7.1 Genetic material construction

With the rapid advancement of gene editing technologies (Cong et al., 2013), obtaining a complete set of mutants for the genes of SUT and SWEET families has become more accessible. For instance, researchers have extensively employed CRISPR/Cas9 knockout techniques to generate mutants for most SWEETs and conducted comprehensive studies on these mutants (Li et al., 2022d). Furthermore, due to the redundancy of sugar transport proteins, it is now possible to create double, triple, or multiple mutants, thereby providing a rich genetic resource for investigating the precise interactions between these sugar transporters.

7.2 Functional analysis at the organ level

To elucidate the function of sugar transporters, measuring the sugar and starch content in different tissues at various developmental stages is essential. It could help uncover the differences in the distribution of sugars and starches across tissues (Sun et al., 2023). Such analyses will lay the foundation for further in-depth investigations into their microscopic functional differences using single-cell transcriptomics and spatial metabolomics techniques.

7.3 Single-cell transcriptomics

Our expanding knowledge highlights the remarkable heterogeneity in each cell, even within the same cell type. This intricate diversity and the various cell types in tissues amplify the need for spatially resolved single-cell omics techniques (Gilmore et al., 2019). In recent years, high-throughput single-cell transcriptomics has significantly progressed (Potter, 2018). Profiling multiple cells allows for a precise examination, unveiling unique insights into developmental processes, transcriptional pathways, and the molecular intricacies within tissues (Denyer et al., 2019). A notable investigation centered on vascular cells within Arabidopsis leaves, culminating in establishing a single-cell transcriptome of leaf vasculature. This comprehensive study successfully identified a minimum of 19 distinct cell clusters. Remarkably, among these clusters, PP cells exhibited a substantial enrichment of transporter genes, with prominent candidates including *SWEET11* and *SWEET12* (Kim et al., 2021a).

Leveraging this technology allows for the precise determination of the expression patterns of each sugar transporter gene at a single-cell resolution. Investigating these patterns is vital, encompassing the examination of single-cell expression patterns of sugar transporter genes, analysis of changes in downstream gene expression within sugar transporter mutants, and exploration of single-cell expression variations in upstream genes of the mutants. Furthermore, employing additional techniques such as *in situ* hybridization and

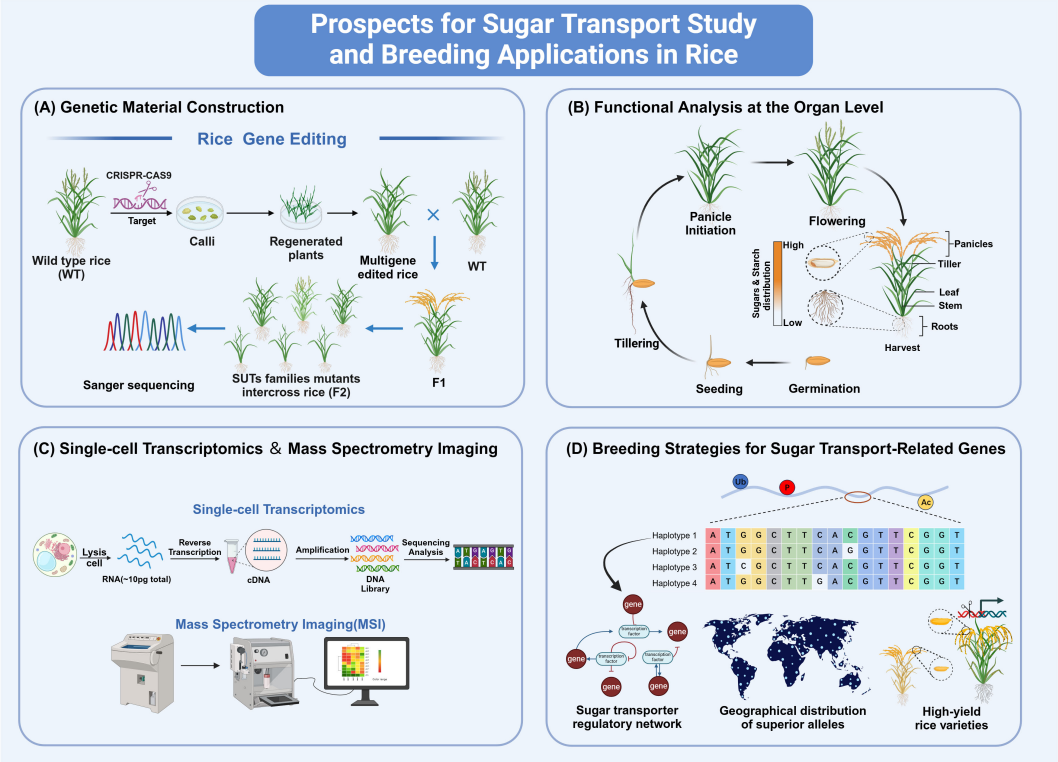


FIGURE 3
Prospects for sugar transport study and breeding applications in rice. **(A)** Gene editing technologies such as CRISPR/Cas9 have simplified the creation of mutants for SUT and SWEET families, facilitating the study of interactions between sugar transport proteins. **(B)** Measuring sugar and starch in tissues at various stages can reveal differences in their distribution. **(C)** Applying Single-cell transcriptomics and Spatial metabolomics allows for precise examination of multiple cells. **(D)** Using tissue-specific promoters, gene editing, and synthetic breeding with rice databases to select superior alleles can optimize sugar transport genes and boost rice yield, which is crucial for breeding. The figure was created with BioRender.com.

immunofluorescence is pivotal and will substantially expedite the study of rice sugar transport (Qin et al., 2021b).

7.4 High-resolution spatial metabolomics technology

In recent decades, MSI has emerged as a powerful tool for non-targeted, label-free chemical imaging. It eliminates the need to pre-select specific molecular species for analysis, making detecting compounds without chemical modification or labeling agents possible. MSI encompasses various techniques such as MALDI-MSI (Matrix-assisted Laser Desorption Ionization Mass Spectrometry Imaging), DESI-MSI (Desorption Electrospray Ionization Mass Spectrometry Imaging), LAESI-MSI (Laser Ablation Electrospray Ionization Mass Spectrometry Imaging), SIMS (Imaging Secondary Ion Mass Spectrometry) and so on (Spengler, 2015; Buchberger et al., 2018).

By employing MSI, we can gain precise insights into the spatiotemporal distribution variations of sugars, including sucrose, fructose, and glucose (Hu et al., 2017). Utilizing MALDI-MSI, a recent study accurately depicted the spatial distribution of free soluble sugars (glucose, fructose, sucrose, and maltose) in rice seeds during the dough stage. These four free soluble sugars exhibited similar spatial distribution patterns, primarily concentrated in the seed cortex and embryo with high abundance (Zhao et al., 2023).

7.5 Breeding strategies for sugar transport-related genes

Merely elevating the expression of sugar transporters may not inevitably augment phloem transport capacity and has the potential to yield undesired outcomes. As demonstrated, the continuous overexpression of *OsDOF11* leads to decreased yield, while the tissue-specific activation of *OsDOF11* significantly enhances yield (Wu et al., 2018; Kim et al., 2021b). This underscores the crucial role of tissue-specific approaches when optimizing genes associated with sugar transport. Consequently, the strategic utilization of tissue-specific and site-specific promoters bears substantial significance.

In addition, the extensive publication of rice germplasm resources and genomic data has laid the foundation for identifying superior alleles of sugar transport-related genes (Wang et al., 2018; Qin et al., 2021a; Wang et al., 2023). Firstly, screening variant sites (including SNPs and InDels) within the promoters and coding regions of sugar transport-related genes and clustering analysis of their Haps (Haplotypes). Secondly, utilizing available databases of the yield-related traits, including 1000-grain weight, panicle grain number, and sugar distribution characteristics, such as 600 rice resources database

(http://ricevarmap.ncpgr.cn/), analyzed their association with haplotypes of sugar transport-related genes for superior allelic variants identification. Thirdly, superior allelic variants enable functional validation and the creation of breeding materials in major rice cultivars. This could be achieved through precise gene editing and integrated breeding techniques, which facilitate the precise management of the balance of source-sink-flow.

In conclusion, the union of knowledge and methodological approaches concerning the cellular and tissue-level partitioning of sugar transporters, the elucidation of sugar transporters regulating networks, and their crosstalk with developmental mechanisms is imperative for a more sophisticated comprehension and precise manipulation of sugar transport mechanisms. This integrative perspective not only refines our grasp of the underlying regulatory models governing sugar transport but also prompts the optimization of rice germplasm. The strategic application of such knowledge is pivotal for enhancing rice yield potential and for stimulating its response to the environment.

Author contributions

JL: Writing – original draft, Writing – review & editing. CH: Writing – original draft, Writing – review & editing. SL: Writing – review & editing. YG: Writing – review & editing. YZ: Writing – review & editing. LZ: Writing – review & editing. XZ: Writing – review & editing. DX: Writing – review & editing. XL: Writing – review & editing. HL: Writing – review & editing. XY: Writing – review & editing. YW: Writing – review & editing. JS: Writing – review & editing. BY: Writing – review & editing. JW: Writing – review & editing. PW: Writing – review & editing. XD: Funding acquisition, Investigation, Supervision, Writing – review & editing. CS: Conceptualization, Funding acquisition, Investigation, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by grants from the Natural Science Foundation of Sichuan Province (2023NSFSC0213, 2024NSFTD0022, and 2022NSFSC1748), National Natural Science Foundation of China (32172022, 31371602, 31401358, and 32260502), Sichuan Science and Technology Program (2020YJ0408), Sichuan Rice Innovation Team of National Modern Agricultural Industry Technology System (CXTD-2024-01).

Acknowledgments

The figures were created with BioRender.com.

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

References

Abelenda, J. A., Bergonzi, S., Oortwijn, M., Sonnewald, S., Du, M., Visser, R. G. F., et al. (2019). Source-sink regulation is mediated by interaction of an FT homolog with a SWEET protein in potato. *Curr. Biol.* 29, 1178–1186 e1176. doi: 10.1016/j.cub.2019.02.018

Alexandratos, N., and Bruinsma, J. (2012). *World agriculture towards 2030/2050: the 2012 revision*. Rome: ESA Working Paper. doi: 10.22004/ag.econ.288998

Anjali, A., Fatima, U., Manu, M. S., Ramasamy, S., and Senthil-Kumar, M. (2020). Structure and regulation of SWEET transporters in plants: An update. *Plant Physiol. Biochem.* 156, 1–6. doi: 10.1016/j.plaphy.2020.08.043

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., et al. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. *Plant Cell* 22, 3864–3876. doi: 10.1105/tpc.110.078964

Aoki, N., Hirose, T., and Furbank, R. T. (2012). "Sucrose transport in higher plants: from source to sink," in *Photosynthesis. Advances in Photosynthesis and Respiration*, vol. 34 . Eds. J. Eaton-Rye, B. Tripathy and T. Sharkey (Springer, Dordrecht). doi: 10.1007/978-94-007-1579-0_28

Aoki, N., Hirose, T., Scofield, G. N., Whitfeld, P. R., and Furbank, R. T. (2003). The sucrose transporter gene family in rice. *Plant Cell Physiol.* 44, 223–232. doi: 10.1093/pcp/pcg030

Bai, A. N., Lu, X. D., Li, D. Q., Liu, J. X., and Liu, C. M. (2016). NF-YB1-regulated expression of sucrose transporters in aleurone facilitates sugar loading to rice endosperm. *Cell Res.* 26, 384–388. doi: 10.1038/cr.2015.116

Bao, J. (2019). "Rice starch," in *Rice* (Amsterdam: Elsevier), 55–108.

Bitterlich, M., Krugel, U., Boldt-Burisch, K., Franken, P., and Kuhn, C. (2014). The sucrose transporter SlSUT2 from tomato interacts with brassinosteroid functioning and affects arbuscular mycorrhiza formation. *Plant J.* 78, 877–889. doi: 10.1111/tpj.12515

Boone, C., Bussey, H., and Andrews, B. J. (2007). Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* 8, 437–449. doi: 10.1038/nrg2085

Botha, C. E., Aoki, N., Scofield, G. N., Liu, L., Furbank, R. T., and White, R. G. (2008). A xylem sap retrieval pathway in rice leaf blades: evidence of a role for endocytosis? *J. Exp. Bot.* 59, 2945–2954. doi: 10.1093/jxb/ern150

Braun, D. M. (2022). Phloem loading and unloading of sucrose: what a long, strange trip from source to sink. *Annu. Rev. Plant Biol.* 73, 553–584. doi: 10.1146/annurev-arplant-070721-083240

Braun, D. M., Wang, L., and Ruan, Y. L. (2014). Understanding and manipulating sucrose phloem loading, unloading, metabolism, and signalling to enhance crop yield and food security. *J. Exp. Bot.* 65, 1713–1735. doi: 10.1093/jxb/ert416

Brown, M. H., Paulsen, I. T., and Skurray, R. A. (1999). The multidrug efflux protein NorM is a prototype of a new family of transporters. *Mol. Microbiol.* 31, 394–395. doi: 10.1046/j.1365-2958.1999.01162.x

Buchberger, A. R., DeLaney, K., Johnson, J., and Li, L. (2018). Mass spectrometry imaging: a review of emerging advancements and future insights. *Anal. Chem.* 90, 240. doi: 10.1021/acs.analchem.7b04733

Büttner, M. (2007). The monosaccharide transporter (-like) gene family in Arabidopsis. *FEBS Lett.* 581, 2318–2324. doi: 10.1016/j.febslet.2007.03.016

Cakir, B., Shiraishi, S., Tuncel, A., Matsusaka, H., Satoh, R., Singh, S., et al. (2016). Analysis of the rice ADP-glucose transporter (OsBT1) indicates the presence of regulatory processes in the amyloplast stroma that control ADP-glucose flux into starch. *Plant Physiol.* 170, 1271–1283. doi: 10.1104/pp.15.01911

Chen, R., Deng, Y., Ding, Y., Guo, J., Qiu, J., Wang, B., et al. (2022). Rice functional genomics: decades' efforts and roads ahead. *Sci. China Life Sci.* 65, 33–92. doi: 10.1007/s11427-021-2024-0

Cho, J. I., Burla, B., Lee, D. W., Ryoo, N., Hong, S. K., Kim, H. B., et al. (2010). Expression analysis and functional characterization of the monosaccharide transporters, OsTMTs, involving vacuolar sugar transport in rice (Oryza sativa). *New Phytol.* 186, 657–668. doi: 10.1111/j.1469-8137.2010.03194.x

CocK, J. H., and Yoshida, S. (1972). Accumulation of ^{14}C-labelled carbohydrate before flowering and its subsequent redistribution and respiration in the rice plant. *J. Crop Sci.* 41, 226–234. doi: 10.1626/jcs.41.226

Comtet, J., Turgeon, R., and Stroock, A. D. (2017). Phloem loading through plasmodesmata: A biophysical analysis. *Plant Physiol.* 175, 904–915. doi: 10.1104/pp.16.01041

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143

Cusick, M. E., Klitgord, N., Vidal, M., and Hill, D. E. (2005). Interactome: gateway into systems biology. *Hum. Mol. Genet.* 14 Spec No. 2, R171–R181. doi: 10.1093/hmg/ddi335

Deng, X., An, B., Zhong, H., Yang, J., Kong, W., and Li, Y. (2019). A novel insight into functional divergence of the MST gene family in rice based on comprehensive expression patterns. *Genes* 10, 239. doi: 10.3390/genes10030239

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M. C. P. (2019). Spatiotemporal developmental trajectories in the arabidopsis root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* 48, 840–852 e845. doi: 10.1016/j.devcel.2019.02.022

Dhungana, S. R., and Braun, D. M. (2021). Sugar transporters in grasses: Function and modulation in source and storage tissues. *J. Plant Physiol.* 266, 153541. doi: 10.1016/j.jplph.2021.153541

Durand, M., Mainson, D., Porcheron, B., Maurousset, L., Lemoine, R., and Pourtau, N. (2018). Carbon source-sink relationship in Arabidopsis thaliana: the role of sucrose transporters. *Planta* 247, 587–611. doi: 10.1007/s00425-017-2807-4

Eom, J. S., Cho, J. I., Reinders, A., Lee, S. W., Yoo, Y., Tuan, P. Q., et al. (2011). Impaired function of the tonoplast-localized sucrose transporter in rice, OsSUT2, limits the transport of vacuolar reserve sucrose and affects plant growth. *Plant Physiol.* 157, 109–119. doi: 10.1104/pp.111.176982

Eom, J. S., Choi, S. B., Ward, J. M., and Jeon, J. S. (2012). The mechanism of phloem loading in rice (Oryza sativa). *Mol. Cells* 33, 431–438. doi: 10.1007/s10059-012-0071-9

Eom, J. S., Luo, D., Atienza-Grande, G., Yang, J., Ji, C., Thi Luu, V., et al. (2019). Diagnostic kit for rice blight resistance. *Nat. Biotechnol.* 37, 1372–1379. doi: 10.1038/s41587-019-0268-y

Eom, J. S., Nguyen, C. D., Lee, D. W., Lee, S. K., and Jeon, J. S. (2016). Genetic complementation analysis of rice sucrose transporter genes in Arabidopsis mutant. *J. Plant Biol.* 59, 231–237. doi: 10.1007/s12374-016-0015-6

Fei, H., Yang, Z., Lu, Q., Wen, X., Zhang, Y., Zhang, A., et al. (2021). OsSWEET14 cooperates with OsSWEET11 to contribute to grain filling in rice. *Plant Sci.* 306, 110851. doi: 10.1016/j.plantsci.2021.110851

Feng, B., Sun, Y., Ai, H., Liu, X., Yang, J., Liu, L., et al. (2018). Overexpression of sucrose transporter OsSUT1 affects rice morphology and physiology. *CJRS* 32, 549–556. doi: 10.16819/j.1001-7216.2018.8015

Fukumorita, T., and Chino, M. (1982). Sugar, amino acid and inorganic contents in rice phloem sap. *Plant Cell Physiol.* 23, 273–283. doi: 10.1093/oxfordjournals.pcp.a076347

Furbank, R. T., Scofield, G. N., Hirose, T., Wang, X.-D., Patrick, J. W., and Offler, C. E. (2001). Cellular localisation and function of a sucrose transporter OsSUT1 in developing rice grains. *Funct. Plant Biol.* 28, 1187–1196. doi: 10.1071/pp01111

Gao, Y., Zhang, C., Han, X., Wang, Z. Y., Ma, L., Yuan, P., et al. (2018). Inhibition of OsSWEET11 function in mesophyll cells improves resistance of rice to sheath blight disease. *Mol. Plant Pathol.* 19, 2149–2161. doi: 10.1111/mpp.12689

Gaxiola, R. A., Sanchez, C. A., Paez-Valencia, J., Ayre, B. G., and Elser, J. J. (2012). Genetic manipulation of a "vacuolar" H(+)-PPase: from salt tolerance to yield enhancement under phosphorus-deficient soils. *Plant Physiol.* 159, 3–11. doi: 10.1104/pp.112.195701

Giaquinta, R. T. (1979). Phloem loading of sucrose: involvement of membrane ATPase and proton transport. *Plant Physiol.* 63, 744–748. doi: 10.1104/pp.63.4.744

Gilmore, I. S., Heiles, S., and Pieterse, C. L. (2019). Metabolic imaging at the single-cell scale: recent advances in mass spectrometry imaging. *Annu. Rev. Anal. Chem. (Palo Alto Calif)* 12, 201–224. doi: 10.1146/annurev-anchem-061318-115516

Hirose, T., Zhang, Z., Miyao, A., Hirochika, H., Ohsugi, R., and Terao, T. (2010). Disruption of a gene for rice sucrose transporter, OsSUT1, impairs pollen function but pollen maturation is unaffected. *J. Exp. Bot.* 61, 3639–3646. doi: 10.1093/jxb/erq175

Hu, X., Fang, C., Lu, L., Hu, Z., Shao, Y., and Zhu, Z. (2017). Determination of soluble sugar profile in rice. *J. Chromatogr. B* 1058, 19–23. doi: 10.1016/j.jchromb.2017.05.001

Hu, Z., Tang, Z., Zhang, Y., Niu, L., Yang, F., Zhang, D., et al. (2021). Rice SUT and SWEET transporters. *Int. J. Mol. Sci.* 22, 11198. doi: 10.3390/ijms222011198

Jones, A. M., Xuan, Y., Xu, M., Wang, R. S., Ho, C. H., Lalonde, S., et al. (2014). Border control–a membrane-linked interactome of Arabidopsis. *Science* 344, 711–716. doi: 10.1126/science.1251358

Jonikas, M. C., Collins, S. R., Denic, V., Oh, E., Quan, E. M., Schmid, V., et al. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* 323, 1693–1697. doi: 10.1126/science.1167983

Julius, B. T., Leach, K. A., Tran, T. M., Mertz, R. A., and Braun, D. M. (2017). Sugar transporters in plants: new insights and discoveries. *Plant Cell Physiol.* 58, 1442–1460. doi: 10.1093/pcp/pcx090

Kim, J. Y., Symeonidi, E., Pang, T. Y., Denyer, T., Weidauer, D., Bezrutczyk, M., et al. (2021a). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* 33, 511–530. doi: 10.1093/plcell/koaa060

Kim, P., Xue, C. Y., Song, H. D., Gao, Y., Feng, L., Li, Y., et al. (2021b). Tissue-specific activation of DOF11 promotes rice resistance to sheath blight disease and increases grain weight via activation of SWEET14. *Plant Biotechnol. J.* 19, 409–411. doi: 10.1111/pbi.13489

Krasensky, J., and Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* 63, 1593–1608. doi: 10.1093/jxb/err460

Kühn, C., Barker, L., Bürkle, L., and Frommer, W.-B. (1999). Update on sucrose transport in higher plants. *J. Exp. Bot.* 50, 935–953. doi: 10.1093/jexbot/50.suppl_1.935

Lalonde, S., Ehrhardt, D. W., Loque, D., Chen, J., Rhee, S. Y., and Frommer, W. B. (2008). Molecular and cellular approaches for the detection of protein-protein interactions: latest techniques and current limitations. *Plant J.* 53, 610–635. doi: 10.1111/j.1365-313X.2007.03332.x

Lalonde, S., Sero, A., Pratelli, R., Pilot, G., Chen, J., Sardi, M. I., et al. (2010). A membrane protein/signaling protein interaction network for Arabidopsis version AMPv2. *Front. Physiol.* 1. doi: 10.3389/fphys.2010.00024

Laloum, T., De Mita, S., Gamas, P., Baudin, M., and Niebel, A. (2013). CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci.* 18, 157–166. doi: 10.1016/j.tplants.2012.07.004

Lee, S. J., Kang, K., Lim, J. H., and Paek, N. C. (2022). Natural alleles of CIRCADIAN CLOCK ASSOCIATED1 contribute to rice cultivation by fine-tuning flowering time. *Plant Physiol.* 190, 640–656. doi: 10.1093/plphys/kiac296

Lemoine, R., Camera, S. L., Atanassova, R., Dédaldéchamp, F., Allario, T., Pourtau, N., et al. (2013). Source-to-sink transport of sugar and regulation by environmental factors. *Front. Plant Sci.* 4, 272. doi: 10.3389/fpls.2013.00272

Li, J., Blue, R., Zeitler, B., Strange, T. L., Pearl, J. R., Huizinga, D. H., et al. (2013). Activation domains for controlling plant gene expression using designed transcription factors. *Plant Biotechnol. J.* 11, 671–680. doi: 10.1111/pbi.12057

Li, J., Kim, Y. J., and Zhang, D. (2022c). Source-to-sink transport of sugar and its role in male reproductive development. *Genes (Basel)* 13, 1323. doi: 10.3390/genes13081323

Li, C., Liu, X. J., Yan, Y., Alam, M. S., Liu, Z., Yang, Z. K., et al. (2022a). OsLHY is involved in regulating flowering through the Hd1- and Ehd1- mediated pathways in rice (Oryza sativa L.). *Plant Sci.* 315, 111145. doi: 10.1016/j.plantsci.2021.111145

Li, K., Tang, S., Zhang, S., Tian, Y., Qu, H., Gu, M., et al. (2023). Rice circadian clock regulator Nhd1 controls the expression of the sucrose transporter gene OsSUT1 and impacts carbon-nitrogen balance. *J. Exp. Bot.* 74, 1460–1474. doi: 10.1093/jxb/erac494

Li, P., Wang, L. H., Liu, H. B., and Yuan, M. (2022d). Impaired-mediated sugar transportation impacts starch metabolism in developing rice seeds. *Crop J.* 10, 98–108. doi: 10.1016/j.cj.2021.04.012

Li, S., Wei, X., Ren, Y., Qiu, J., Jiao, G., Guo, X., et al. (2017). OsBT1 encodes an ADP-glucose transporter involved in starch synthesis and compound granule formation in rice endosperm. *Sci. Rep.* 7, 40124. doi: 10.1038/srep40124

Li, Z., Wei, X., Tong, X., Zhao, J., Liu, X., Wang, H., et al. (2022e). The OsNAC23-Tre6P-SnRK1a feed-forward loop regulates sugar homeostasis and grain yield in rice. *Mol. Plant* 15, 706–722. doi: 10.1016/j.molp.2022.01.016

Li, D., Xu, R., Lv, D., Zhang, C., Yang, H., Zhang, J., et al. (2020). Identification of the core pollen-specific regulation in the rice osSUT3 promoter. *Int. J. Mol. Sci.* 21, 1909. doi: 10.3390/ijms21061909

Li, G., Zhou, C., Yang, Z., Zhang, C., Dai, Q., Huo, Z., et al. (2022b). Low nitrogen enhances apoplastic phloem loading and improves the translocation of photoassimilates in rice leaves and stems. *Plant Cell Physiol.* 63, 991–1007. doi: 10.1093/pcp/pcac066

Liu, D., Xu, L., Wang, W., Jia, S., Jin, S., and Gao, J. (2020). OsRRM, an RNA-binding protein, modulates sugar transport in rice (Oryza sativa L.). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.605276

Ma, B., Zhang, L., and He, Z. (2023). Understanding the regulation of cereal grain filling: The way forward. *J. Integr. Plant Biol.* 65, 526–547. doi: 10.1111/jipb.13456

Ma, L., Zhang, D., Miao, Q., Yang, J., Xuan, Y., and Hu, Y. (2017). Essential role of sugar transporter osSWEET11 during the early stage of rice grain filling. *Plant Cell Physiol.* 58, 863–873. doi: 10.1093/pcp/pcx040

Marger, M. D., and Saier, M. H. Jr. (1993). A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem. Sci.* 18, 13–20. doi: 10.1016/0968-0004(93)90081-w

Mason, T., and Maskell, E. (1928). Studies on the transport of carbohydrates in the cotton plant: II. The factors determining the rate and the direction of movement of sugars. *Ann. Bot.* 42, 571–636. doi: 10.1093/oxfordjournals.aob.a090131

Mathan, J., Singh, A., and Ranjan, A. (2021a). Sucrose transport and metabolism control carbon partitioning between stem and grain in rice. *J. Exp. Bot.* 72, 4355–4372. doi: 10.1093/jxb/erab066

Mathan, J., Singh, A., and Ranjan, A. (2021b). Sucrose transport in response to drought and salt stress involves ABA-mediated induction of OsSWEET13 and OsSWEET15 in rice. *Physiol. Plant* 171, 620–637. doi: 10.1111/ppl.13210

Matsukura, C., Saitoh, T., Hirose, T., Ohsugi, R., Perata, P., and Yamaguchi, J. (2000). Sugar uptake and transport in rice embryo. Expression of companion cell-specific sucrose transporter (OsSUT1) induced by sugar and light. *Plant Physiol.* 124, 85–93. doi: 10.1104/pp.124.1.85

Mengzhu, L., Gaopeng, W., Yue, W., Yi, R., Ganghua, L., Zhenghui, L., et al. (2020). Function analysis of sucrose transporter OsSUT4 in sucrose transport in rice. *Chin. J. Rice Sci.* 34, 491. doi: 10.16819/j.1001-7216.2020.0316

Miras, M., Pottier, M., Schladt, T. M., Ejike, J. O., Redzich, L., Frommer, W. B., et al. (2022). Plasmodesmata and their role in assimilate translocation. *J. Plant Physiol.* 270, 153633. doi: 10.1016/j.jplph.2022.153633

Mohammad, A., Muhammad, A., and Muhammad, S. (1993). "Genetic improvement in physiological traits of rice yield," in *Genetic improvement of field crops* (Boca Raton, Florida: CRC Press), 413–455.

Morii, M., Sugihara, A., Takehara, S., Kanno, Y., Kawai, K., Hobo, T., et al. (2020). The dual function of osSWEET3a as a gibberellin and glucose transporter is important for young shoot development in rice. *Plant Cell Physiol.* 61, 1935–1945. doi: 10.1093/pcp/pcaa130

Nadal, M., Sawers, R., Naseem, S., Bassin, B., Kulicke, C., Sharman, A., et al. (2017). An N-acetylglucosamine transporter required for arbuscular mycorrhizal symbioses in rice and maize. *Nat. Plants* 3, 1–7. doi: 10.1038/nplants.2017.73

Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M., and Moriyama, Y. (2006). The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends Pharmacol. Sci.* 27, 587–593. doi: 10.1016/j.tips.2006.09.001

Patrick, J. W., Botha, F. C., and Birch, R. G. (2013). Metabolic engineering of sugars and simple sugar derivatives in plants. *Plant Biotechnol. J.* 11, 142–156. doi: 10.1111/pbi.12002

Pizzio, G. A., Paez-Valencia, J., Khadilkar, A. S., Regmi, K., Patron-Soberano, A., Zhang, S., et al. (2015). Arabidopsis type I proton-pumping pyrophosphatase expresses strongly in phloem, where it is required for pyrophosphate metabolism and photosynthate partitioning. *Plant Physiol.* 167, 1541–1553. doi: 10.1104/pp.114.254342

Potter, S. S. (2018). Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.* 14, 479–492. doi: 10.1038/s41581-018-0021-7

Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., et al. (2021a). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* 184, 3542–3558 e3516. doi: 10.1016/j.cell.2021.04.046

Qin, P., Zhang, G., Hu, B., Wu, J., Chen, W., Ren, Z., et al. (2021b). Leaf-derived ABA regulates rice seed development via a transporter-mediated and temperature-sensitive mechanism. *Sci. Adv.* 7, eabc8873. doi: 10.1126/sciadv.abc8873

Quesada, A., Galvan, A., and Fernandez, E. (1994). Identification of nitrate transporter genes in Chlamydomonas reinhardtii. *Plant J.* 5, 407–419. doi: 10.1111/j.1365-313x.1994.00407.x

Regmi, K. C., Zhang, S., and Gaxiola, R. A. (2016). Apoplasmic loading in the rice phloem supported by the presence of sucrose synthase and plasma membrane-localized proton pyrophosphatase. *Ann. Bot.* 117, 257–268. doi: 10.1093/aob/mcv174

Riesmeier, J. W., Willmitzer, L., and Frommer, W. B. (1992). Isolation and characterization of a sucrose carrier cDNA from spinach by functional expression in yeast. *EMBO J.* 11, 4705–4713. doi: 10.1002/j.1460-2075.1992.tb05575.x

Ryoo, N., Eom, J. S., Kim, H. B., Vo, B. T., Lee, S. W., Hahn, T. R., et al. (2013). Expression and functional analysis of rice plastidic maltose transporter, OsMEX1. *J. Korean Soc. Appl. Bl.* 56, 149–155. doi: 10.1007/s13765-012-3266-z

Salvi, P., Agarrwal, R., Gandass, N., Manna, M., Kaur, H., and Deshmukh, R. (2022). Sugar transporters and their molecular tradeoffs during abiotic stress responses in plants. *Physiol. Plant* 174, e13652. doi: 10.1111/ppl.13652

Scofield, G. N., Aoki, N., Hirose, T., Takano, M., Jenkins, C. L., and Furbank, R. T. (2007a). The role of the sucrose transporter, OsSUT1, in germination and early seedling growth and development of rice plants. *J. Exp. Bot.* 58, 483–495. doi: 10.1093/jxb/erl217

Scofield, G. N., Hirose, T., Aoki, N., and Furbank, R. T. (2007b). Involvement of the sucrose transporter, OsSUT1, in the long-distance pathway for assimilate transport in rice. *J. Exp. Bot.* 58, 3155–3169. doi: 10.1093/jxb/erm153

Scofield, G. N., Hirose, T., Gaudron, J. A., Furbank, R. T., Upadhyaya, N. M., and Ohsugi, R. (2002). Antisense suppression of the rice transporter gene, OsSUT1, leads to impaired grain filling and germination but does not affect photosynthesis. *Funct. Plant Biol.* 29, 815–826. doi: 10.1071/PP01204

Siao, W., Chen, J. Y., Hsiao, H. H., Chung, P., and Wang, S. J. (2011). Characterization of osSUT2 expression and regulation in germinating embryos of rice seeds. *Rice* 4, 39–49. doi: 10.1007/s12284-011-9063-1

Singh, J., Das, S., Jagadis Gupta, K., Ranjan, A., Foyer, C. H., and Thakur, J. K. (2023). Physiological implications of SWEETs in plants and their potential applications in improving source-sink relationships for enhanced yield. *Plant Biotechnol. J.* 21, 1528–1541. doi: 10.1111/pbi.13982

Singh, J., James, D., Murali Achary, V. M., Patel, M. K., Thakur, J. K., Reddy, M., et al. (2021). Coordinated overexpression of OsSUT1, OsSWEET11 and OsSWEET14 in rice impairs carbohydrate metabolism that has implications in plant growth, yield and susceptibility to Xanthomonas oryzae pv oryzae (Xoo). *bioRxiv* 2021. doi: 10.1101/2021.01.07.425507v1.abstract

Song, W., Hao, Q., Cai, M., Wang, Y., Zhu, X., Liu, X., et al. (2020). Rice OsBT1 regulates seed dormancy through the glycometabolism pathway. *Plant Physiol. bioch.* 151, 469–476. doi: 10.1016/j.plaphy.2020.03.055

Sosso, D., Luo, D., Li, Q. B., Sasse, J., Yang, J., Gendrot, G., et al. (2015). Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. *Nat. Genet.* 47, 1489–1493. doi: 10.1038/ng.3422

Spengler, B. (2015). Mass spectrometry imaging of biomolecular information. *Anal. Chem.* 87, 64–82. doi: 10.1021/ac504543v

Srivastava, A. C., Dasgupta, K., Ajieren, E., Costilla, G., McGarry, R. C., and Ayre, B. G. (2009). Arabidopsis plants harbouring a mutation in AtSUC2, encoding the predominant sucrose/proton symporter necessary for efficient phloem transport, are able to complete their life cycle and produce viable seed. *Ann. Bot.* 104, 1121–1128. doi: 10.1093/aob/mcp215

Srivastava, A. C., Ganesan, S., Ismail, I. O., and Ayre, B. G. (2008). Functional characterization of the Arabidopsis AtSUC2 Sucrose/H+ symporter by tissue-specific complementation reveals an essential role in phloem loading but not in long-distance transport. *Plant Physiol.* 148, 200–211. doi: 10.1104/pp.108.124776

Sun, L., Deng, R., Liu, J., Lai, M., Wu, J., Liu, X., et al. (2022). An overview of sucrose transporter (SUT) genes family in rice. *Mol. Biol. Rep.* 49, 5685–5695. doi: 10.1007/s11033-022-07611-x

Sun, Y., Reinders, A., LaFleur, K. R., Mori, T., and Ward, J. M. (2010). Transport activity of rice sucrose transporters OsSUT1 and OsSUT5. *Plant Cell Physiol.* 51, 114–122. doi: 10.1093/pcp/pcp172

Sun, C., Wang, Y., Yang, X., Tang, L., Wan, C., Liu, J., et al. (2023). MATE transporter GFD1 cooperates with sugar transporters, mediates carbohydrate partitioning and controls grain-filling duration, grain size and number in rice. *Plant Biotechnol. J.* 21, 621–634. doi: 10.1111/pbi.13976

Sun, A. J., Xu, H. L., Gong, W. K., Zhai, H. L., Meng, K., Wang, Y. Q., et al. (2008). Cloning and expression analysis of rice sucrose transporter genes OsSUT2M and OsSUT5Z. *J. Integr. Plant Biol.* 50, 62–75. doi: 10.1111/j.1744-7909.2007.00596.x

Sun, C., Zhang, K., Zhou, Y., Xiang, L., He, C., Zhong, C., et al. (2021). Dual function of clock component OsLHY sets critical day length for photoperiodic flowering in rice. *Plant Biotechnol. J.* 19, 1644–1657. doi: 10.1111/pbi.13580

Takanashi, K., Shitan, N., and Yazaki, K. (2014). The multidrug and toxic compound extrusion (MATE) family in plants. *Plant Biotechnol.* 31, 417–430. doi: 10.5511/plantbiotechnology.14.0904a

Upadhyay, N., Kar, D., Deepak Mahajan, B., Nanda, S., Rahiman, R., Panchakshari, N., et al. (2019). The multitasking abilities of MATE transporters in plants. *J. Exp. Bot.* 70, 4643–4656. doi: 10.1093/jxb/erz246

van Bel, A. J. (1996). Interaction between sieve element and companion cell and the consequences for photoassimilate distribution. Two structural hardware frames with associated physiological software packages in dicotyledons? *J. Exp. Bot.* 47, 1129–1140. doi: 10.1093/jxb/47.Special_Issue.1129

Vandeputte, G., and Delcour, J. A. (2004). From sucrose to starch granule to starch physical behaviour: a focus on rice starch. *Carbohydr. Polym.* 58, 245–266. doi: 10.1016/j.carbpol.2004.06.003

Wang, T., He, W., Li, X., Zhang, C., He, H., Yuan, Q., et al. (2023). A rice variation map derived from 10 548 rice accessions reveals the importance of rare variants. *Nucleic Acids Res.* 51, gkad840. doi: 10.1093/nar/gkad840

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, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat. Genet.* 40, 1370–1374. doi: 10.1038/ng.220

Wang, G., Wu, Y., Ma, L., Lin, Y., Hu, Y., Li, M., et al. (2021). Phloem loading in rice leaves depends strongly on the apoplastic pathway. *J. Exp. Bot.* 72, 3723–3738. doi: 10.1093/jxb/erab085

Wen, S., Neuhaus, H. E., Cheng, J., and Bie, Z. (2022). Contributions of sugar transporters to crop yield and fruit quality. *J. Exp. Bot.* 73, 2275–2289. doi: 10.1093/jxb/erac043

White, P. J., and Ding, G. (2023). "Long-distance transport in the xylem and phloem," in *Marschner's Mineral Nutrition of Plants* (Amsterdam: Elsevier), 73–104.

Williams, L. E., Lemoine, R., and Sauer, N. (2000). Sugar transporters in higher plants–a diversity of roles and complex regulation. *Trends Plant Sci.* 5, 283–290. doi: 10.1016/s1360-1385(00)01681-2

Wipf, D., Pfister, C., Mounier, A., Leborgne-Castel, N., Frommer, W. B., and Courty, P. E. (2021). Identification of putative interactors of arabidopsis sugar transporters. *Trends Plant Sci.* 26, 13–22. doi: 10.1016/j.tplants.2020.09.009

Wirth, J., Chopin, F., Santoni, V., Viennois, G., Tillard, P., Krapp, A., et al. (2007). Regulation of root nitrate uptake at the NRT2.1 protein level in Arabidopsis thaliana. *J. Biol. Chem.* 282, 23541–23552. doi: 10.1074/jbc.M700901200

Wu, Y., Lee, S. K., Yoo, Y., Wei, J., Kwon, S. Y., Lee, S. W., et al. (2018). Rice transcription factor osDOF11 modulates sugar transport by promoting expression of sucrose transporter and SWEET genes. *Mol. Plant* 11, 833–845. doi: 10.1016/j.molp.2018.04.002

Xiong, Y., Ren, Y., Li, W., Wu, F., Yang, W., Huang, X., et al. (2019). NF-YC12 is a key multi-functional regulator of accumulation of seed storage substances in rice. *J. Exp. Bot.* 70, 3765–3780. doi: 10.1093/jxb/erz168

Xu, Q., Yin, S., Ma, Y., Song, M., Song, Y., Mu, S., et al. (2020). Carbon export from leaves is controlled via ubiquitination and phosphorylation of sucrose transporter SUC2. *Proc. Natl. Acad. Sci. U.S.A.* 117, 6223–6230. doi: 10.1073/pnas.1912754117

Yanagisawa, S., and Sheen, J. (1998). Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. *Plant Cell* 10, 75–89. doi: 10.1105/tpc.10.1.75

Yang, J., Luo, D., Yang, B., Frommer, W. B., and Eom, J. S. (2018). SWEET11 and 15 as key players in seed filling in rice. *New Phytol.* 218, 604–615. doi: 10.1111/nph.15004

Yong, Z., Kotur, Z., and Glass, A. D. (2010). Characterization of an intact two-component high-affinity nitrate transporter from Arabidopsis roots. *Plant J.* 63, 739–748. doi: 10.1111/j.1365-313X.2010.04278.x

Yuan, M., Chu, Z., Li, X., Xu, C., and Wang, S. (2009). Pathogen-induced expressional loss of function is the key factor in race-specific bacterial resistance conferred by a recessive R gene xa13 in rice. *Plant Cell Physiol.* 50, 947–955. doi: 10.1093/pcp/pcp046

Yuan, M., and Wang, S. (2013). Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms. *Mol. Plant* 6, 665–674. doi: 10.1093/mp/sst035

Zhang, C., and Turgeon, R. (2018). Mechanisms of phloem loading. *Curr. Opin. Plant Biol.* 43, 71–75. doi: 10.1016/j.pbi.2018.01.009

Zhang, Y.-M., Yu, H.-X., Ye, W.-W., Shan, J.-X., Dong, N.-Q., Guo, T., et al. (2021b). A rice QTL GS3. 1 regulates grain size through metabolic-flux distribution between flavonoid and lignin metabolons without affecting stress tolerance. *Commun. Biol.* 4, 1171. doi: 10.1038/s42003-021-02686-x

Zhang, S., Zhang, Y., Li, K., Yan, M., Zhang, J., Yu, M., et al. (2021a). Nitrogen mediates flowering time and nitrogen use efficiency via floral regulators in rice. *Curr. Biol.* 31, 671–683 e675. doi: 10.1016/j.cub.2020.10.095

Zhao, Y., Hu, J., Zhang, Y., Tao, H., Li, L., He, Y., et al. (2023). Unveiling targeted spatial metabolome of rice seed at the dough stage using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry imaging. *Food Res. Int.* 174, 113578. doi: 10.1016/j.foodres.2023.113578

Zhou, C., Lin, Q., Ren, Y., Lan, J., Miao, R., Feng, M., et al. (2023). A CYP78As–SMG4–COPII pathway promotes grain size in rice. *Plant Cell* 35, 4325. doi: 10.1093/plcell/koad239

Zhou, M., Deng, X., Jiang, Y., Zhou, G., and Chen, J. (2023). Genome-wide identification and an evolution analysis of tonoplast Monosaccharide Transporter (TMT) genes in seven gramineae crops and their expression profiling in rice. *Genes* 14, 1140. doi: 10.3390/genes14061140

Zhou, Y., Liu, L., Huang, W., Yuan, M., Zhou, F., Li, X., et al. (2014). Overexpression of OsSWEET5 in rice causes growth retardation and precocious senescence. *PloS One* 9, e94210. doi: 10.1371/journal.pone.0094210

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 September 2024
DOI 10.3389/fpls.2024.1441555

OPEN ACCESS

EDITED BY
Ming Luo,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

REVIEWED BY
Zitong Li,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
Chenwu Xu,
Yangzhou University, China

*CORRESPONDENCE
Hai Lan
lanhai_maize@163.com
Hailan Liu
lhlzju@hotmail.com

†These authors have contributed equally to this work

RECEIVED 31 May 2024
ACCEPTED 21 August 2024
PUBLISHED 09 September 2024

CITATION
Xiang Y, Xia C, Li L, Wei R, Rong T, Liu H and Lan H (2024) Genomic prediction of yield-related traits and genome-based establishment of heterotic pattern in maize hybrid breeding of Southwest China. *Front. Plant Sci.* 15:1441555. doi: 10.3389/fpls.2024.1441555

Genomic prediction of yield-related traits and genome-based establishment of heterotic pattern in maize hybrid breeding of Southwest China

Yong Xiang†, Chao Xia†, Lujiang Li, Rujun Wei, Tingzhao Rong, Hailan Liu* and Hai Lan*

Maize Research Institute/State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu, Sichuan, China

When genomic prediction is implemented in breeding maize (*Zea mays* L.), it can accelerate the breeding process and reduce cost to a large extent. In this study, 11 yield-related traits of maize were used to evaluate four genomic prediction methods including rrBLUP, HEBLP|A, RF, and LightGBM. In all the 11 traits, rrBLUP had similar predictive accuracy to HEBLP|A, and so did RF to LightGBM, but rrBLUP and HEBLP|A outperformed RF and LightGBM in 8 traits. Furthermore, genomic prediction-based heterotic pattern of yield was established based on 64620 crosses of maize in Southwest China, and the result showed that one of the parent lines of the top 5% crosses came from temp-tropic or tropic germplasm, which is highly consistent with the actual situation in breeding, and that heterotic pattern (Reid+ × Suwan+) will be a major heterotic pattern of Southwest China in the future.

KEYWORDS

genomic prediction, maize, yield-related traits, heterotic pattern of yield, Southwest China

Introduction

Maize (*Zea mays* L.), one of the most important cereal crops throughout the world, has been widely used as food, biofuel, feed and raw materials of many industrial products (Yang et al., 2011a). In order to increase maize production, molecular marker techniques such as RFLP, SSR, and SNP have been developed to improve agronomic traits of economic importance (Stevens, 2008; Prasanna et al., 2010; Inghelandt et al., 2010). However, conventional marker-assisted selection (MAS) can only utilize the markers tightly linked

to large or moderate-effect QTL and fails to deal with quantitative traits controlled by minor-polygene (Heffner et al., 2009; Jannink et al., 2010; Chai et al., 2012; Hao et al., 2014).

Meuwissen et al. (2001) pioneered the technique of genomic prediction to solve this problem. First, training population with genotypic and phenotypic information is utilized to compute genetic effect of each marker at the level of genome-wide markers, and then the genomic estimated breeding values (GEBV) of each individual in candidate population with only genotypic information are obtained. The fact that all marker information is included in genomic prediction contributes to higher predictive accuracy and genetic gain. Applied successfully in the breeding programs of animals and plants such as dairy cattle, beef cattle, pigs, sheep, maize, wheat, and rice (Hayes et al., 2009; Heffner et al., 2011; Meuwissen et al., 2013; Fristche-Neto et al., 2018; Cui et al., 2020; Alemu et al., 2024), many genomic prediction methods have been developed and can be categorized as (1) parametric methods such as Genomic Best Linear Unbiased Prediction (GBLUP), ridge regression Best Linear Unbiased Prediction (rrBLUP), Bayes, Haseman-Elston regression Best Linear Prediction (HEBLP), and RHPP (Meuwissen et al., 2001; VanRaden, 2008; Endelman, 2011; Habier et al., 2011; Xu et al., 2014; Liu and Chen, 2017; Liu and Yu, 2023), and (2) non-parametric methods such as Random Forest (RF), Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM) and deep learning (DL) (Ogutu et al., 2011; Montesinos-López et al., 2018; Bellot et al., 2018; Yan et al., 2021).

Maize is the first crop to apply genomic prediction (Bernardo and Yu, 2007). Riedelsheimer et al. (2012) used 56110 SNPs to predict combining abilities of 7 biomass- and bioenergy-related traits of maize and found that prediction accuracies ranged from 0.72 to 0.81. Zhao et al. (2012) analyzed European maize elite breeding population and found that prediction accuracies of grain moisture and grain yield were respectively 0.90 and 0.58. Navarro et al. (2017) used 30000 markers to predict flower time traits including days to anthesis (DA) and days to silking (DS) for 4471 maize landraces via rrBLUP and found that the average prediction accuracy was 0.45. Li et al. (2020) utilized 8 genomic prediction methods to predict 10 traits of maize and found that the prediction accuracy ranged from 0.382 to 0.795. Xiao et al. (2021) utilized 8652 F1 hybrids from maize CUBIC population to predict days to tasseling (DTT), plant height (PH) and ear weight (EW) and found that the prediction accuracies of DTT, PH, and EW were 0.76, 0.81, and 0.66 respectively. Studies above demonstrated that genomic prediction was a highly effective technique to improve maize.

Southwest China is one of the three main production areas of maize in the country. In order to increase yield and adapt to complex ecological condition in this region, the breeders introduced tropical or subtropical maize germplasm such as ETO, Suwan, Dentado Amarillo, Tuson and Tuxpeno into the local maize germplasm to broaden the narrow genetic base (Zhang et al., 2016; Leng et al., 2019). Furthermore, a number of inbred lines containing tropical or subtropical maize germplasm were used to obtain excellent hybrid. In this study, we generated 2077 hybrids derived from random crosses of 360 inbred lines to perform genomic prediction analysis of 11 yield-related traits and furthermore established genomic prediction-based heterotic pattern of yield.

Materials and methods

Phenotypic data collection of maize

In this study, 360 maize inbred lines of wide genetic background and origin were collected, including the most representative inbred lines in Southwest China, inbred lines exchanged from the other breeding units of Sichuan, temperate inbred lines introduced from North China, tropic inbred lines introduced from the other breeding units, and inbred lines cultivated by our institute (Supplementary Table S1). From 64,620 crosses obtained through pairing of the 360 inbred lines, 2077 were randomly selected to perform field evaluation. All hybrids were planted in Field data of 11 traits including row number per ear, ear length, ear diameter, grain number per row, grain length, grain width, grain thickness, hundred grain weight, weight per unit volume, and yield (kg/mu) for 2077 crosses were collected at Gasa town, Jinghong city (21°95′N, 100°75′E, 520 meters above sea level), Yunnan province, China, in 2019. All hybrids were planted in one-row plot using a complete block design. Ten plants were planted in each row. The row length was 2.5 m and row spacing was 0.8 m. The plant density was about 3300 plants/mu. Field management followed standard procedures. For quality control, five plants in the middle of each row were selected to obtain the phenotypes. No further phenotypic corrections were performed on the basis of the other hybrid.

Genotypic data collection of maize inbred lines

Genomic DNA of the 360 inbred lines were extracted via CTAB method, and their qualities were evaluated with a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific). High-quality DNA was genotyped via 48K liquid phase gene chip and sequenced on an Illumina Nova 150-bp paired-end sequencing platform in China Golden Marker (Beijing, China). Quality control of raw reads was performed via Trimmomatic-0.36 (Lohse et al., 2012) with default parameter, and then the filtered data were mapped to B73_RefGen_v4 genome with BWA software (Li and Durbin, 2009). The SNPs were called with GATK (McKenna et al., 2010) and further filtered with VCFtools (Danecek et al., 2011). The parameters used in SNP filters ration were: –minDP 4, –minGQ 10, –max-missing 0.8, –maf 0.05.

Construction of phylogenetic tree

45425 SNPs were used to generate distance matrix using VCF2Dis (https://github.com/BGI-shenzhen/VCF2Dis) and then phylogenetic tree was constructed based on distance matrix via FastME2.1.6.4 (Lefort et al., 2015).

Genotyping of maize hybrids

The genotypes of 2077 hybrid were obtained based on those of the inbred lines. 33009 of 45425 SNPs were kept according to the

following criteria: (1) there are at least two kinds of genotype at each locus; (2) the percentage of "NA" at each locus is not more than 0.3; (3) the percentage of rare alleles at each locus is not less than 0.1.

Estimation of SNP heritability

The genetic model of a quantitative trait can be written as

$$y = Xb + Zu + e$$

where y is an $n \times 1$ vector for the phenotypic values; X is a $n \times p$ incidence matrix for fixed effects; b is a $p \times 1$ vector of fixed effects; p is the number of fixed effects; Z is $n \times m$ genotype matrix; m is the number of markers; u is a $m \times 1$ vector of additive effects following $N(0, I\sigma_u^2)$; e is a $n \times 1$ vector of residual error following $N(0, I\sigma_e^2)$. The SNP heritability was calculated as:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

We estimated σ_g^2 and σ_e^2 via the program GCTA v1.94.1, which consists of two steps: (1) computing the genetic relationship matrix (GRM) of all individuals based on all SNPs; (2) estimating additive genetic variance of a trait based on above GRM via the restricted maximum likelihood (REML) (Yang et al., 2011b).

Evaluation of genomic prediction of 11 yield-related traits

Two parametric genomic prediction methods including rrBLUP (Endelman, 2011) and HEBLP|A (Liu and Chen, 2017) and two non-parametric genomic prediction methods including RF and LightGBM were used in this study.

rrBLUP is one of the first methods for genomic prediction and has been taken as a baseline model (Meuwissen et al., 2001). The formula of rrBLUP is written as the following:

$$y = \mu + Zu + e$$

in which y is the vector of the phenotype; μ is the vector of mean value; Z is the genotype matrix; u is the vector of additive effects of the causal loci following $N(0, I\sigma_u^2)$ where σ_u^2 is the additive genetic variance; e is the vector of residual error following $N(0, I\sigma_e^2)$. The BLUP solution for u is written as $\hat{u} = Z^T(ZZ^T + \lambda I)^{-1}(y - \hat{\mu})$ where $\lambda = \frac{\sigma_e^2}{\sigma_u^2}$. The ridge parameter λ is computed on the tenfold cross-validated partial likelihood that minimized the residuals between predicted and observed phenotypes within the training population.

HEBLP|A is highly efficient in computation due to its estimation of heritability via Haseman-Elston regression (Liu and Chen, 2017). HEBLP|A and rrBLUP have similar model, but are different in the strategy of computing λ. HEBLP|A uses IBS-based Haseman-Elston (HE) to estimate σ_u^2 and then $\lambda = \frac{1-\sigma_u^2}{\sigma_u^2}$.

RF is one of the most popular and powerful machine learning algorithms and has already been applied to a wide variety of genomic problems (Ogutu et al., 2011). RF uses bootstrap sample of the training population to build B decision trees and then averages the output of the ensemble of B trees to predict candidate individuals. The formula is as follows:

$$\hat{f}^B(x) = \frac{1}{B}\sum_{i=1}^{B} T(x, \psi_i)$$

in which ψ_i is the ith RF trees.

LightGBM, a very popular machine learning algorithm, was developed by Microsoft in 2017 and has been widely used in dealing with extremely large data with ultra-high efficiency (Ke et al., 2017). LightGBM uses a gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) to reduce computational time. The basic process of the LightGBM is described as follows (Mienye and Sun, 2022):

1. Merge mutually exclusive features x_i based on training population $S = (x_i, y_i)$ using the EFB.
2. Initialize $\Theta_0(x) = \text{argmin}_c \sum_i^n L(y_i, c)$ where $L(y, c)$ is the loss function.
3. For $t = 1, \ldots, T$:
   i. Compute the absolute values of gradients:
   $$r_i = \left[\frac{\delta L(y_i, \Theta(x_i))}{\delta \Theta(x_i)}\right]_{\Theta(x)=\Theta_{t-1}(x)}$$
   ii. Resample the training population using the GOSS: $D = A + B$ where A is samples with larger gradients and B is samples with small gradients.
   iii. Calculate information gain values.
   iv. Obtain a new decision tree $\Theta_t(x)'$ on D.
   v. Update $\Theta_t(x) = \Theta_{t-1}(x) + \Theta_t(x)'$.
4. Output: $\hat{\theta}(x) = \Theta_T(x)$.

For RF and LightGBM methods, we utilized "randomForest" and "LightGBM" function in R package with the default parameters. The correlation coefficient between the phenotypes and the predicted genotypic values was considered as the prediction accuracy. 10 replications were used to evaluate the prediction accuracy of genomic prediction methods.

Genomic prediction of yield of all crossing combinations

The 360 inbred lines generated 64620 crossing combinations in total. We utilized 2077 hybrids with the phenotypic and genotypic data to predict the yield performance of 64620 crossing combinations via HEBLP|A under additive model.

Results

Phylogenetic tree analysis of the 360 maize inbred lines

As was demonstrated in the phylogenetic tree, the 360 maize inbred lines fell into three groups including temperate germplasm (90 inbred lines), tropical germplasm (106 inbred lines), and temp-tropic germplasm (164 inbred lines) (Figure 1). The temperate group contains temperate inbred lines from North China such as Zheng58, Chang7-2,

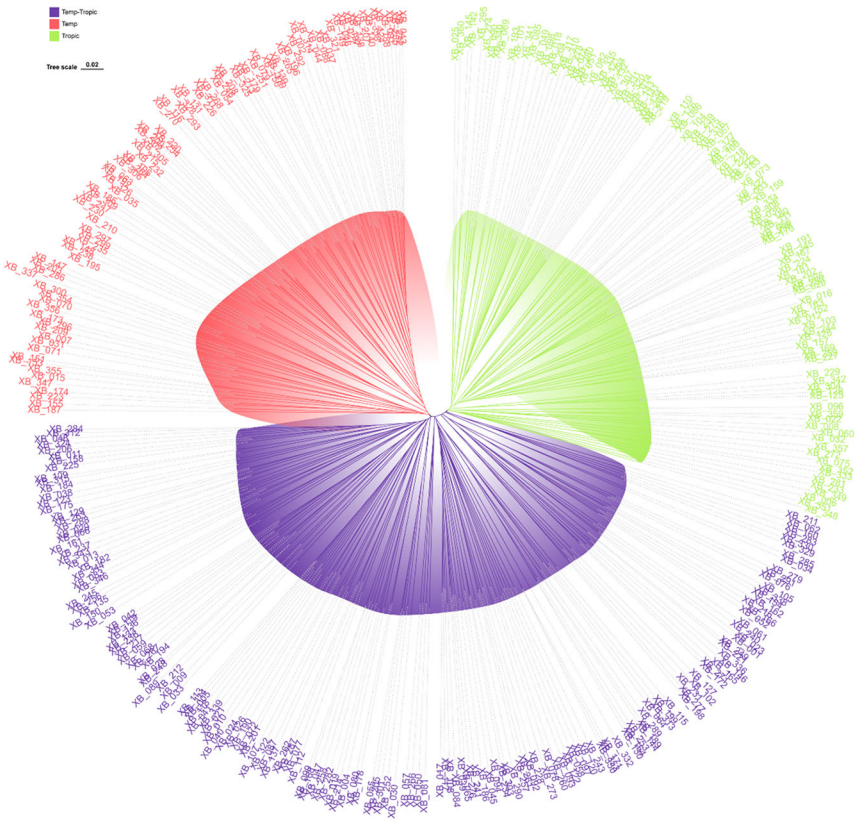


FIGURE 1
Phylogenetic tree for 360 maize inbred lines using 45425 SNP markers. Temperate germplasm (red), Temp-Tropic germplasm (purple), and Tropic germplasm (green).

Mo17, Dan340, PH6WC, PH4CV, and Jing724 and those from Southwest China such as Y9614 and 4011. Most of the inbred lines in the tropical group have mixed genetic background of Suwan, Tuxpeno, and CIMMTY, with the parent lines of main hybrids in Southwest China QR723, WG646, ZNC442, Xian21A, Rekangbai67, XZ50612, and 7031 being examples. As for temp-tropic group, most inbred lines such as SCML0849, CL11, R18, GH35, F19, SH1070, Q1 come from PB inbred lines and improved Reid inbred lines.

Evaluation of genomic prediction for the 11 yield-related traits in maize hybrids

The SNP heritability of the 11 yield-related traits of 2077 hybrids was estimated via GCTA analysis under the additive model and ranged from 0.058 for grain thickness to 0.883 for row number per ear (Figure 2). The result indicated that most traits were mainly controlled by additive effects.

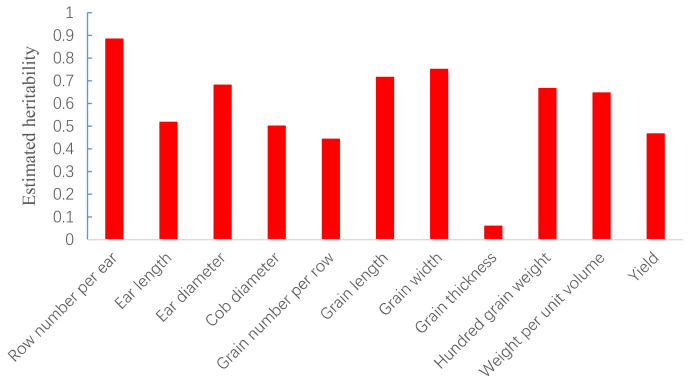


FIGURE 2
SNP-based heritability estimation of 11 traits in maize.

Evaluation of genomic prediction for the 11 yield-related traits was performed via rrBLUP, HEBLP|A, RF, and LightGBM (Table 1) with the size of training population and candidate population being 1800 and 277 respectively. The prediction accuracies ranged from 0.232 to 0.833 across traits and models. The row number per ear showed the highest prediction accuracy (average value across all methods being 0.796) and grain thickness had the lowest prediction accuracy (average value cross all methods being 0.302). The Pearson's correlation coefficient between heritability and average prediction accuracy was significant ($R = 0.929$ and $P < 0.001$), indicating that the heritability influenced prediction accuracy to a great extent, and the result was consistent with previous researches (Daetwyler et al., 2008; Liu and Chen, 2018). Among the four methods, rrBLUP and HEBLP|A have similar prediction accuracy in all the 11 traits (for example, $r_{rrBLUP} = 0.833 \pm 0.007$ and $r_{HEBLP|A} = 0.825 \pm 0.007$ in row number per ear), and so do RF and LightGBM (for example, $r_{RF} = 0.752 \pm 0.008$ and $r_{LightGBM} = 0.772 \pm 0.008$ in row number per ear). The parametric methods (rrBLUP and HEBLP|A) outperformed non-parametric methods (RF and LightGBM) in 8 traits (Row number per ear, Ear length, Ear diameter, Grain number per row, Grain length, Grain width, Hundred grain weight, and Weight per unit volume), and the latter outperformed the former only in grain thickness, which may be mainly controlled by non-additive effects such as dominance, epistasis and genotype-by-environment interaction. Both two categories of methods had similar predictive accuracy in 2 traits (Cob diameter and Yield).

Prediction of GEBVs of yield for all crossing combinations

GEBVs for yield of the 64620 crosses generated by the 360 inbred lines were obtained via 2077 hybrids with phenotypic and genotypic values under additive model via HEBLP|A and were sorted in descending order. The average GEBV of all crosses was 341.50 for yield, and that of the top 200 crosses, the top 5% (3231 ones) and the top 10% (6462 ones) was 558.35, 491.78 and 469.20 respectively. If breeders use the crosses of the best predictive accuracy in cultivation (the top 10% for example), a significant genetic gain (469.20 – 341.50 = 127.7) can be expected.

To evaluate the general combining ability (GCA) of an inbred line, its mean GEBV of the 359 crosses was obtained by crossing the remaining 359 ones with it, and consequently there were 44 inbred lines with mean GEBVs over 400kg, among which some most popular inbred lines in Southwest China were found, such as ZNC442, QR273, YA8201, Xian21A, Rekangbai67, XZ50612, and F19. There are 29 tropic inbred lines (proportion is 65.9%) and 11 temp-tropic ones (proportion is 25%) among the 44 inbred lines, indicating that inbred lines with tropic and temp-tropic germplasm play an important role in the maize hybrid breeding of Southwest China. In contrast, the mean GEBVs of some pure temperate inbred lines introduced from North China such as 478, Zheng58, Chang7-2, Dan340, 91227, M54, and NH60 were quite low, but their improved version demonstrated high GCA when tropic germplasm was added. For example, comparing the inbred line 91227 with LX2715, its improved version via introducing tropic germplasm, we found a significant increase of the mean GEBV from 297.2kg to 349.3kg. Therefore, tropic or temp-tropic germplasms can help pure temperate inbred lines from North China to adapt to ecological environment, reducing their defects in Southwest China such as low resistance to diseases and pests when they are used in maize breeding. Generally speaking, most of the local temperate inbred lines from Southwest China contain a certain proportion of tropic germplasm through long-term improvement. In addition, the result of some representative inbred lines in Southwest China including PH6WC (temperate), GH35, SCML0849, and NG5 (temp-tropic), and WG646, ZNC442, XZ50612, QR273, and XL8242 (tropic) has showed that tropic inbred lines have better performance than temperate and temp-tropic ones (Figure 3). In particular, the mean

TABLE 1 Comparison of genomic predictability among rrBLUP, HEBLP|A, RF, and LightGBM for 11 yield related traits of maize based on 10 simulations.

Trait	Training	Candidate	rrBLUP	HEBLP|A	RF	LightGBM
Row number per ear	1800	277	0.833 ± 0.007	0.825 ± 0.007	0.752 ± 0.008	0.772 ± 0.008
Ear length	1800	277	0.627 ± 0.026	0.628 ± 0.025	0.584 ± 0.020	0.585 ± 0.024
Ear diameter	1800	277	0.749 ± 0.004	0.727 ± 0.006	0.693 ± 0.007	0.712 ± 0.006
Cob diameter	1800	277	0.679 ± 0.013	0.653 ± 0.016	0.641 ± 0.010	0.650 ± 0.010
Grain number per row	1800	277	0.561 ± 0.012	0.561 ± 0.013	0.509 ± 0.020	0.495 ± 0.015
Grain length	1800	277	0.677 ± 0.009	0.670 ± 0.009	0.599 ± 0.010	0.613 ± 0.010
Grain width	1800	277	0.767 ± 0.009	0.769 ± 0.009	0.675 ± 0.009	0.706 ± 0.007
Grain thickness	1800	277	0.250 ± 0.045	0.232 ± 0.042	0.382 ± 0.055	0.342 ± 0.062
Hundred grain weight	1800	277	0.742 ± 0.011	0.726 ± 0.013	0.649 ± 0.06	0.683 ± 0.010
Weight per unit volume	1800	276	0.783 ± 0.008	0.763 ± 0.009	0.714 ± 0.012	0.747 ± 0.007
Yield	1800	264	0.680 ± 0.006	0.653 ± 0.007	0.650 ± 0.012	0.650 ± 0.007

The correlation coefficient between the phenotypes and the predicted genotypic values was considered as the prediction accuracy. The values after ± represent the corresponding standard error.

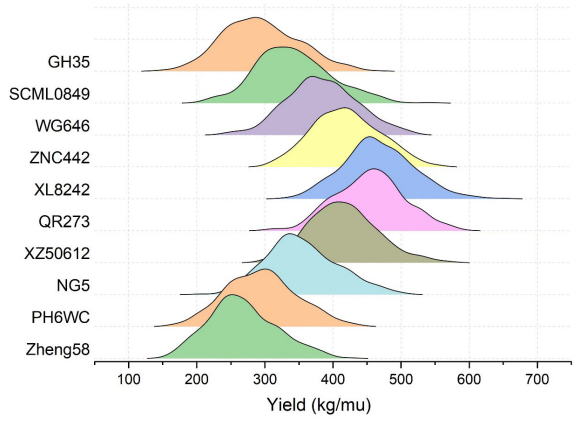


FIGURE 3
The GEBV distribution of the 359 crosses corresponding the representative inbred lines in Southwest China. Zheng58 (temperate line from North China) was taken as control inbred line. Temperate inbred line in Southwest China (PH6WC), temp-tropic inbred lines in Southwest China (GH35, SCML0849, and NG5), and tropic inbred lines in Southwest China (WG646, ZNC442, XZ50612, QR273, and XL8242).

GEBV of the 359 crosses corresponding to XL8242 (a newest tropic inbred line cultivated by our unit) ranked the first among the 360 inbred lines and has been used to breed some hybrids successfully (Table 2).

We constructed six heterotic patterns with three germplasms (Figure 4) and the results showed that the mean GEBV of Tropic × Tropic, Tropic × Temp-Tropic, Tropic × Temp, Temp-Tropic × Temp-Tropic, Temp-Tropic × Temp and Temp × Temp is 391.9 ± 45.0, 366.4 ± 46.5, 353.4 ± 45.6, 327.0 ± 42.2, 312.6 ± 41.8, and 290.1 ± 41.6 respectively (Figure 5). Among the 5% top crosses, 21% are Tropic × Tropic mode and 66.4% are Temperate (or Temp-Tropic) × Tropic mode, indicating that Temperate (or Temp-Tropic) × Tropic mode is a major heterotic pattern. There are some approved commercial varieties among the 5% top crosses including Youdi899, Chuanqing9, Yayu68, Rongyuqingzhu1, Shidi6, Qixiangqingzhu6, Rongkefengzan, Nonghua606, Chuandan99, and Ronghe99 (Table 2). Take Youdi899 for example, deriving from SCML0849 and XL8242, it ranked 184 in all crosses and had an excellent performance in Southwest China (Figure 6). In 2020-2022 years' national spring maize cultivar regional trials (middle and high altitude region) in Southwest China, its mean yield of two years is 814.3kg and the production increased by 14.8% compared with the control variety. In 2021-2022 years' autonomous production test, its mean yield was 783.1kg and the production increased by 11.6% compared with the control variety. After the identification of field inoculation diseases, this variety was highly resistant to northern maize leaf blight and southern maize leaf blight, moderately resistant to ear rot, maize stalk rot, grey leaf spot, southern maize rust, and susceptible to northern maize leaf blight. Tropic × Tropic mode account for relative high proportion because Jinghong City of Yunnan Province, the place where the training population of maize was planted, belongs to tropic-subtropic region, and the result conforms well to practical breeding. In the future, we will further investigate heterotic mode of maize in other regions of Southwest China (including middle and low

TABLE 2 Basic information of the maize varieties from 5% top crosses based on genomic prediction of yield.

Predictive rank	Name of variety	Approval number	Male parent	Female parent	Heterotic pattern
184	Youdi899	National test20231073	SCML0849	XL8242	Tropic × Temp-Tropic
249	Chuanqing9	Sichuan test20222061	LX2715	XL8242	Tropic × Temp-Tropic
1294	Yayu68	Yunnan test201105	F06	YA8201	Tropic × Temp-Tropic
1352	Rongyuqingzhu1	National test20180178	SCML5409	Xian21A	Tropic × Temp-Tropic
1752	Shidi6	Yunnan test2022102	Y9614	XB_320	Tropic × Temp
2010	Qixiangqingzhu6	Yunnan test2023161	XB_267	XL8242	Tropic × Temp-Tropic
2139	Rongkefengzan	Yunnan test2022086	NG0715	XL8242	Tropic × Temp
2255	Nonghua606	Guangxi test2017004	Y0921	ZNC442	Tropic × Temp-Tropic
2520	Chuandan99	National test20210096	ZNC442	SCML0849	Tropic × Temp-Tropic
2525	Ronghe99	National test20220514	SCML0849	XL2142	Tropic × Temp-Tropic

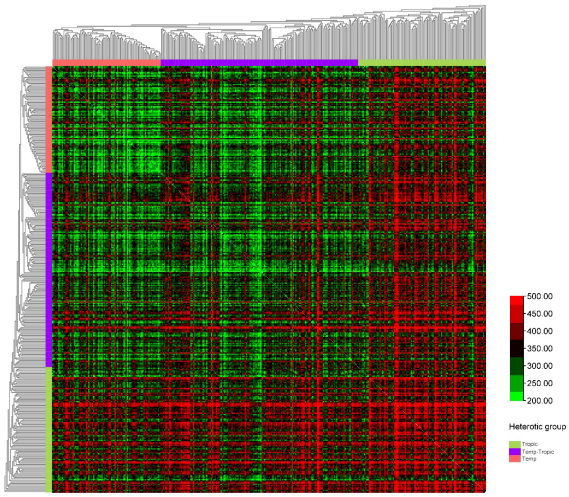


FIGURE 4
Heterotic pattern in maize hybrid breeding of Southwest China based on genomic prediction of yield (kg/mu).

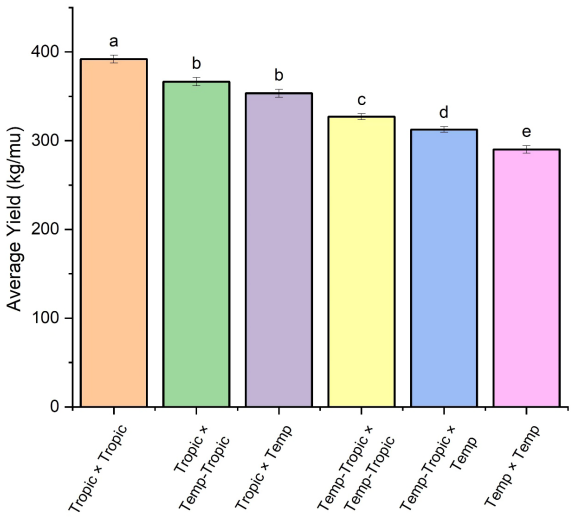


FIGURE 5
Average yield of different heterotic groups based on genomic prediction. The lowercase letters (a, b, c, d, and e) have significant difference.

FIGURE 6
Phenotype of maize variety Youdi899. **(A)** Plant architecture. **(B)** Ear.

altitude region and middle and high altitude region) via genomic prediction.

Discussion

It is critical for modern plant breeding to predict superior individuals with high accuracy (Heffner et al., 2009; Voss-Fels et al., 2019). When there is a large number of inbred lines in the process of maize breeding, it is difficult to identify superior hybrid crosses from $\frac{N*(N-1)}{2}$ crosses (N represents number of inbred lines) via conventional breeding methods because it is time consuming and too expensive. Being a useful technique to select superior hybrid crosses, genomic prediction has been successfully applied to maize breeding currently (Albrecht et al., 2011; Zhao et al., 2012; Riedelsheimer et al., 2012; Technow et al., 2014; Li et al., 2020; Wang et al., 2020; Luo et al., 2023).

To our knowledge, this is the first study about a large-scale application of genomic prediction in Southwest China's maize hybrid breeding. We evaluated the genomic prediction accuracies of 11 yield-related traits in an F1 population with 2077 F1 maize hybrids and found that the predictive accuracies of most traits except grain thickness were higher than 55%. It is indicated that genomic prediction is a promising tool in maize breeding programs. Comparison of four methods demonstrated that parametric ones (rrBLUP and HEBLP|A) outperformed non-parametric ones (RF and LightGBM) in 8 out of 11 traits, which is consistent with previous studies that parametric methods outperformed machine-learning methods when the traits were mainly controlled by additive effects (Abdollahi-Arpanahi et al., 2020; Alves et al., 2020; Yu et al., 2023).

Increasing yield is the most important objective in hybrid maize production (Li et al., 2011; Peng et al., 2011; Tian et al., 2024), but it is difficult to improve yield trait via phenotypic selection or conventional MAS because it is a complex trait controlled by a large number of quantitative trait loci (QTL) with small effects (Ndlovu et al., 2022). In this study, the predictive accuracy of yield was about 65% when four genomic prediction methods were evaluated with 2077 maize crosses. According to the prediction of the yield performance of all 64620 crosses, one parent of the top 5% crosses has at least a tropic or temp-tropic inbred lines, and 86.4% of those top 5% have at least a tropic inbred line, which is highly consistent with practical breeding. It is indicated that tropic inbred lines play an important role in selecting superior hybrids in Southwest China.

With serious increase of diseases and pests of maize at present, it is more and more difficult for Reid × Suwan, the conventional heterotic pattern of Southwest China (Ni et al., 1996; Pan et al., 2020) to tackle the problem. Therefore, we proposed a new heterotic pattern (Reid+ × Suwan+) in which Reid+ represents the introgression from tropical germplasm or other temperate germplasm to Reid inbred lines, and Suwan+ represents the introgression from other tropical germplasm to Suwan inbred lines. It will become a major heterotic pattern of Southwest China in the future.

Data availability statement

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding authors.

Author contributions

HLL: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. YX: Investigation, Writing – review & editing. CX: Data curation, Formal analysis, Funding acquisition, Investigation, Writing – review & editing. LL: Data curation, Resources, Writing – review & editing. RW: Resources, Writing – review & editing. TR: Writing – review & editing. HL: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by National Key Research and Development Project of China (Grant No. 2023YFD1201101), Major Science and Technology Project of Sichuan Province (Grant No. 2022ZDZX0013), the National Natural Science Foundation of China (Grant No. 32271984 and 32101673), the Natural Science Foundation of Sichuan Province (Grant No. 2023NSFSC0222 and 2022NSFSC1774), and Science and Technology Program of Liangshan Prefecture (Grant No. 24YYYJ0183 and 24YYYJ0184).

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.2024.1441555/full#supplementary-material

References

Abdollahi-Arpanahi, R., Gianola, D., and Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Selection Evol.* 52, 12. doi: 10.1186/s12711-020-00531-z

Albrecht, T., Wimmer, V., Auinger, H. J., Erbe, M., Knaak, C., Ouzunova, M., et al. (2011). Genome-based prediction of testcross values in maize. *Theor. Appl. Genet.* 123, 339–350. doi: 10.1007/s00122-011-1587-7

Alemu, A., Åstrand, J., Montesinos-López, O. A., y Sánchez, J. I., Fernández-Gónzalez, J., Tadesse, W., et al. (2024). Genomic selection in plant breeding: Key factors shaping two decades of progress. *Mol. Plant* 17, 552–578. doi: 10.1016/j.molp.2024.03.007

Alves, A. A. C., da Costa, R. M., Bresolin, T., Júnior, G. A. F., Espigolan, R., Ribeiro, A. M. F., et al. (2020). Genome-wide prediction for complex traits under the presence of dominance effects in simulated populations using GBLUP and machine learning methods. *J. Anim. Sci.* 98, skaa179. doi: 10.1093/jas/skaa179

Bellot, P., de los Campos, G., and Pérez-Enciso, M. (2018). Can deep learning improve genomic prediction of complex human traits. *Genetics* 210, 809–819. doi: 10.1534/genetics.118.301298

Bernardo, R., and Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47, 1082–1090. doi: 10.2135/cropsci2006.11.0690

Chai, Y., Hao, X., Yang, X., Allen, W. B., Li, J., Yan, J., et al. (2012). Validation of DGAT1-2 polymorphisms associated with oil content and development of functional markers for molecular breeding of high-oil maize. *Mol. Breed.* 29, 939–949. doi: 10.1007/s11032-011-9644-0

Cui, Y., Li, R., Li, G., Zhang, F., Zhu, T., Zhang, Q., et al. (2020).). Hybrid breeding of rice via genomic selection. *Plant Biotechnol. J.* 18, 57–67. doi: 10.1111/pbi.13170

Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PloS One* 3, e3395. doi: 10.1371/journal.pone.0003395

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Fristche-Neto, R., Akdemir, D., and Jannink, J. L. (2018). Accuracy of genomic selection to predict maize single-crosses obtained through different mating designs. *Theor. Appl. Genet.* 131, 1153–1162. doi: 10.1007/s00122-018-3068-8

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinf.* 12, 186. doi: 10.1186/1471-2105-12-186

Hao, X., Li, X., Yang, X., and Li, J. (2014). Transferring a major QTL for oil content using marker-assisted backcrossing into an elite hybrid to increase the oil content in maize. *Mol. Breed.* 34, 739–748. doi: 10.1007/s11032-014-0071-x

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92, 433–443. doi: 10.3168/jds.2008-1646

Heffner, E. L., Jannink, J. L., and Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4, 65–75. doi: 10.3835/plantgenome2010.12.0029

Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512

Inghelandt, D. V., Melchinger, A. E., Lebreton, C., and Stich, B. (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* 120, 1289–1299. doi: 10.1007/s00122-009-1256-2

Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings Funct. Genomics* 9, 166–177. doi: 10.1093/bfgp/elq001

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). LightGBM: a highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* 30, 3149–3157.

Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* 32, 2798–2800. doi: 10.1093/molbev/msv150

Leng, Y., Lv, C., Li, L., Xiang, Y., Xia, C., Wei, R., et al. (2019). Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). *Mol. Breed.* 39, 38. doi: 10.1007/s11032-019-0946-y

Li, G., Dong, Y., Zhao, Y., Tian, X., Würschum, T., Xue, J., et al. (2020). Genome-wide prediction in a hybrid maize population adapted to Northwest China. *Crop J.* 8, 830–842. doi: 10.1016/j.cj.2020.04.006

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, J. Z., Zhang, Z. W., Li, Y. L., Wang, Q. L., and Zhou, Y. G. (2011). QTL consistency and meta-analysis for grain yield components in three generations in maize. *Theor. Appl. Genet.* 122, 771–782. doi: 10.1007/s00122-010-1485-4

Liu, H., and Chen, G. B. (2017). A fast genomic selection approach for large genomic data. *Theor. Appl. Genet.* 130, 1277–1284. doi: 10.1007/s00122-017-2887-3

Liu, H., and Chen, G. B. (2018). A new genomic prediction method with additive-dominance effects in the least-squares framework. *Heredity* 121, 196–204. doi: 10.1038/s41437-018-0099-5

Liu, H., and Yu, S. (2023). A dimensionality-reduction genomic prediction method without direct inverse of the genomic relationship matrix for large genomic data. *Plant Cell Rep.* 42, 1825–1832. doi: 10.1007/s00299-023-03069-8

Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., et al. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40, W622–W627. doi: 10.1093/nar/gks540

Luo, P., Wang, H., Ni, Z., Yang, R., Wang, F., Yong, H., et al. (2023). Genomic prediction of yield performance among single-cross maize hybrids using a partial diallel cross design. *Crop J.* 11, 1884–1892. doi: 10.1016/j.cj.2023.09.009

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Meuwissen, T., Hayes, B., and Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1, 221–237. doi: 10.1146/annurev-animal-031412-103705

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Mienye, I. D., and Sun, Y. (2022). A survey of ensemble learning: concepts, algorithms, applications, and prospects. *IEEE Access* 10, 99129–99149. doi: 10.1109/ACCESS.2022.3207287

Montesinos-López, A., Montesinos-López, O. A., Gianola, D., Crossa, J., and Hernández-Suárez, C. M. (2018). Multi-environment genomic prediction of plant traits using deep learners with a dense architecture. *G3: Genes Genomes Genet.* 8, 3813–3828. doi: 10.1534/g3.118.200740

Navarro, J. A. R., Wilcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., et al. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* 49, 476–480. doi: 10.1038/ng.3784

Ndlovu, N., Spillane, C., McKeown, P. C., Cairns, J. E., Das, B., and Gowda, M. (2022). Genome–wide association studies of grain yield and quality traits under optimum and low–nitrogen stress in tropical maize (Zea mays L.). *Theor. Appl. Genet.* 135, 4351–4370. doi: 10.1007/s00122-022-04224-7

Ni, X., Liu, L., and Lei, B. (1996). Study on the selection of maize inbred line S_{37} suited to mountain area maize breeding. *J. Sichuan Agric. Univ.* 14, 366–370. doi: 10.16036/j.issn.1000-2650.1996.03.011

Ogutu, J. O., Piepho, H. P., and Schulz-Streeck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc.* 5, S11. doi: 10.1186/1753-6561-5-S3-S11

Pan, G. T., Yang, K. C., Li, W. C., Huang, Y. B., Gao, S. B., Lan, H., et al. (2020). A review of the research and application of heterotic groups and patterns of maize breeding in Southwest China. *J. Maize Sci.* 28, 1–8. doi: 10.13597/j.cnki.maize.science.20200101

Peng, B., Li, Y., Wang, Y., Liu, C., Liu, Z., Tan, W., et al. (2011). QTL analysis for yield components and kernel-related traits in maize across multi-environments. *Theor. Appl. Genet.* 122, 1305–1320. doi: 10.1007/s00122-011-1532-9

Prasanna, B. M., Pixley, K., Warburton, M. L., and Xie, C. X. (2010). Molecular marker-assisted breeding options for maize improvement in Asia. *Mol. Breed.* 26, 339–356. doi: 10.1007/s11032-009-9387-3

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44, 217–220. doi: 10.1038/ng.1033

Stevens, R. (2008). Prospects for using marker-assisted breeding to improve maize production in Africa. *J. Sci. Food Agric.* 88, 745–755. doi: 10.1002/jsfa.3154

Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H., and Melchinger, A. E. (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197, 1343–1355. doi: 10.1534/genetics.114.165860

Tian, J., Wang, C., Chen, F., Qin, W., Yang, H., Zhao, S., et al. (2024). Maize smart-canopy architecture enhances yield at high densities. *Nature*. 632, 576-584. doi: 10.1038/s41586-024-07669-6

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* 132, 669–686. doi: 10.1007/s00122-018-3270-8

Wang, X., Zhang, Z., Xu, Y., Li, P., Zhang, X., and Xu, C. (2020). Using genomic data to improve the estimation of general combining ability based on sparse partial diallel cross designs in maize. *Crop J.* 8, 819–829. doi: 10.1016/j.cj.2020.04.012

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., et al. (2021). The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* 22, 148. doi: 10.1186/s13059-021-02370-7

Xu, S., Zhu, D., and Zhang, Q. (2014). Predicting hybrid performance in rice using genomic best linear unbiased prediction. *PNAS* 111, 12456–12461. doi: 10.1073/pnas.1413750111

Yan, J., Xu, Y., Cheng, Q., Jiang, S., Wang, Q., Xiao, Y., et al. (2021). LightGBM: accelerated genomically designed crop breeding through ensemble learning. *Genome Biol.* 22, 271. doi: 10.1186/s13059-021-02492-y

Yang, X., Gao, S., Xu, S., Zhang, Z., Prasanna, B. M., Li, L., et al. (2011a). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* 28, 511–526. doi: 10.1007/s11032-010-9500-7

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011b). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011

Yu, G., Cui, Y., Jiao, Y., Zhou, K., Wang, X., Yang, W., et al. (2023). Comparison of sequencing-based and array-based genotyping platforms for genomic prediction of maize hybrid performance. *Crop J.* 11, 490–498. doi: 10.1016/j.cj.2022.09.004

Zhang, X., Zhang, H., Li, L., Lan, H., Ren, Z., Liu, D., et al. (2016). Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. *BMC Genomics* 17, 697. doi: 10.1186/s12864-016-3041-3

Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H. P., Longin, F. H., et al. (2012). Accuracy of genomic selection in European maize elite breeding populations. *Theor. Appl. Genet.* 124, 769–776. doi: 10.1007/s00122-011-1745-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 October 2024
DOI 10.3389/fpls.2024.1476829

OPEN ACCESS

EDITED BY
Yi-Hong Wang,
University of Louisiana at Lafayette,
United States

REVIEWED BY
Bo Peng,
Xinyang Normal University, China
Jie Hu,
Fujian Agriculture and Forestry University,
China

*CORRESPONDENCE
Yonghua Xie
xieyh272811@163.com
Duo Xia
xiaduo0610@foxmail.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 06 August 2024
ACCEPTED 12 September 2024
PUBLISHED 01 October 2024

CITATION
Xu P, Qin Y, Ma M, Liu T, Ruan F, Xue L, Cao J, Xiao G, Chen Y, Fu H, Zhou G, Xie Y and Xia D (2024) Genome-wide association study reveals the genetic basis of rice resistance to three herbicides.
Front. Plant Sci. 15:1476829.
doi: 10.3389/fpls.2024.1476829

Genome-wide association study reveals the genetic basis of rice resistance to three herbicides

Peizhou Xu[1†], Yuhe Qin[2†], Maosen Ma[1], Tengfei Liu[1], Fenhua Ruan[1], Le Xue[1], Jiying Cao[1], Guizong Xiao[1], Yun Chen[1], Hongyan Fu[1], Gege Zhou[1], Yonghua Xie[3*] and Duo Xia[1*]

[1]State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Rice Research Institute, Sichuan Agricultural University, Chengdu, China, [2]Department of Research and Development, Luzhou Taifeng Seed Industry Co., Ltd., Luzhou, Sichuan, China, [3]Department of Research and Development, Zoeve Seed Co., Ltd., Chengdu, Sichuan, China

Crop resistance to herbicides is crucial for agricultural productivity and sustainability amidst escalating challenges of weed resistance. Uncovering herbicide resistant genes is particularly important for rice production. In this study, we tested the resistance to three commonly used herbicides: glufosinate, glyphosate and mesotrione of 421 diverse rice cultivars and employed genome-wide association studies (GWAS) to unravel the genetic underpinnings of resistance to these three herbicides in rice. We discovered that cultivated rice exhibited rich variation in resistance to the three herbicides, and the differences among subpopulations were significant. Six identified associations harboring candidate genes for resistance to these herbicides were significant. Among them, *RGlu6* and *RGly8* were the major QTL for resistance to glufosinate and glyphosate, respectively. The favorable alleles of *RGlu6* and *RGly8* were primarily present in *japonica* cultivars that originated from Europe, highlighting the geographic and genetic diversity of herbicide resistance and emphasizing the localized selection pressures in European rice varieties. Moreover, our findings might suggest that traditional target genes may not contain tolerant alleles in nature, and alternative mechanisms with novel loci associated with resistance may work. By mapping the genes for herbicide resistance, our results may help develop new strategies to combat the dual challenges on effective weed management and herbicide sustainability.

KEYWORDS

rice, herbicide resistance, genome-wide association studies, genetic diversity, glufosinate, glyphosate, mesotrione

Introduction

Rice is a crucial staple food for more than half of the global population, playing an essential role in agricultural production, especially in Asian countries (FAO, http://www.fao.org). Weeds pose a horrible threat to rice by competing for nutrients, water, sunlight, and space, and by increasing the incidence of diseases and insect pests (Akbar

et al., 2011). This competition can lead to substantial reductions in rice yield and grain quality, up to over 40% due to weed damage (Burgos et al., 2014). The shift from transplanting to direct seeding, driven by urbanization and labor shortages has exacerbated weed problems (Chaudhary et al., 2023). Traditional methods for weed control, such as tillage, irrigation, hand-weeding etc., are labor-intensive and resource-demanding. Using herbicides is the most effective way to control weeds, but many herbicides also damage or kill rice plants (Jin et al., 2022). The development and use of herbicide-resistant cultivars for an effective control of weeds without harming the plant have therefore become critical in modern agricultural practices (Kawahara et al., 2013).

Glufosinate, glyphosate and mesotrione are three widely used herbicides and significant progress has been made in understanding and developing resistant plants to these herbicides. As a major competitive inhibit target of glufosinate, glutamine synthase (GS) plays a pivotal role in plants' resistance to glufosinate and mutations in GS genes could enhance plants' resistance to glufosinate (Tian et al., 2015). Glyphosate is a broad-spectrum post-emergent herbicide and mainly targets 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), a key enzyme in the biosynthesis of aromatic amino acids, and phenolics, and gain of function of EPSPS genes could endow plants with resistance to glyphosate (Sammons and Gaines, 2014). It has been reported that acetyltransferases (ACEs) synthesizing genes can be induced by mesotrione, indicating a role for ACEs in resistance to mesotrione of crops (Chen et al., 2022). In addition, other genes like *OsACC1* which encodes acetyl-CoA carboxylase and *OsALS1* which encodes acetolactate synthase were reported to be responsible for resistance to herbicides like aryloxyphenoxypropionates (APPs) and pyrimidinyl carboxy (PC) herbicides, respectively (Okuzaki et al., 2007; Xu et al., 2021). Researchers have successfully created herbicide-resistant rice varieties through artificial mutagenesis, genome editing, introducing and overexpression of these genes (Li et al., 2016; Maeda et al., 2019; Chen et al., 2021; Ren et al., 2023; Zhang et al., 2023). Despite these advancements, the genetic resources for herbicide resistance are still limited in rice. The evolution of herbicide-resistant weeds due to the extensive use of chemical control strategies has further complicated this issue. For instance, over 255 weed species worldwide have developed resistance to at least one herbicide, including 42 species resistant to glyphosate (Perotti et al., 2019). This situation underscores the need for continuous research and development of new herbicide resistance mechanisms and the identification of novel resistance genes to ensure sustainable rice production.

Genome-wide association studies (GWAS) have emerged as a powerful tool for dissecting the genetic basis of complex quantitative traits in plants (Zhou et al., 2017, 2021a, 2021b). This method enables the identification of genetic variants associated with specific traits by analyzing the entire genome (Xia et al., 2022). In this study, we employed GWAS to investigate the genetic basis of resistance to three different herbicides in a diverse rice germplasm population. This approach allowed us to identify key genetic loci associated with herbicide resistance, providing valuable insights into the development of new herbicide-tolerant rice varieties. By leveraging the genetic diversity present in natural populations, GWAS offers a robust framework for enhancing our understanding of herbicide resistance and for breeding rice varieties with improved resistance profiles.

Materials and methods

Plant materials and phenotyping

The 421 cultivated rice varieties used in this study were selected from diverse germplasms collected domestically and internationally, and the names, geographical origin, and subpopulation information of which have been previously reported (Zhao et al., 2015). In 2022, these rice varieties were planted at the South Breeding Base of Sichuan Agricultural University in Lingshui County, Hainan Province, China. The seeds of each were individually harvested in April 2023, uniformly air-dried and placed in the seed storage facility in the Rice Research Institute of Sichuan Agricultural University in Chengdu, Sichuan Province.

Sixty grains of each accession were sown in a flowerpot (15 cm diameter and 16 cm height) in July 2023 with three replications for herbicide evaluation in the Chengdu Campus of Sichuan Agricultural University. Half-lethal dose of glyphosate, glufosinate and mesotrione solutions was sprayed, respectively at the 3-4 leaf stage as 2.25 L/hm^2 of 41% glyphosate (Fuhua Chemical), 3.00 L/hm^2 of 20% glufosinate (BASF) and 1.20 L/hm^2 of 10% mesotrione (Hubei Jiahui Xingcheng Biotechnology Co., Ltd.). The response of each pot to herbicide injury was scored with 0-4 scale five days after the application of glyphosate and glufosinate, but seven days after the application of mesotrione, respectively, where 0 was for the plants with no apparent injury, level 1 for slight effect on seedlings, level 2 for moderate effect, level 3 for severe effect and 4 with the most severe injury.

Genome wide association study

GWAS analyses for herbicide resistance traits were performed for the entire panel and for *indica* and *japonica* group separately, using mixed linear models to take population structure and relative kinship into consideration for statistical association. SNPs with a minor allele frequency (MAF) greater than 5% and a missing rate less than 15% were selected for the association analysis, and 6.3 million variants, including 5.5 million SNPs and 800,000 InDels, were used for GWAS, as well as estimating population structure and kinship coefficients. The thresholds of genome-wide significance were determined using a modified Bonferroni correction as described by Li et al. (2012), where the total number of SNPs (M) used for threshold calculation was replaced by the effective number of SNPs (Me), in mixed linear models provided by EMMAX program (Kang et al., 2010). The physical locations of the SNPs and InDels were identified based on the Rice Annotation version 7.0 of the variety Nipponbare (Kawahara et al., 2013) from the website http://rice.uga.edu/.

The Bayesian clustering program fast Structure (Raj et al., 2014) was used to calculate different levels of K (K = 2-5), and the

command choose_K.py was employed to identify the model complexity for maximizing the marginal likelihood. LD heatmaps and gene haplotypes were constructed using SNPs and InDels (MAF >0.03) from the 1 Mb regions around each peak SNP using R package 'LD heatmap' (Shin et al., 2006).

Variant-function-based association analyses

We annotated the functional variations within a 1 Mb region surrounding the associated loci using snpEff software (Cingolani et al., 2012), and classified all polymorphisms in the candidate region into four groups, as for 1) those predicted to induce amino acid exchanges or to change splicing junctions (GT or AG at the beginning or end of an intron, respectively), 2) those located at the 5′ flanking sequences of genes (e.g., within 2kb of the first ATG, typically the promoter region), 3) those located within a gene but did not meet the criteria for Group I or II (for example, located in a coding region but not predicted to change an amino acid, an intron, or a 3′ noncoding sequence), and 4) those located outside of coding regions. By identifying the most significant variant in Group I, potential candidate genes were determined.

Gene haplotype and geographic distribution analyses

The haplotype of each candidate gene was analyzed by downloading the genotypes of variant sites located 2kb upstream of the transcription start site and within the coding sequence of the gene from the RiceVarMap 2.0 (ricevarmap.ncpgr.cn) (Zhao et al., 2021). To improve the efficiency of the haplotype analysis, those variant sites with a minor allele frequency less than 0.05 and a missing rate greater than 10% were filtered out. The genotypes of remaining high-quality variants were used for haplotype classification. To better display the distribution pattern of haplotypes, only those haplotypes containing more than 10 varieties were shown. The information on geographical distribution of the 421 accessions was also downloaded from the RiceVarMap website. The world map was plotted using the 'maps' package in R (Becker and Wilks, 1995), and the characteristics of different haplotypes on geographical distribution were displayed in pie chart form using the 'ape' package in R (Paradis and Schliep, 2019).

Statistical analyses

Violin plots, correlations, and multiple comparisons were constructed using phenotypic means from three replicates for each accession. The *P* values for Pearson's correlation coefficients were calculated by two-sided *t*-tests using the cor.test() function in R (Ihaka and Gentleman, 1996). Multiple comparisons of phenotypic data among different subgroups or haplotypes were conducted using the duncan.test() command in the 'agricolae' package in R, with a significant probability at 0.01. The 'vioplot' package in R was used to create violin plots for the multiple comparisons (https://github.com/TomKellyGenetics/vioplot).

Result

The distribution of three herbicide resistance in rice germplasm

By grading the herbicide resistance, we observed significant differences among the 421 accessions responding to the three herbicides, indicating the feasibility of the grading standards (Figures 1A; Supplementary Figure S1). The cultivars generally exhibited low resistance to all three herbicides, with average resistance ratings around level 3. Less than 5% cultivars exhibited relatively high resistance (resistance ratings smaller than level 2 to the three herbicides, and a *Temperate Japonica* cultivar HB-6-2 was the relatively highest glufosinate and glyphosate resistant cultivar in the tested materials (Supplementary Figure S1; Supplementary Tables S4, S5). There is a significant positive correlation between the resistance to the three herbicides in cultivated rice (r^2 = 0.26,0.13,0.14), although the correlation coefficients are relatively low (Figure 1B). Among them, the positive correlation between glufosinate resistance and glyphosate resistance is the most significant, with a correlation coefficient of 0.26 (p= 7.5 × 10^{-5}).

Different subpopulations varied in resistance to the three herbicides

A total of 6.3 million high-quality variants, including 5.5 million SNPs and 0.8 million InDels, queried from RiceVarMap (http://ricearmap.ncpgr.cn), were used to estimate population structure and *Kinship* coefficients, as well as GWAS. When *K* equaled 2, the 421 accessions in the whole panel were divided into 300 *Indica*, 106 *Japonica* and 15 *Intermediate* ones. And When *K* equaled 5, the 300 *Indica* accessions were further divided into 38 *Aus*, 143 *Indica I* (*IndI*), 119 *Indica II* (*Ind II*), while the 106 *Japonica* accessions were divided into 60 *Temperate Japonica* (*TeJ*) and 46 *Tropical Japonica* (*TrJ*). Based on this population structure, we noticed a varying resistance to different herbicides of different subgroups (Figure 1C). For glufosinate and glyphosate, the *TeJ* subgroup exhibited significantly higher resistance than other subgroups (Figure 1C); for halosulfuron-methyl, the *Aus* subgroup showed significantly lower resistance than other subgroups. Thus, the cultivars with different genetic backgrounds demonstrated diverse herbicide resistance.

Loci associated with herbicide resistance

The strict inbreeding mating habits of rice have led to significant subpopulation structure and considerable linkage disequilibrium (LD), which reduce the mapping resolution of association studies and increase the occurrence of type I errors

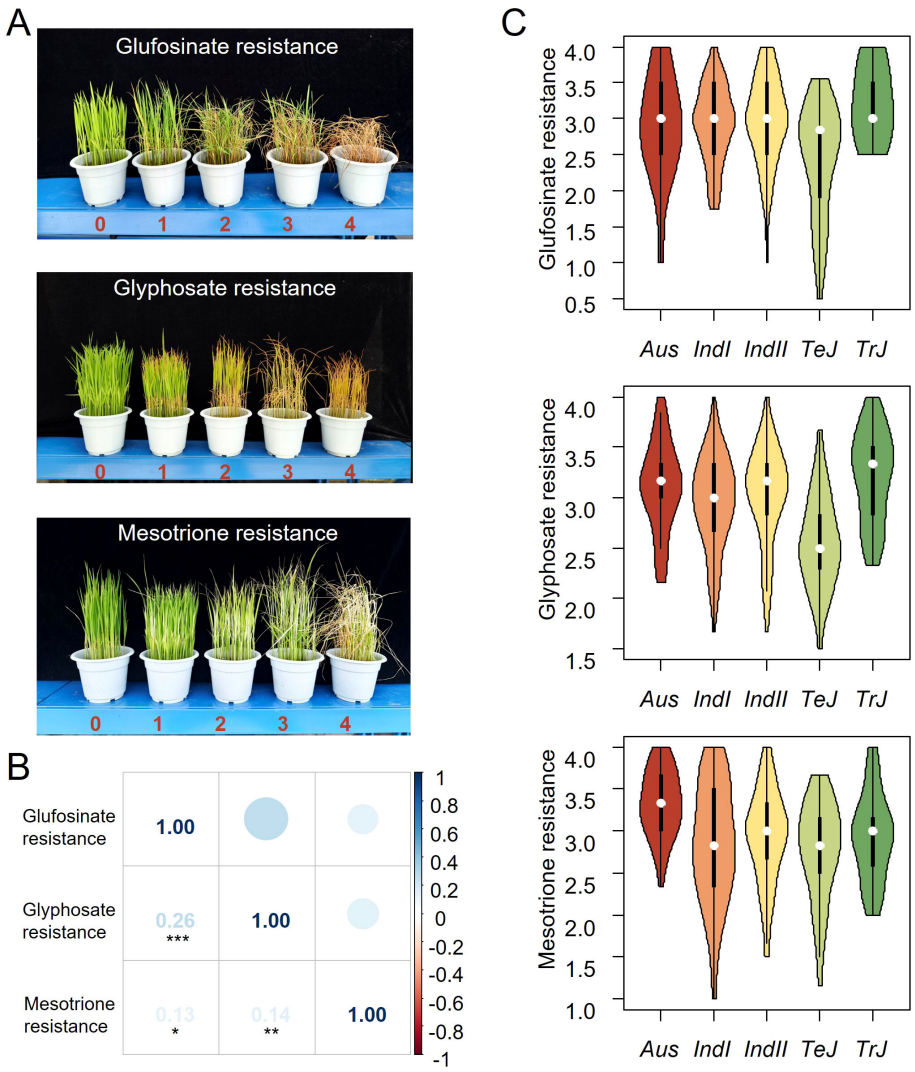


FIGURE 1
Identification, phenotypic distribution, and correlation of resistance to three herbicides in cultivated rice. **(A)** Grading criteria for tolerance to three herbicides. **(B)** Correlation among three herbicide tolerances in cultivated rice. **(C)** Phenotypic distribution of tolerance to three herbicides across different subgroups in cultivated rice.

(Zhou et al., 2017). To address these issues, we used a mixed model to correct the interferences from subpopulation structure by incorporating the first three principal components (PCs) as covariates in the model, and conducted genome-wide association studies (GWAS) in the *Indica* group (303 accessions), *Japonica* group (181 accessions), and the entire group (all varieties) (see Methods). The calculated genome-wide significance thresholds, based on a nominal level of 0.05, were $P = 6.6 \times 10^{-8}$, 8.7×10^{-8}, and 2.0×10^{-7} for the whole panel, *Indica* and *Japonica*, respectively (Supplementary Table S1).

Five significant loci associated with resistance to glufosinate and glyphosate, but none to mesotrione were identified in the whole panel (Figure 2; Supplementary Figures S2, S3). Detailed association results for each herbicide in each subpopulation are presented in Table 1. Three loci associated with glufosinate resistance in the whole panel totally explained 25.76% of the phenotypic variance, with effect of each varying from 15.95% to 19.20% (Supplementary Table S2). For glyphosate resistance, two loci accounted for 26.69% of the phenotypic variance in total, and 5.15% and 20.06% alone, indicating a significant role of the additive effects on the variation of glyphosate resistance (Supplementary Table S3).

Functional and haplotype analyses of *RGlu6* for rice glufosinate resistance

Since *RGlu6* explained the highest phenotypic variation (R^2 = 19.20%, Supplementary Table S2), we chose it as a major QTL for rice glufosinate resistance and conducted variant-function-based association analysis of the associated loci regions on *RGlu6* (see Methods). By analyzing the function of variants within 500 kb upstream and downstream of the peak SNP and the linkage disequilibrium (LD) blocks in that segment (Figures 3A, B), we analyzed the candidate genes and functional variants of *RGlu6*. LD

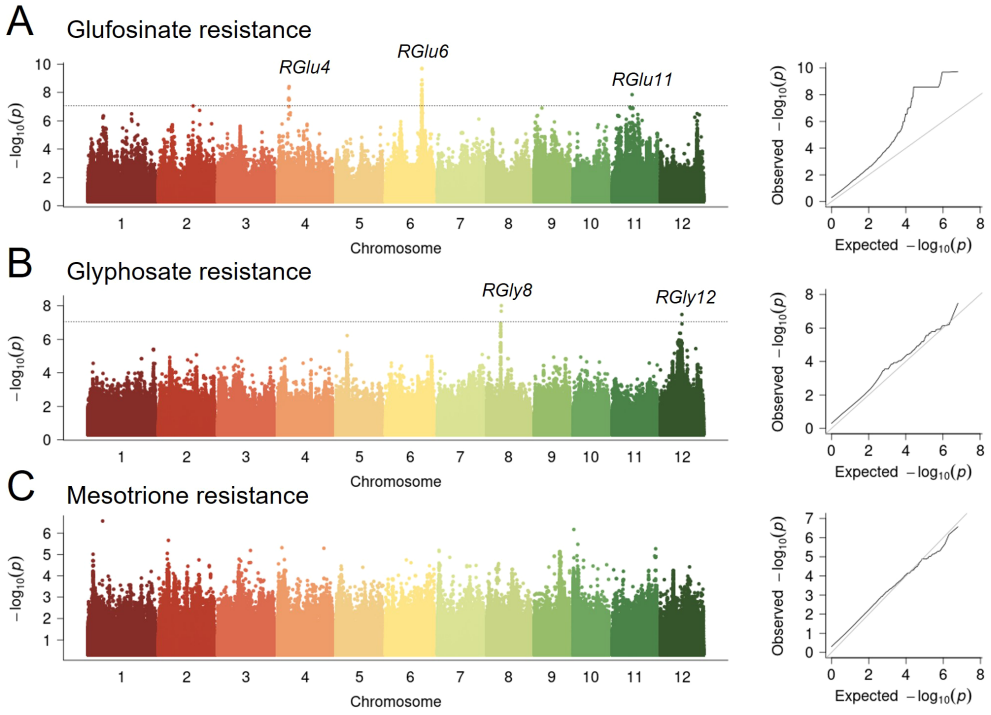


FIGURE 2
Genome-wide association results for resistance to three herbicides in 421 cultivated rice. Manhattan plots and quantile–quantile plots depicting GWAS results using a mixed model for tolerance to glufosinate **(A)**, glyphosate **(B)** and mesotrione **(C)**. Associations identified in all accessions (lower), indica subspecies (middle), and japonica subspecies (upper). The x axis depicts the physical location of SNPs across the 12 chromosomes of rice and the y axis depicts the -$\log_{10}$ (*P* value).

decay around *RGlu6* was rapid with no obvious LD blocks, where the most significant functional variant (vg0623197432) was only 6 kb away from the peak SNP (Table 1). A significant glufosinate resistance difference between type A and type G of vg0623197432 variant was observed (p= 1.8×10^{-10}, Figure 3C). Since variant vg0623197432 was located on the exon of *LOC_Os06g39070*, which encodes a UDP-glucosyl transferase (UGT), we consider it as a candidate gene for *RGlu6*.

Using the genotypes of the 2 kb promoter and coding region variations of *LOC_Os06g39070* with 61 SNPs and 13 InDels, we categorized *RGlu6* of these 421 cultivars into 5 haplotypes, namely H1 - H5 (Figure 3D). Haplotype H1 was predominantly present in *TeJ* and *TrJ*, H2 and H3 mainly in *IndI* and *IndII*, H4 in *Aus*, and H5 only in *TeJ*, respectively. Multiple comparison demonstrated no significant differences were among haplotypes H1-H4, while haplotype H5 exhibited significantly higher resistance to glufosinate than others, indicating beneficial allelic variations of H5 haplotype (Figure 3E). Thus, we classified *RGlu6* into two allelic genotypes with merging the indistinct H1-H4 into $RGlu^A$ genotype, and considering H5 as $RGlu^G$ genotype. By analyzing the geographic distribution of these allelic genotypes (Supplementary Table S4), we observed that the accessions with the $RGlu6^G$

TABLE 1 SNPs and candidate genes significantly associated with herbicides resistance.

Trait	Allele[a]	Chr	Pos[b]	*P* value	QTL	Dis(kb)[c]	Candidate gene[d]	Description
Glufosinate	A/G	4	7431857	3.81E-09	*RGlu4*	-40	*LOC_Os04g13130*	hypothetical protein
Glufosinate	A/G	6	23190901	2.07E-10	*RGlu6*	6	*LOC_Os06g39070*	UDP-glucosyl transferase
Glufosinate	A/C	11	13551277	1.87E-08	*RGlu11*	9	*LOC_Os11g08190*	expressed protein
Glyphosate	A/G	8	8954806	1.11E-08	*RGly8*	-7	*LOC_Os08g14850*	resistance protein
Glyphosate	G/C	12	12637368	5.10E-08	*RGly12*	40	*LOC_Os12g22460*	hypothetical protein
Mesotrione	T/C	1	9366585	9.69E-07	*RMes1*	6	*LOC_Os01g16510*	expressed protein

[a]Major allele/minor allele; underlined base is the reference allele. Position in base pairs for the lead SNP according to the Nipponbare reference genome version 7. Distance from the candidate gene to the peak SNP. Plausible biological candidate gene in the locus or the nearest annotated gene to the lead SNP.

10.3389/fpls.2024.1476829

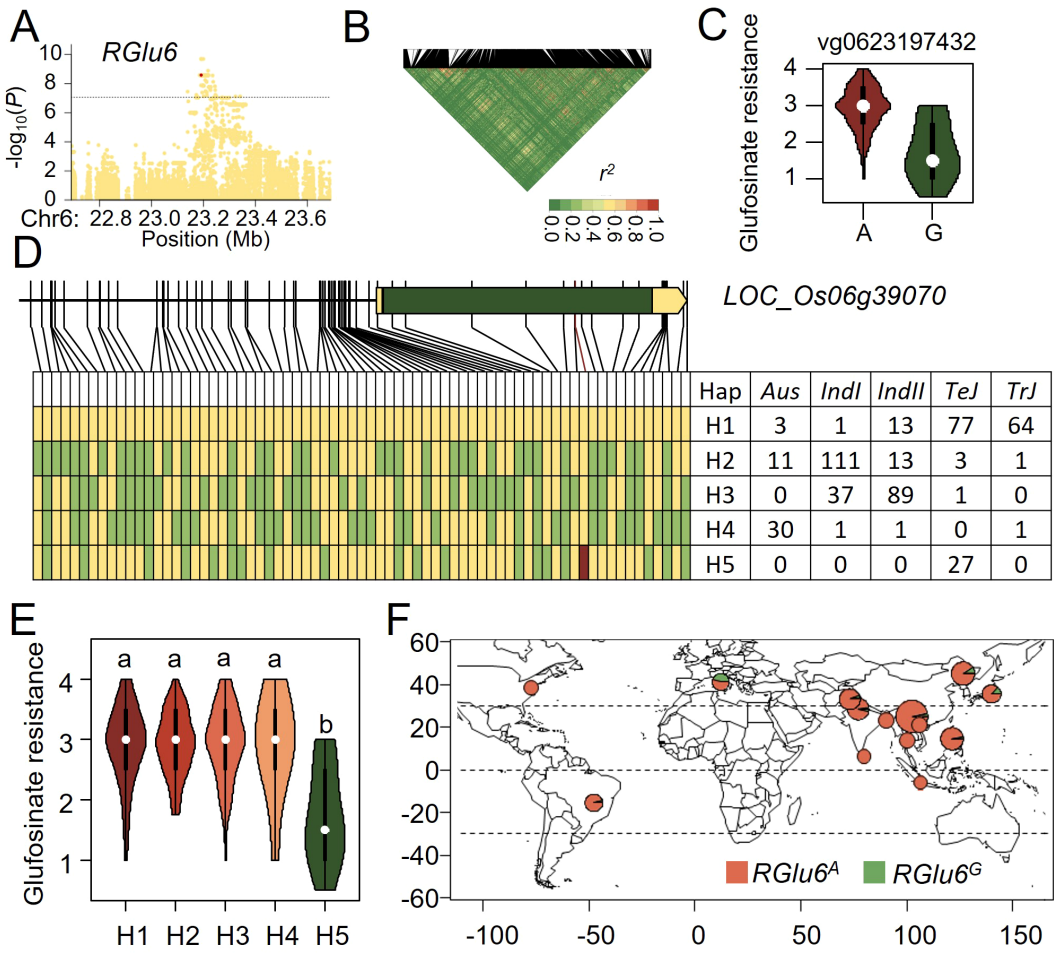


FIGURE 3
Natural variation of *RGlu6* affect glufosinate resistance in rice. **(A)** Manhattan plot for the 1 Mb region around the glufosinate resistance-associated locus *RGlu6*. **(B)** LD heatmap for the region around *RGlu6*. **(C)** Potential functional variants at *RGlu6* and their effects on glufosinate tolerance. **(D)** Haplotype analysis of the candidate gene at *RGlu6*. Yellow represents the reference genome genotype, green represents alternate genotypes, and red represents functional variants. **(E)** Phenotypic distribution of the five haplotypes at *RGlu6*. **(F)** Geographic distribution of accessions with different *RGlu6* allele in the GWAS panel.

genotype mainly originated from Europe (Figure 3F), which indicates that $RGlu6^G$ may have arisen from a mutation and was retained locally through natural selection.

Functional and haplotype analyses of *RGly8* for rice glyphosate resistance

Analyzing candidate genes at the *RGly8* locus for rice glyphosate resistance in a rapid LD decay around *RGly8*, we found that vg0808951557 was the most significant functional variant located only 7 kb from the peak SNP (Figures 4A, B). The T to C mutation of vg0808951557 which was in the fourth exon of *LOC_Os08g14850*, led to an amino acid changed from serine to proline (Figure 4C). Since *LOC_Os08g14850* encodes a resistance protein and was reported to be involved biotic and abiotic stresses, we considered it as a candidate gene for *RGly8*.

Seventy four SNPs and five InDels were used to analyze the haplotypes of *RGly8* and seven haplotypes, namely H1 - H7 were categorized based on the glyphosate resistance and genotypes of the 421 accessions (Figure 4D). Haplotypes H1-H5 were predominantly present in *Indica*, while H6 and H7 were mainly present in *Japonica*. The superior C genotype of vg0808951557 was exclusively found in the H6 haplotype. Multiple comparisons revealed significant differences for glyphosate resistance among these seven haplotypes, with H6 haplotype exhibiting a significantly higher glyphosate resistance than other haplotypes (Figure 4E). Therefore, we designated the H6 haplotype as the allelic genotype $RGly8^C$, and the other haplotypes as $RGly8^T$. Analysis of the geographic distribution of these alleles showed that the superior allele $RGly8^C$ primarily originated from Europe, the same as that of $RGlu6^G$ (Figure 4F; Supplementary Table S4).

These results suggested that resistance to both glufosinate and glyphosate may have been locally selected in Europe, leading to an

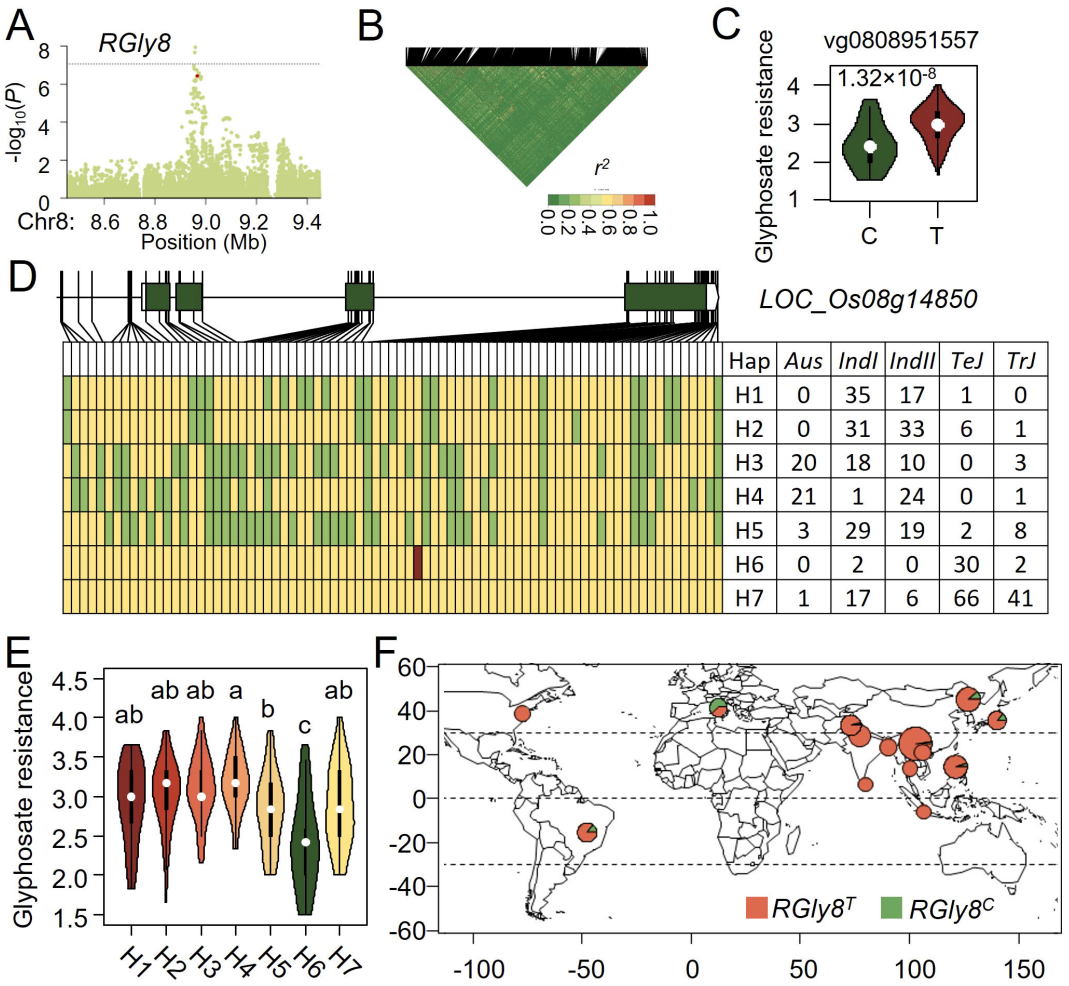


FIGURE 4
Natural variation of *RGly8* affect glyphosate resistance in rice. **(A)** Manhattan plot for the 1 Mb region around the glyphosate resistance-associated locus *RGly8*. **(B)** LD heatmap for the region around *RGly8*. **(C)** Potential functional variants at *RGlu6* and their effects on glyphosate tolerance. **(D)** Haplotype analysis of the candidate gene at *RGly8*. Yellow represents the reference genome genotype, green represents alternate genotypes, and red represents functional variants. **(E)** Phenotypic distribution of the five haplotypes at *RGly8*. **(F)** Geographic distribution of accessions with different *RGly8* allele in the GWAS panel.

increase in the frequency of superior resistance alleles in rice accessions originated from that region.

Discussion

The intense application of herbicides has effectively assisted in weeds control but meanwhile accelerated the evolution of resistance to herbicides of both crops and weeds (Peterson et al., 2018). Discovering and developing herbicide-resistant crop is one of the most promising ways for weed control and crop production. In this study, we tested the resistance to three frequently herbicides, namely glufosinate, glyphosate and mesotrione, of 421 diverse rice accessions. Varying degrees of resistance to these herbicides were discovered (Supplementary Figure S1), indicating the possibility of natural herbicide-resistant rice varieties. Besides, a relatively higher resistance in average to these three herbicides in *TeJ* compared to other subpopulations were observed (Figure 1C). Since TeJ mainly originated from Europe, a higher herbicides selection pressure in Europe might be the cause (Peterson et al., 2018).

GS and EPSPS are the major competitively inhibited targets of glufosinate and glyphosate, respectively, and rice varieties endowed with resistance to these two herbicides were mostly genetically modified on genes encoding GS and EPSPS (Achary et al., 2020). *OsGS1;1* and *OsGS2*, encoding glutamine synthetase, were reported to be involved in rice growth and resistance to abiotic stresses, but the resistance to glufosinate has not been reported (Cai et al., 2010; James et al., 2018). Our study aimed to identify natural variant loci associated with resistance to either of glufosinate and glyphosate in rice through GWAS, but none was detected in their target gene regions (Figures 2A, B, Supplementary Figures S2, S3). This result suggested that either these target genes might not possess alleles to confer natural herbicide resistance, or such natural resistance alleles may be either too weak or too rare to be detected by GWAS.

Since the 421 geographically and genetically diverse rice cultivars displayed various resistance to the three herbicides

(Figure 1C, Supplementary Figure S1) and QTL which were reported to be associated with resistance to these herbicides did not overlap with neither EPSPS genes nor GS genes (Bao et al., 2022; Xu et al., 2023), alternative mechanisms of plant herbicide resistance beyond just reducing the direct interaction between the target gene protein and the herbicide might exist. In the association analysis for glufosinate resistance, we defined *LOC_Os06g39070* which encodes a UDP-glucosyl transferase (UGT) as a candidate gene. UGT is a class of enzymes widely found in plants, whose primary function is to transfer glucose units from uridine diphosphate glucose (UDP-glucose) to a variety of acceptor molecules, including hormones, toxins, and secondary metabolites (Lairson et al., 2008; Tiwari et al., 2016). In rice, UGT genes participates in response to abiotic and biotic stresses, and gain of function of OsUGTs could improve rice resistance to various stresses through elevating scavenging reactive oxygen species (ROS) and plant hormones (Wang et al., 2021, 2023; He et al., 2023). Thus, *LOC_Os06g39070* might reduce the toxicity of glufosinate through hormones or ROS scavenging processes. In the association analysis for glyphosate resistance, *LOC_Os08g14850* which encodes resistance protein CC-NBS-LRR was detected (Figures 2B, 4). Previous studies have shown that resistance proteins play a role not only in plant responses to biotic stress but also in responses to abiotic stress (Atkinson and Urwin, 2012). *LOC_Os08g14850*/*SCR8* has been reported to be involved in rice's resistance to bacterial blight and mutant *scr8* displayed a defected root growth phenotype (Hu et al., 2021). Taking these together, *LOC_Os08g14850* might be a potential target for enhancing rice resistance to glyphosate.

Haplotype analysis of *RGlu6* and *RGly8* revealed that superior alleles of *RGlu6* and *RGly8* are predominantly present in *TeJ* (Figures 3D, E, 4D, E). Geographic analysis of $RGlu6^{G}$ and $RGly^{C}$ showed that either of them were originated from Europe (Figures 3F, 4F). These results indicated a possible severe herbicide pressure in Europe (Devos et al., 2006).

In summary, this study reveals the genetic basis of rice resistance to three herbicides through genome-wide association analysis, offering new insights for deciphering new mechanisms of herbicide resistance and breeding herbicide-resistant rice varieties.

Conclusion

Enhancing rice resistance to herbicides is the most effective way to maintain rice production against weeds. In this study, we tested the resistance to three major herbicides, namely glufosinate, glyphosate and mesotrione of 421 diverse rice cultivars and identified 5 significant associations for rice resistance to glufosinate and glyphosate. By candidate gene-based GWAS, we predicted two genes, *LOC_Os06g39070* and *LOC_Os08g14850* for rice resistance to glufosinate and glyphosate, and found superior alleles for these two genes against glufosinate and glyphosate, respectively. Our results will shed light on the genetic diversity of rice resistance to herbicides and provide promising targets for enhancing rice resistance to herbicides.

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

PX: Conceptualization, Methodology, Supervision, Writing – original draft. YQ: Funding acquisition, Supervision, Writing – review & editing. MM: Data curation, Investigation, Writing – original draft. TL: Formal analysis, Investigation, Writing – review & editing. FR: Investigation, Writing – review & editing. LX: Investigation, Writing – review & editing. JC: Investigation, Writing – original draft. GX: Investigation, Writing – original draft. YC: Investigation, Software, Writing – original draft. HF: Investigation, Writing – original draft. GZ: Investigation, Writing – original draft. YX: Funding acquisition, Supervision, Writing – review & editing. DX: Data curation, Funding acquisition, Investigation, Resources, Writing – review & editing, Conceptualization, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Science and Technology Department of Sichuan Province, China (Project No. 2023NSFSC1933, 2023NSFSC0215, and 2024NSFSC0333), the Sichuan National Science Foundation Innovation Research Group (Project No 2023NSFSC 1996), and the Bureau of Science and Technology of Chengdu City, Sichuan China (Project No 2022-YF09-00036-SN and 2023-YF08-00005-SN).

Conflict of interest

Author YQ was employed by Luzhou Taifeng Seed Industry Co., Ltd. Author YX was employed by Zoeve Seed Co., Ltd.

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1476829/full#supplementary-material

References

Achary, V. M. M., Sheri, V., Manna, M., Panditi, V., Borphukan, B., Ram, B., et al. (2020). Overexpression of improved *EPSPS* gene results in field level glyphosate tolerance and higher grain yield in rice. *Plant Biotechnol. J.* 18, 2504–2519. doi: 10.1111/pbi.v18.12

Akbar, N., Jabran, K., and Ali, M. A. (2011). Weed management improves yield and quality of direct seeded rice. *Aust. J. Crop Sci.* 5, 688–694.

Atkinson, N. J., and Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot.* 63, 3523–3543. doi: 10.1093/jxb/ers100

Bao, J., Gao, Y., Li, Y., Wu, S., Li, J., Dong, Z., et al. (2022). Genetic analysis and fine mapping of *zmGHT1* conferring glufosinate herbicide tolerance in maize (*Zea mays* L.). *Int. J. Mol. Sci.* 23, 11481. doi: 10.3390/ijms231911481

Becker, R. A., and Wilks, A. R. (1995). *Constructing a Geographical Database, Statistics Research Report [95.2]*. Murray Hill, NJ, USA: AT&T Bell Laboratories.

Burgos, N. R., Singh, V., Tseng, T. M., Black, H., Young, N. D., Huang, Z., et al. (2014). The impact of herbicide-resistant rice technology on phenotypic diversity and population structure of United States weedy rice. *Plant Physiol.* 166, 1208–1220. doi: 10.1104/pp.114.242719

Cai, H. M., Xiao, J. H., Zhang, Q. F., and Lian, X. M. (2010). Co-suppressed *glutamine synthetase2* gene modifies nitrogen metabolism and plant growth in rice. *Chin. Sci. Bull.* 55, 823–833. doi: 10.1007/s11434-010-0075-9

Chaudhary, A., Venkatramanan, V., Kumar Mishra, A., and Sharma, S. (2023). Agronomic and environmental determinants of direct seeded rice in south asia. *Circul. Econo. Sustain.* 3, 253–290. doi: 10.1007/s43615-022-00173-x

Chen, L., Gu, G., Wang, C., Chen, Z., Yan, W., Jin, M., et al. (2021). Trp548Met mutation of *acetolactate synthase* in rice confers resistance to a broad spectrum of ALS-inhibiting herbicides. *Crop J.* 9, 750–758. doi: 10.1016/j.cj.2020.11.003

Chen, Z. J., Liu, J., Zhang, N., and Yang, H. (2022). Identification, characterization and expression of rice (*Oryza sativa*) acetyltransferase genes exposed to realistic environmental contamination of mesotrione and fomesafen. *Ecotoxicol. Environ. Saf.* 233, 113349. doi: 10.1016/j.ecoenv.2022.113349

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 6, 80–92. doi: 10.4161/fly.19695

Devos, Y., Reheul, D., De Schrijver, A., Cors, F., and Moens, W. (2006). Management of herbicide-tolerant oilseed rape in Europe: a case study on minimizing vertical gene flow. *Environ. Biosafety Res.* 4, 217–222. doi: 10.1051/ebr:2005001

He, Y., Sun, S., Zhao, J., Huang, Z., Peng, L., Huang, C., et al. (2023). UDP-glucosyltransferase *OsUGT75A* promotes submergence tolerance during rice seed germination. *Nat. Commun.* 14, 2296. doi: 10.1038/s41467-023-38085-5

Hu, P., Wen, Y., Wang, Y., Wu, H., Wang, J., Wu, K., et al. (2021). Identification and characterization of *short crown root 8*, a temperature-sensitive mutant associated with crown root development in rice. *Int. J. Mol. Sci.* 22, 9868. doi: 10.3390/ijms22189868

Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314. doi: 10.1080/10618600.1996.10474713

James, D., Borphukan, B., Fartyal, D., Ram, B., Singh, J., Manna, M., et al. (2018). Concurrent overexpression of *OsGS1;1* and *OsGS2* genes in transgenic rice (*Oryza sativa* L.): Impact on tolerance to abiotic stresses. *Front. Plant Sci.* 9, 1–19. doi: 10.3389/fpls.2018.00786

Jin, M., Chen, L., Deng, X. W., and Tang, X. (2022). Development of herbicide resistance genes and their application in rice. *Crop J.* 10, 26–35. doi: 10.1016/j.cj.2021.05.007

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348. doi: 10.1038/ng.548

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 4. doi: 10.1186/1939-8433-6-4

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77, 521–555. doi: 10.1146/annurev.biochem.76.061005.092322

Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., et al. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nat. Plants* 2, 1–6. doi: 10.1038/nplants.2016.139

Li, M.-X., Yeung, J. M. Y., Cherny, S. S., and Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant *p*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131, 747–756. doi: 10.1007/s00439-011-1118-2

Maeda, H., Murata, K., Sakuma, N., Takei, S., Yamazaki, A., Karim, M. R., et al. (2019). A rice gene that confers broad-spectrum resistance to b-triketone herbicides. *Science* 365, 393–396. doi: 10.1126/science.aax0379

Okuzaki, A., Shimizu, T., Kaku, K., Kawai, K., and Toriyama, K. (2007). A novel mutated acetolactate synthase gene conferring specific resistance to pyrimidinyl carboxy herbicides in rice. *Plant Mol. Biol.* 64, 219–224. doi: 10.1007/s11103-007-9146-y

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633

Perotti, V. E., Larran, A. S., Palmieri, V. E., Martinatto, A. K., Alvarez, C. E., Tuesca, D., et al. (2019). A novel triple amino acid substitution in the EPSPS found in a high-level glyphosate-resistant Amaranthus hybridus population from Argentina. *Pest Manage. Sci.* 75, 1242–1251. doi: 10.1002/ps.2019.75.issue-5

Peterson, M. A., Collavo, A., Ovejero, R., Shivrain, V., and Walsh, M. J. (2018). The challenge of herbicide resistance around the world: a current summary. *Pest Manage. Sci.* 74, 2246–2259. doi: 10.1002/ps.2018.74.issue-10

Raj, A., Stephens, M., and Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350

Ren, Y., Liu, B., Jiang, H., Cheng, W., Tao, L., Wu, K., et al. (2023). Precision editing of *GLR1* confers glufosinate resistance without yield penalty in rice. *Plant Biotechnol. J.* 21, 2417–2419. doi: 10.1111/pbi.v21.12

Sammons, R. D., and Gaines, T. A. (2014). Glyphosate resistance: State of knowledge. *Pest Manage. Sci.* 70, 1367–1377. doi: 10.1002/ps.2014.70.issue-9

Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Software Code Snippets* 16, 1–9. doi: 10.18637/jss.v016.c03

Tian, Y. S., Xu, J., Zhao, W., Xing, X. J., Fu, X. Y., Peng, R. H., et al. (2015). Identification of a phosphinothricin-resistant mutant of rice glutamine synthetase using DNA shuffling. *Sci. Rep.* 5, 1–11. doi: 10.1038/srep15495

Tiwari, P., Sangwan, R. S., and Sangwan, N. S. (2016). Plant secondary metabolism linked glycosyltransferases: An update on expanding knowledge and scopes. *Biotechnol. Adv.* 34, 714–739. doi: 10.1016/j.biotechadv.2016.03.006

Wang, T., Li, X., Liu, X., Yang, X., Li, Y.j., and Hou, B.k. (2023). Rice glycosyltransferase gene *UGT2* functions in salt stress tolerance under the regulation of bZIP23 transcription factor. *Plant Cell Rep.* 42, 17–28. doi: 10.1007/s00299-022-02933-3

Wang, T., Ma, Y.q., Huang, X.x., Mu, T.j., Li, Y.j., Li, X.k., et al. (2021). Overexpression of *OsUGT3* enhances drought and salt tolerance through modulating ABA synthesis and scavenging ROS in rice. *Environ. Exp. Bot.* 192, 104653. doi: 10.1016/j.envexpbot.2021.104653

Xia, D., Zhou, H., Wang, Y., Ao, Y., Li, Y., Huang, J., et al. (2022). *qFC6*, a major gene for crude fat content and quality in rice. *Theor. Appl. Genet.* 135, 2675–2685. doi: 10.1007/s00122-022-04141-9

Xu, S., Fei, Y., Wang, Y., Zhao, W., Hou, L., Cao, Y., et al. (2023). Identification of a seed vigor–related QTL cluster associated with weed competitive ability in direct–seeded rice (*Oryza sativa* L.). *Rice* 16, 45. doi: 10.1186/s12284-023-00664-x

Xu, R., Liu, X., Li, J., Qin, R., and Wei, P. (2021). Identification of herbicide resistance *OsACC1* mutations via in planta prime-editing-library screening in rice. *Nat. Plants* 7, 888–892. doi: 10.1038/s41477-021-00942-w

Zhang, C., Zhong, X., Li, S., Yan, L., Li, J., He, Y., et al. (2023). Artificial evolution of *OsEPSPS* through an improved dual cytosine and adenine base editor generated a novel allele conferring rice glyphosate tolerance. *J. Integr. Plant Biol.* 65, 2194–2203. doi: 10.1111/jipb.13543

Zhao, H., Li, J., Yang, L., Qin, G., Xia, C., Xu, X., et al. (2021). An inferred functional impact map of genetic variants in rice. *Mol. Plant* 14, 1584–1599. doi: 10.1016/j.molp.2021.06.025

Zhao, H., Yao, W., Ouyang, Y., Yang, W., Wang, G., Lian, X., et al. (2015). RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res.* 43, D1018–D1022. doi: 10.1093/nar/gku894

Zhou, H., Li, P., Xie, W., Hussain, S., Li, Y., Xia, D., et al. (2017). 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, H., Xia, D., Li, P., Ao, Y., Xu, X., Wan, S., et al. (2021a). Genetic architecture and key genes controlling the diversity of oil composition in rice grains. *Mol. Plant* 14, 456–469. doi: 10.1016/j.molp.2020.12.001

Zhou, H., Xia, D., Zhao, D., Li, Y., Li, P., Wu, B., et al. (2021b). The origin of *Wxla* provides new insights into the improvement of grain quality in rice. *J. Integr. Plant Biol.* 63, 878–888. doi: 10.1111/jipb.13011

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 October 2024
DOI 10.3389/fpls.2024.1451749

OPEN ACCESS

EDITED BY
Ming Luo,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

REVIEWED BY
Hanna Bolibok-Bragoszewska,
Warsaw University of Life Sciences, Poland
Hao Zhou,
Sichuan Agricultural University, China

*CORRESPONDENCE
Jochen C. Reif
reif@ipk-gatersleben.de

RECEIVED 19 June 2024
ACCEPTED 30 August 2024
PUBLISHED 02 October 2024

CITATION
Berkner MO, Jiang Y, Reif JC and Schulthess AW (2024) Trait-customized sampling of core collections from a winter wheat genebank collection supports association studies.
Front. Plant Sci. 15:1451749.
doi: 10.3389/fpls.2024.1451749

Trait-customized sampling of core collections from a winter wheat genebank collection supports association studies

Marcel O. Berkner, Yong Jiang, Jochen C. Reif* and Albert W. Schulthess

Breeding Research Department, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany

Subsampling a reduced number of accessions from *ex situ* genebank collections, known as core collections, is a widely applied method for the investigation of stored genetic diversity and for an exploitation by breeding and research. Optimizing core collections for genome-wide association studies could potentially maximize opportunities to discover relevant and rare variation. In the present study, eight strategies to sample core collections were implemented separately for two traits, namely susceptibility to yellow rust and stem lodging, on about 6,300 accessions of winter wheat (*Triticum aestivum* L.). Each strategy maximized different parameters or emphasized another aspect of the collection; the strategies relied on genomic data, phenotypic data or a combination thereof. The resulting trait-customized core collections of eight different sizes, covering the range between 100 and 800 accession samples, were analyzed based on characteristics such as population stratification, number of duplicate genotypes and genetic diversity. Furthermore, the statistical power for an association study was investigated as a key criterion for comparisons. While sampling extreme phenotypes boosts the power especially for smaller core collections of up to 500 accession samples, maximization of genetic diversity within the core collection minimizes population stratification and avoids the accumulation of less informative duplicate genotypes when increasing the size of a core collection. Advantages and limitations of different strategies to create trait-customized core collections are discussed for different scenarios of the availability of resources and data.

KEYWORDS

core collections, genebank genomics, association study, plant genetic resources, wheat

1 Introduction

About 7.4 million accessions of crop varieties, landraces, and crop-wild relative species are preserved in more than 1,750 *ex-situ* genebanks worldwide (FAO, 2010). Cereal crops dominate genebank collections (Ramirez-Villegas et al., 2022), with wheat (*Triticum* spp.), rice (*Oryza* spp.), and barley (*Hordeum* spp.) being on the leading positions with more than 850 thousand, 770 thousand, and 460 thousand accessions, respectively (FAO, 2010). These accessions are suspected to harbor beneficial alleles which are associated with an agronomic trait and got accidentally vanished from breeding germplasm due to selective bottlenecks (Tanksley and McCouch, 1997). For instance, 76% and 71% of the total variation in landraces of barley and respectively, wheat is already preserved in genebank collection (Ramirez-Villegas et al., 2022). The identification of donor genotypes retaining these beneficial alleles is vital to exploit the potential of genebanks for breeding. Nevertheless, this task remains difficult even though the results of phenotypic screenings and cost-effective high throughput genotyping are becoming progressively available for many collections. Genome-wide association studies (GWAS) benefit from, or even require, high-quality phenotyping and genotyping. At the scale of an entire collection, the all-encompassing evaluation of every single accession remains however unrealistic in terms of resources required and the sheer number of accessions. This aspect is especially valid for traits demanding expensive phenotyping protocols or for low-heritable traits as well as high-density genotyping such as whole-genome-sequencing. Consequently, more precise examination should be restricted to a targeted selection of accessions: a core-collection (CC).

Since the concept of CC was introduced (Frankel, 1984), CC have been generated based on many genebank collection for crops such as barley (Muñoz-Amatriaín et al., 2014; Milner et al., 2019), groundnut (*Arachis hypogaea* L.) (Upadhyaya et al., 2003), pepper (*Capsicum* spp.) (Lee et al., 2016), potato (*Solanum tuberosum* L.) (Esnault et al., 2016), soybean (*Glycine max* (L.) Merr.) (Bandillo et al., 2015; Haupt and Schmid, 2020), and wheat (Pascual et al., 2020; Phogat et al., 2021). Analyzing the literature spanning a recent decade, Gu and collaborators (Gu et al., 2023) discovered CC for 146 species of crops, ornamentals and trees. In general, there are two contrasting aims for selecting a CC: to represent the diversity harbored in the total collections or to maximize the diversity represented (Marita et al., 2000). Regardless of the aim, sampling strategies of CC have relied on various types of pre-existing or newly generated data such as genotypic (Esnault et al., 2016; Milner et al., 2019; Haupt and Schmid, 2020; Guo et al., 2022) or phenotypic data (Upadhyaya et al., 2003; Phogat et al., 2021); occasionally, involving a stratification based in passport data (Upadhyaya et al., 2003). Moreover, sampling of a CC may face restrictions due to the attributes of the genebank collection, the available resources, and the nature of the crop species *per se*. Certainly, there is not a uniform and universal sampling strategy for a CC (Gu et al., 2023).

Even though many CCs have been analyzed with modern biometric approaches such as GWAS (Muñoz-Amatriaín et al., 2014; Bandillo et al., 2015; Milner et al., 2019), hardly any CC has been reported to be specifically optimized for this approach. Therefore, sampling strategies for CC needs to be re-evaluated and possibly redefined in order to increase the statistical power for the identification of marker-trait-associations that are meaningful for breeding. Based on theoretical knowledge, such a CC should display specific characteristics which are summarized below.

First, the selection of accessions should be enriched with distinct alleles associated with the trait of interest and not yet present in the breeding germplasm. Extremely rare alleles with moderate effect-sizes will probably remain unnoticed in an association study; however, common variants are most likely already present in breeding programs and therefore, not of interest for the identification of donor genotypes. Generally, this also implies that distinct sets of accessions would be sampled for every trait since genetic architectures are in the best case only partially shared and mostly mismatched between different non-correlated traits. To distinguish from the classical CC, this study will refer to a trait-customized core collection (TCCC).

Second, the characteristics of such a TCCC should enable high statistical power in GWAS. The statistical power is defined as the probability to prevent a type 2 error; meaning the probability to correctly identify a marker's association with the trait. The estimates of statistical errors depend on the probability distributions of the null hypothesis and the alternative hypothesis as well as the chosen level of significance. While choosing a more relaxed significance level will increase the power of GWAS, the specificity of a GWAS will decline. Therefore, changing the probability distributions by the constitution of the examined panel remains the only adequate way of adjusting the statistical power. Using a Q + K model, Wang and Xu (2019) demonstrated how the power can be deduced from the non-centrality parameter of the χ^2-distribution. Proven based on simulated data, the determinants of the power are the number of genotypes in the sample, the polygenic contribution to the phenotypic variance, and the effect size of the targeted quantitative trait locus (Wang and Xu, 2019). In contrast, the present work aimed at a practical-oriented evaluation based on real datasets. Sampling extreme phenotypes, also known as selective genotyping (Van Gestel et al., 2000; Xing and Xing, 2009), is known to leverage the statistical power in GWAS (Xing and Xing, 2009; Guey et al., 2011) and was therefore implemented in five out of eight sampling strategies for TCCCs tested in the present report.

Third, population structure and genetic relationship confounded with the variation in the targeted trait can strongly interfere with GWAS results (Myles et al., 2009). Related genotypes have many genetic variants in common; vice versa, genotypes from different subpopulations share very few variants. While all of these shared or unshared variants explain the relatedness, only few variants are truly associated with the trait of interest. Therefore, a higher number of false positive associations will be discovered if population structure and genetic relationships are ignored. Myles and collaborators (Myles et al., 2009) concluded that the covariance of phenotypic similarity and relatedness describes this impact on association mapping and therefore, the correlations between phenotypic and genotypic distance can arguably be considered as a measure of population stratification. The selection of a TCCC

should limit the impact of population stratification, for example by selecting accessions from the entire diversity space of the collections. Active selection against stratification is, however, complex and may negatively influence other aims such as a high number of novel beneficial variants in the TCCC.

Increasing the number of accessions in the TCCC could mitigate the negative impact of population stratification; simultaneously, the power of GWAS will as a tendency increase with the number of accessions (Wang and Xu, 2019). Moreover, the number of distinct positively associated alleles in the TCCC should approach a collection-specific maximum as the number of accessions in the TCCC increases. In general, an unlimited increase in the number of accessions would therefore be an easy, though unrealistically expensive approach to address all problems identified above. Not only the costs of high-quality phenotyping and genotyping, but also the cost for maintaining a TCCC in pure constant conditions on the long-term, urge the need to limit the size of a TCCC to the effective minimum. In addition, a larger TCCC will increase the chance of sampling accessions which are genetic duplicates. Some genebank collections have been reported to harbor duplicate accessions with for instance an abundance of 37% (Schulthess et al., 2022a) and even up to 54% (Singh et al., 2019), respectively. Undoubtedly, the size of a TCCC has therefore an extreme impact. Sizes of CC have mostly been defined relatively to the size of the original genebank collection. Brown (1989a) recommended a proportion of 10% of all available accessions and found to cover about 70% of the entire genetic variations with this CC size. Depending on the collection, these proportions can in practice vary a lot (Van Hintum et al., 2000) and can be translated into a few or more than thousand accessions (Gu et al., 2023). We advance the hypothesis that the composition of the TCCC, as defined by the underlying sampling strategy, determines the success of the identification of marker-trait-associations and the corresponding donor genotypes, given a constant size of the TCCC.

The main goal of the present study was to compare the effect of different sampling strategies on the probability to identify donor genotypes for two agronomic traits in the winter wheat collections of the German Federal *Ex situ* Genebank hosted at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK Genebank). In particular, the objectives were to (1) elaborate eight different selection strategies, relying on phenotypic and/or genotypic information, to sample accessions for TCCCs of increasing size for two traits, namely yellow rust susceptibility (YR) and stem lodging (SL), (2) study the suitability of the strategies to identify association in GWAS based on power estimates of trait-specific panels of markers, genetic and phenotypic diversity in the TCCC as well as population stratification, and (3) investigate the impact of the size of the TCCC on the power of quantitative trait locus detection.

2 Materials and methods

2.1 Plant material

The presented research relied on a previously published subset of the *Triticum* collection stored at the IPK Genebank (Schulthess et al., 2022a, b). In total, 7,651 accessions were selected with the intention to include most of the available winter wheat accessions. While the vast majority were *Triticum aestivum* L., nine accessions were classified as other species of the *Triticum* genus.

2.2 Genotypic data

Genotyping-by-sequencing profiles for all 7,651 accessions were used as previously published by Schulthess and collaborators (Schulthess et al., 2022a, b). Concisely, genotypic profiles were generated by engaging a genotyping-by-sequencing approach as follows. On the field, 7,745 distinct morphotypes were identified among all 7,651 genebank accessions. Each morphotype was transferred into an isolate line by bagging one characteristic ear and propagation in an ear-to-row fashion; the isolate line is referred to as accession sample. For each isolate line, DNA was extracted from a 10-day-old seedling and subsequently digested using PstI and MspI restriction enzymes followed by ligation to adapters comprising barcode sequences. Digested samples were pooled into genotyping-by-sequencing libraries and sequencing was performed either on an Illumina Hiseq-2500 or a NovaSeq 6000 system. The resulting reads were trimmed to a minimum read length of 30 bp followed by SNP calling based on the wheat reference genome var. Chinese Spring v. 1.0 (IWGSC, 2018). The resulting SNP matrix was further filtered: only markers with less than 10% missing values, any of both alleles in homozygous state in at least 10 genotypes, and heterozygosity ≤1% were retained for downstream analyses. Finally, missing genotypic data was imputed marker-wise as the predominant allele among all accession samples.

2.3 Phenotypic records and performance estimation

Two phenotypic traits were considered in the present work: susceptibility to yellow rust infections (YR) and stem lodging (SL). Both traits were measured by following the standard protocols of the German Federal Plant Variety Office (Bundessortenamt, 2020) and based on 1-9 scoring scales, with 1 corresponding to YR resistance or SL tolerance, and 9 indicating extreme YR or SL susceptibility. YR and SL were recorded in 12 and 13 field experiments, respectively, with ten experiments having records for both traits simultaneously. YR was based on natural infections and recorded when sufficient disease pressure was observed for an entire experiment while SL was measured after heading stage. Field experiments involved two German locations, Gatersleben (latitude 51° 49' 19.74" N, longitude 11° 17' 11.80" E) and Schackstedt (latitude 51° 43' 0" N, longitude 11° 37' 0" E), as well as seven consecutive growing seasons within the 2014-2020 period. Each experiment was performed to large-scale test 1,428-1,698 entries under an alpha-lattice design and considering a 0.4 m^2 plot as experimental unit. All experiments were performed considering two replications, but in two experiments phenotypes were recorded in only one replication. Experiments were partially connected through overlapping common entries. Crop management of all experiments

considered only the chemical control against weeds, while no fertilizers were additionally applied. Data quality assessment as well as the computation of variance components for heritability estimation and the average performance across experiments, best linear unbiased estimates (BLUEs) for each trait and entry were performed based on linear mixed models as implemented in the R software package ASReml-R (v. 3) (Butler et al., 2009). Further details on methods underlying these analyses can be found in past works (Schulthess et al., 2022a, b), where YR phenotypic records were originally published. SL data has not been either presented or assessed previously. Out of the 7,745 accession samples with genotyping-by-sequencing profiles, 6,300 and 6,251 have BLUEs for YR and SL, respectively, and were used for downstream analyses.

2.4 Evaluation of genetic diversity and population structure

Genetic diversity within groups of accessions was analyzed based on genetic distances. With this intention, pairwise Rogers' distances were calculated (Rogers, 1972) for all 7,745 accession samples and condensed by principal coordinate (PCo) analysis (Gower, 1966). Population structure was visualized by plotting first and second PCos against each other.

2.5 Tested sampling strategies

Eight different strategies to sample a TCCC were contrasted in the present study. The tested strategies differed with respect to the data sources required for the selection decision as outlined in Table 1. While four sampling strategies relied on phenotypic information, one strategy required exclusively genomic profiles and additional two strategies engaged both types of data. For the latter, phenotypic data, as BLUEs per accession, were linked to the existent genotypic data through accession samples. Only one strategy did not need any data: completely random sampling (All_random) of accessions was included as benchmark for the most simplistic approach. In the second sampling strategy (All_Pdiv), the range of phenotypic values of all accession samples was subdivided into 10 equally-large quantiles, while the total sample size was subdivided into individual drawings which were randomly assigned to the 10 defined quantiles. During implementation, restrictions due to fewer representatives toward the extremes of distributions must be considered. Therefore, if the number of drawings was larger than the total number of representatives within the respective quantile, the surplus was randomly distributed to the other quantiles. Thereafter, accession samples were randomly sampled within the quantiles according to the number of drawings. In this way, the whole range of phenotypic diversity was covered as equally as possible. For the third sampling strategy (All_Gdiv), the R software package corehunter (v. 3.2.1) (De Beukelaer and Davenport, 2018) was used to maximize the genetic diversity within the TCCC. The algorithm, as implemented in the "sampleCore" function, was applied with the specification EN (entry-to-nearest-entry distance) in combination with MR (modified Roger's distance) and allowing a maximum of 20 iterations without any improvement. This function run in combination with rJava (v. 1.0-6) (Urbanek, 2021).

TABLE 1 Description of the eight strategies applied for sampling of trait-customized core collections.

Name	Description	Phenotypic data	Genotypic data
All_random	Completely random sampling from all accession samples	No	No
All_Pdiv	Sampling phenotypes equally covering the entire range of the phenotypic diversity	Yes	No
All_Gdiv	Sampling from all accession samples in order to maximize genetic diversity	No	Yes
1T_rank	Sampling from the tail of best-performing accession based on phenotypic ranking	Yes	No
2T_rank	Sampling the most contrasting phenotypes from two tails based on phenotypic ranking	Yes	No
2T_random	Random sampling of accession samples from two phenotypic tails	Yes	No
2T_Gdiv&Gsim	Maximizing genetic diversity within the tail of best-performing accession and sampling of related genotypes from the contrasting tail	Yes	Yes
2T_Gdiv&Gdiv	Maximizing genetic diversity independently within both contrasting phenotypic tails	Yes	Yes

Depicted are the abbreviated names of the strategies, a short description of the underlying procedure as well as an indication about the required data.

In the fourth sampling strategy (1T_rank), all accessions were ranked based on the phenotypic values and the most resistant accessions, thus with the lowest BLUEs, were successively selected. Similarly, the sampling strategy (2T_rank) involved a ranking of accessions based on their phenotypic values. However, accessions were chosen from the lowest and highest phenotypic range in a 3:1 ratio. For both traits, low phenotypic scores are the breeding target and thus, this last step allows the accumulation of beneficial variants within the TCCC.

For the remaining three strategies, a positive and a negative tail were defined based on the distribution of BLUEs. Each phenotypic tail contained the 10% of all accession which have the lowest and highest phenotypic values. Tails with low and high values are

referred to as positive and negative tail, respectively. Random sampling in the positive and negative tail was performed for the strategy called 2T_random, while the tails were considered in a 3:1 ratio. For the last two tested sampling strategies, the R software package corehunter (v. 3.2.1) (De Beukelaer and Davenport, 2018) was engaged to maximize the genetic diversity among the accessions chosen from the positive tail. For this maximization, the same software specifications were used as described above for All_Gdiv. For the seventh sampling strategy (2T_Gdiv&Gsim), accessions were sampled from the negative tail in a 3:1 ratio in such a way that one accession from the negative tail was chosen due to the lowest Rogers' distances with three accessions of the positive tail. This stepwise selection was repeated until all accessions from the positive tail were covered. For the eighth sampling strategy (2T_Gdiv&Gdiv), accessions were initially sampled from the negative tail based on a low genetic distance in exactly the same number as accessions were present in the positive tail. In a second step, the number of accessions in the pre-selection from the negative tail was reduced in order to achieve the 3:1 ratio between the samples from the positive and negative tail. For the latter step, the R software package corehunter (v. 3.2.1) (De Beukelaer and Davenport, 2018) was used as described above to sample a lower number of accessions in parallel to maximizing the genetic diversity within the sample from the negative tail.

The described sampling strategies were implemented for eight different sizes of TCCCs, measured in the number of accession samples included. The sizes of TCCCs increased with a step size of 100. The minimum size was set at 100 because this is the widely-accepted threshold in GWAS to yield publishable results; in contrast the maximum of 800 reflects the size of the phenotypic tails and the 3:1 sampling ratio between both tails in some sampling strategies. The sampling and later evaluation was performed with 50 independent replications for each combination of strategy and size of TCCC.

2.6 Evaluation of trait-customized core collections

The characteristics of the TCCCs differed depending on the compositions of selected accession samples. Unbiased comparisons of TCCCs and the associated sampling strategies were ensured by only contrasting TCCCs of the same size. The comparison was based on six criteria which were namely the phenotypic distribution, the correlation between phenotypic and genotypic distances, the number of duplicate genotypes based on identity-by-state values between sampled accessions, the genetic distinctness of the sampled TCCC from the remaining accessions samples (F_{st}), the statistical power of GWAS (Wang and Xu, 2019), and the average minor-allele-frequency (MAF) within the TCCC and within the trait-specific panels of markers which are later on referred to as Top10_MTAs.

For each trait, the Euclidean distances were computed separately for all pairs of accession samples in the TCCC or in the entire collection and the correlation between these values and the pairwise Rogers' distances from genomic data served as diagnostic measure for population stratification. Identity-by-state values were calculated for all pairwise comparisons within the respective TCCC with the "snpgdsIBS" function of the R software package SNPRelate (v. 1.24.0) (Zheng et al., 2012). A threshold of identity-by-state values > 99% was applied to declare two accession samples as duplicates, as similarly done by Schulthess and collaborators (2022a). The pairwise F_{st} was calculated between all accession samples in the TCCC and all remaining accession samples of the genebank collection. The pairwise F_{st} was calculated with the "fs.dosage" function of the R software package hierfstat (v. 0.5-11) (Goudet and Jombart, 2022). The effective population size (N_e) was calculated within each TCCC based on (Waples, 2006):

$$N_e = \frac{1}{3(r^2_{linkage} - \frac{1}{n})} \quad (1)$$

in which n is the number of sampled genotypes and $r^2_{linkage}$ denotes the mean linkage disequilibrium of unlinked markers in the sampled genotypes. In order to ensure the unlinked state, linkage disequilibrium was calculated as the average squared correlation between markers located on different chromosomes.

The statistical power of GWAS was estimated following the theoretical approach of Wang and Xu (2019). More precisely, we first performed GWAS in the total set of accession samples and took the 10 most significant markers as hypothetical QTL for each trait, referred as trait-specific Top10_MTAs. The computations for GWAS relied on the "GWAS" function of the R software package rrBLUP (v.4.6.1) (Endelman, 2011); the minimum MAF was set to 1%. Then, the approach of Wang and Xu (2019) was used to estimate the power of detecting these markers in the TCCCs. For a GWAS study relying on a Q+K model and involving n genotypes, the power of detecting a specific marker can be estimated based on the non-centrality parameter, δ, of a χ^2-distribution, as follows (Wang and Xu, 2019):

$$\delta = (\lambda + 1)\sum_{j=1}^{n} \frac{1}{(d_j\lambda + 1)} \frac{r^2_{marker}}{1 - r^2_{marker}} \quad (2)$$

in which d_j are the eigenvalues of the kinship matrix, r^2_{marker} is the proportion of phenotypic variance explained by the marker estimated within the TCCC and λ represents the ratio of polygenic variance to the residual variance. The genetic variance and residual variance were calculated with the "kin.blup" function of the R software package rrBLUP (v.4.6.1) (Endelman, 2011) and considering all accession samples within the TCCC with BLUEs for the trait of interest. A significance level of 0.05 was applied in combination with the simpleM method (Gao et al., 2008) to account for the bias due to multiple testing. The R software package ggplot2 (v. 3.4.4) (Wickham, 2016) was used for the visualization of the results. Manhattan plots were created with the R software package qqman (v. 0.1.4) (Turner, 2018). All computations were performed in the R environment (v. 3.4.4, v. 4.0.2) (R Core Team, 2020).

3 Results

3.1 Phenotypic distributions and trait's association with population structure

Fifteen large-scale field experiments provided data with very high heritabilities for YR (h^2 = 0.82) and SL (h^2 = 0.86)

(Supplementary Table S1) as well as BLUEs for 6,300 and 6,251 accession samples for YR and SL, respectively. The distributions of the traits were distinct: while BLUEs of YR followed a more symmetric distribution (mean = 4.95, median = 4.79, standard deviation = 1.40), SL had an L-shaped distribution which was strongly skewed towards lower values (mean = 3.18, median = 2.61, standard deviation = 2.03) (Figure 1).

Population stratification can strongly bias GWAS results and thus, the association of the phenotypic distributions with the population structure was investigated for both traits in the entire genebank collection. Strong correlations with the first PCo from genomic distances were observed for both YR (r = 0.42) and SL (r = 0.52), thus indicating a clear association of population structure with trait variation (Table 2). The separation along the first PCo can also be seen graphically based on the location of the phenotypic tails, one with low and another with high BLUEs, in the diversity space (Figure 1). The correlations with the second, third, fourth and fifth PCo were clearly lower in magnitude however still statistically significant (p-values < 0.001).

3.2 Selection of a marker panel showing association in the entire population

Based on all available data, GWAS was performed to identify likely associated markers that were later investigated within the sampled TCCCs. For both traits, some markers were identified as having a high probability of being associated with the trait of interest (Supplementary Figure S1) and the 10 markers with the highest -$\log_{10}(p)$-value were selected for a trait-specific marker panel and referred to as Top10_MTAs. For both traits, these markers were located on several distinct chromosomes (Supplementary Figure S2). The markers of the Top10_MTAs panels were later categorized based on the effect size, i.e. large,

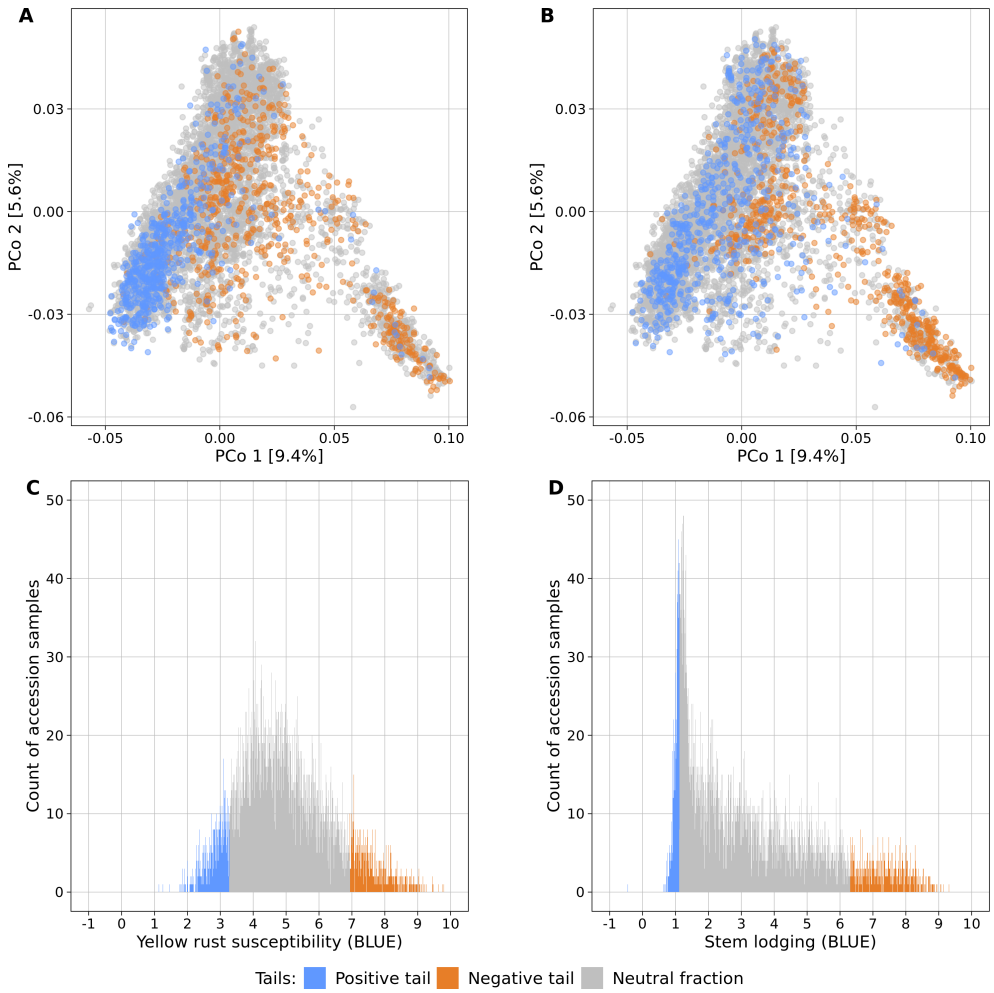


FIGURE 1
Genetic diversity and phenotypic distributions of the two contrasting phenotypic tails each containing 10% of the accessions samples with the lowest (blue) and highest (orange) phenotypic values of yellow rust susceptibility **(A, C)** and stem lodging **(B, D)**. The remaining 80% of all accession samples are shown in gray color. The phenotypic distributions are plotted in relation to the genetic diversity **(A, B)** which is displayed as a biplot of the first and second principal coordinates (PCo) from the Rogers' distances between 7,745 accession samples of the IPK winter wheat collection. Histograms display the abundance of best linear unbiased estimates (BLUE) for yellow rust susceptibility **(C)** and stem lodging **(D)** along the phenotypic range. The phenotypic values exceed the 1-9 range of the phenotyping scale due to the non-orthogonal structure of characterization experiments across locations and years and the computation of linear mixed models.

10.3389/fpls.2024.1451749

TABLE 2 Correlation coefficients for the relation between the five most-explanatory principal coordinates (PCo) calculated from the Rogers' distances between 7,745 accession samples of the IPK winter wheat collection and the best linear unbiased estimates for the traits yellow rust susceptibility (YR for 6,300 accession samples) and stem lodging (SL for 6,251 accession samples).

	PCo 1	PCo 2	PCo 3	PCo 4	PCo 5
	(9.42%)	(5.63%)	(3.00%)	(2.06%)	(1.89%)
YR	0.4225***	0.0473***	-0.2555***	-0.1322***	-0.0984***
SL	0.5196***	0.0878***	0.0444***	0.3167***	-0.1050***

The percentages in brackets display the amount of molecular variance explained by the respective PCo.
***p ≤ 0.001.

medium, and small r^2 values, and the MAF (common, rare) (Supplementary Table S2). With a size of 10 markers, the Top10_MTAs marker panels comprised multiple markers per category; however, not all combinations of effect size and MAF were present in these panels. Common markers with a medium to large effect are often already known to breeders and incorporated in the germplasm. Four and five markers were of this type for YR and SL, respectively. Markers with a small effect demand much more effort from breeding to achieve a measurable improvement of the germplasm; three small effect markers ($r^2 \leq 0.001$) were found for both traits. Markers with medium to large effects which are rather rare are arguably the primarily target of genebank genomics. For YR, markers 6A_135235117 and 2B_57683037 were classified as such while markers 2D_186808122 and 2D_186808096 were identified for SL (Supplementary Table S2). The latter two markers could not be distinguished for the calculation of the r^2 value due to complete collinearity.

3.3 Phenotypic distributions and population stratification resulting from sampling strategies

The phenotypic distributions of the sampled TCCCs showed the expected strong differences between the sampling strategies. As an example, the distributions are presented for two TCCC sizes: 300 and 600 accession samples (Figure 2). The sampling strategies can be clustered into three groups: first, the random sampling (All_random) as well as the sampling along the entire phenotypic range (All_Pdiv) resulted in TCCCs that already cover the available phenotypic diversity despite of the limited number of samples in TCCCs. Similarly, All_Gdiv led to TCCCs that cover the phenotypic range of the entire collection with increasing size of the TCCC. Second, samples resulting from 1T_rank represent a small phenotypic range in combination with overall low values. Third, the phenotypic distribution associated with the strategies which relied on two contrasting phenotypic tails combine in general a broad range of phenotypic values with a high proportion of accessions with low BLUEs and entirely lack intermediate phenotypes. Differences among the latter four strategies can mostly be seen in TCCCs of smaller size since almost the entire positive tail was included for extremely large TCCCs regardless of the strategy.

Within the sampled TCCCs, correlations between phenotypic and genotypic distances were investigated as a possible indicator of population stratification. For both traits, random sampling of accessions (All_random) resulted in a positive correlation which was fairly constant across different sizes of TCCCs. These correlations could be seen as a benchmark for the comparison with the other strategies since they represent the correlation present in the entire population (Figure 3). For the entire collection, the base correlations amount to 0.16 and 0.32 for YR and SL, respectively and any correlation lower than this is arguably a moderation of population stratification in the respective TCCC. For both traits, strong positive correlations were found for the two-tail sampling strategies which disregard genotypic diversity (2T_rank and 2T_random). These strategies did not only yield higher correlation values than All_random but provided also the highest correlation per trait in general: applying 2T_rank to the YR data resulted in a correlation of 0.70 for a TCCC size of 100; for SL, an overall maximum of 0.46 was found in 300 accession samples with the same strategy. Therefore, these strategies probably fortify population stratification. While 2T_Gdiv&Gsim and 2T_Gdiv&Gdiv resulted in TCCCs with low or even negative correlations for smaller TCCCs, positive correlations were found for larger TCCCs with a continuous increase in magnitude as a function of TCCC size. Minimizing phenotypic diversity including only low BLUEs (1T_rank) as well as maximizing genetic diversity regardless of the phenotype (All_Gdiv) results in TCCCs having moderate positive correlations which are lower compared with All_random.

3.4 Genetic diversity and representativity of the sampled TCCC

The genetic diversity within the TCCC and the representativity for the entire genebank collection were evaluated with three indicators that allow to rank the different sampling strategies for a fixed TCCC size. Based on the estimated effective population size (N_e), the sampling strategies could be clustered into two groups (Supplementary Tables S3, S4): Sampling strategies that maximized the genetic diversity by engaging the maximizing algorithm were contrasting to strategies that do not rely on genotypic information. For all evaluated sizes of TCCCs, the All_Gdiv sampling strategy led by far to TCCCs with the highest N_e values. When sampling TCCCs containing 300 accession samples, this sampling strategy led to maximum N_e values of 361 and 355 for YR (Supplementary Table S3) and SL (Supplementary Table S4), respectively. For both 2T_Gdiv&Gsim and 2T_Gdiv&Gdiv, the effect of the genetic maximization on the effective population size was more dominant for SL but less pronounced for YR. Another perspective on the genetic diversity covered within TCCCs is given by the abundance of duplicate genotypes defined by the identity-by-state measure (Figure 4). Differences in the number of duplicates between TCCCs became more obvious for TCCCs of larger size. For YR and SL, differences could be observed starting with TCCC sizes of 200 and 300 accessions, respectively. As a general tendency, sampling strategies which rely on genotypic information sampled a lower number of duplicate genotypes which also demonstrated the

10.3389/fpls.2024.1451749

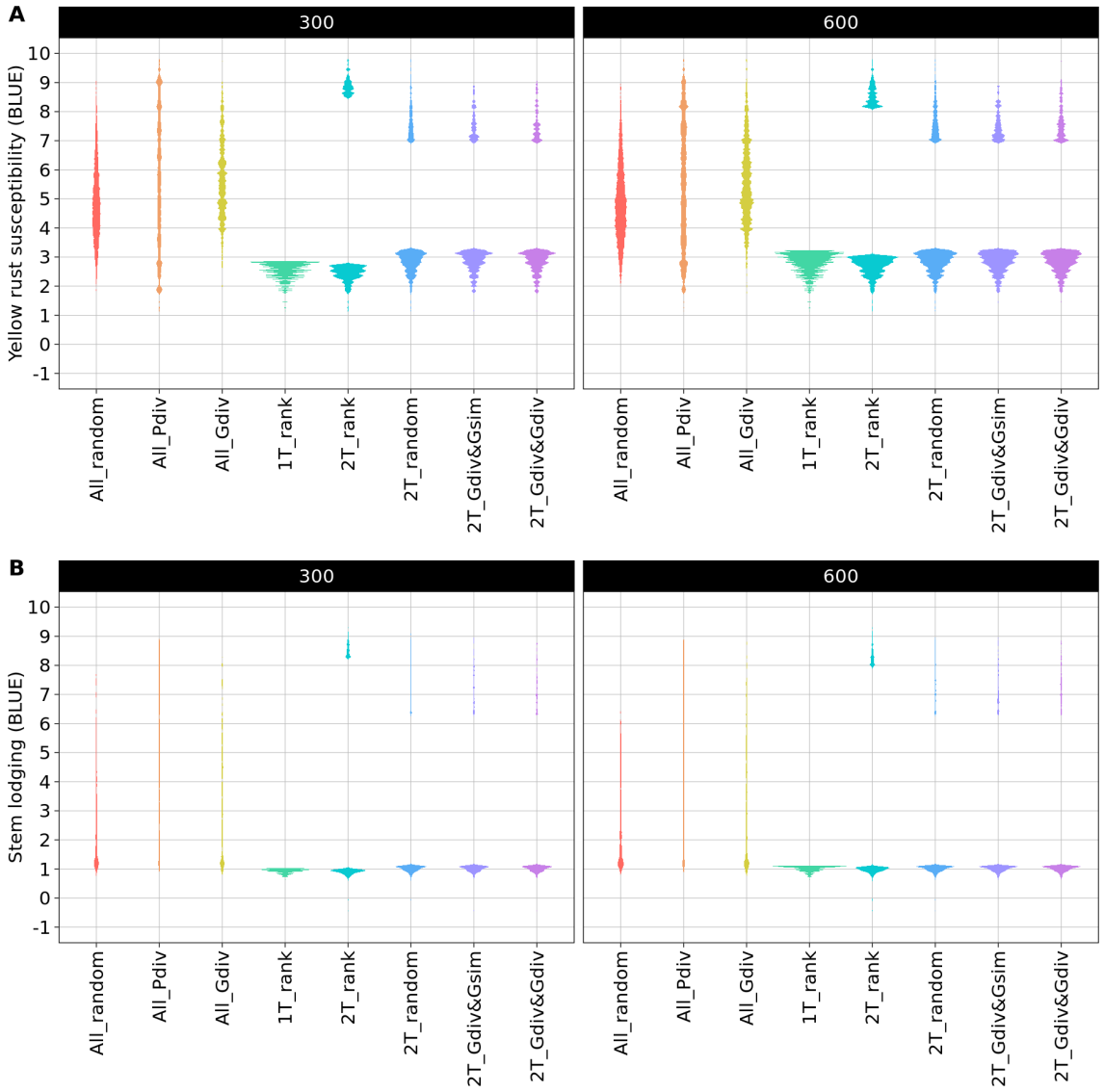


FIGURE 2
Comparison of phenotypic distributions of yellow rust susceptibility **(A)** and stem lodging **(B)** within the trait-customized core collection generated by eight different sampling strategies. Shown are the averaged distributions of best linear unbiased estimates (BLUE) seperately for 300 and 600 included accession samples. The phenotypic values exceed the 1-9 range of the phenotyping scale due to the solving of the linear mixed model.

opposing causal relationship with the higher N_e values. Regardless of the trait, 1T_rank sampled by far the highest number of duplicate genotypes. A high proportion of duplicates was still sampled by applying strategies 2T_rank and 2T_random, especially for YR.

The genetic representativity of the TCCCs was evaluated based on the pairwise F_{st} of the TCCC and all remaining accession samples of the entire collection (Figure 5). Indicated by a value in proximity to zero, random sampling (All_random) led on the one hand to a perfect genetic representation of the entire collection. On the other hand, a distinct genetic distribution in the TCCC was sampled for YR with 1T_rank and for SL with All_Gdiv, while applying 1T_rank for SL resulted in moderate values. These results not only emphasize the impact of considering the phenotypic diversity for the sampling strategy but also highlight the need to focus on trait-customized approaches.

3.5 Statistical power and further requirements for association studies

While the aforementioned measures describe the entire sampled TCCC, measures indicating the capability of identifying association were investigated based on the two trait-specific Top10_MTAs marker panels (Supplementary Table S2). As a prerequisite to identify associations, respective markers need to be in a polymorphic state within the TCCC. The ratio of polymorphic markers within the Top10_MTAs differed between the sampling strategies (Supplementary Tables S5, S6). As expected, more Top10_MTAs markers appeared polymorphic with an increasing size of the TCCC and restrictions in the ratio of polymorphic markers were only present in TCCCs of smaller sizes. While sampling strategy 1T_rank resulted in the overall lowest ratio of

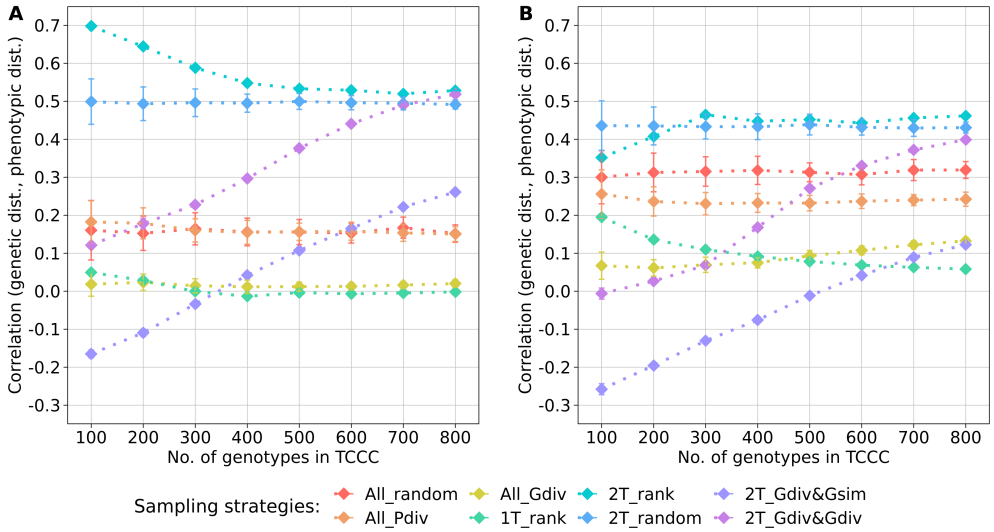


FIGURE 3
Correlation between the genetic distance (Rogers' distance) and the phenotypic distance (Euclidean distance) depending on the number of accession samples included in trait-customized core collections for yellow rust susceptibility **(A)** and stem lodging **(B)**, respectively. Values are depicted separately for eight different sampling strategies representing mean values of 50 independent replications; whiskers display the standard deviations.

polymorphic markers and the sampling strategies 2T_rank, 2T_Gdiv&Gsim and 2T_Gdiv&Gdiv prevented such restriction for all sizes of TCCCs and both traits.

A key parameter of the comparison was the statistical power for the identification of the Top10_MTAs markers. For both traits, the average power across these markers followed a general pattern (Figure 6). Highest power resulted from sampling opposite phenotypic extremes without considering the genotypic information - strategies 2T_rank and 2T_random. Sampling strategies which rely on phenotypic as well as genotypic information showed strong increase in the power with increasing numbers of accession samples. Random sampling (All_random) and sampling from the entire phenotypic range (All_Pdiv) led to moderate high statistical power for 100 accession samples, followed by a strong drop when increasing to 200 accession samples but which did not alter much with increasing sizes of TCCCs further. Sampling only accessions with beneficial phenotypes (1T_rank) predominantly resulted in TCCCs with unfavorable power estimates: the power decreased from a peak value at a TCCC size of 100 with increasing sizes of TCCCs for SL, the power remained in

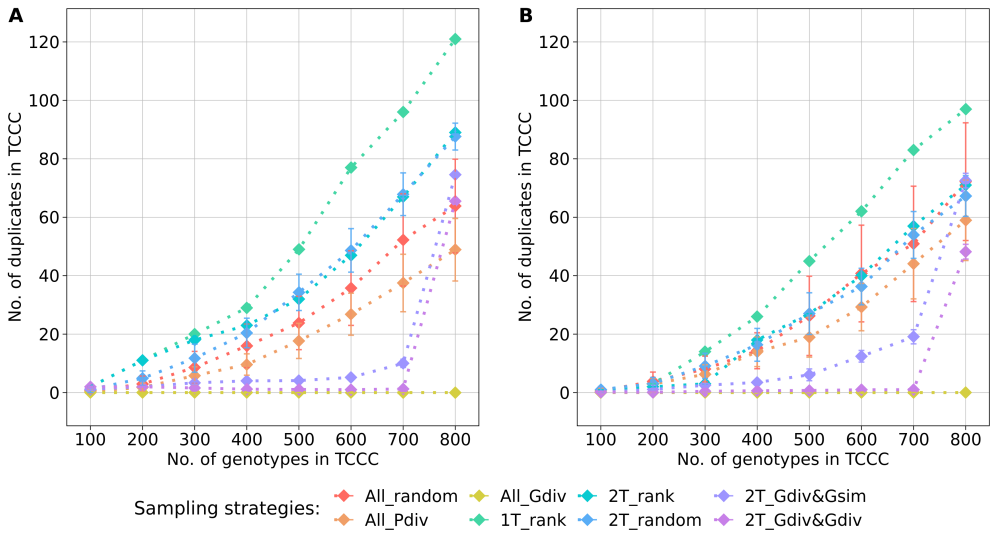


FIGURE 4
Number of duplicate genotypes, defined by an identity-by-state value higher than 0.99, depending on the number of accession samples included in trait-customized core collections (TCCC) for yellow rust susceptibility **(A)** and stem lodging **(B)**, respectively. Values are depicted separately for eight different sampling strategies representing mean values of 50 independent replications; whiskers display the standard deviations.

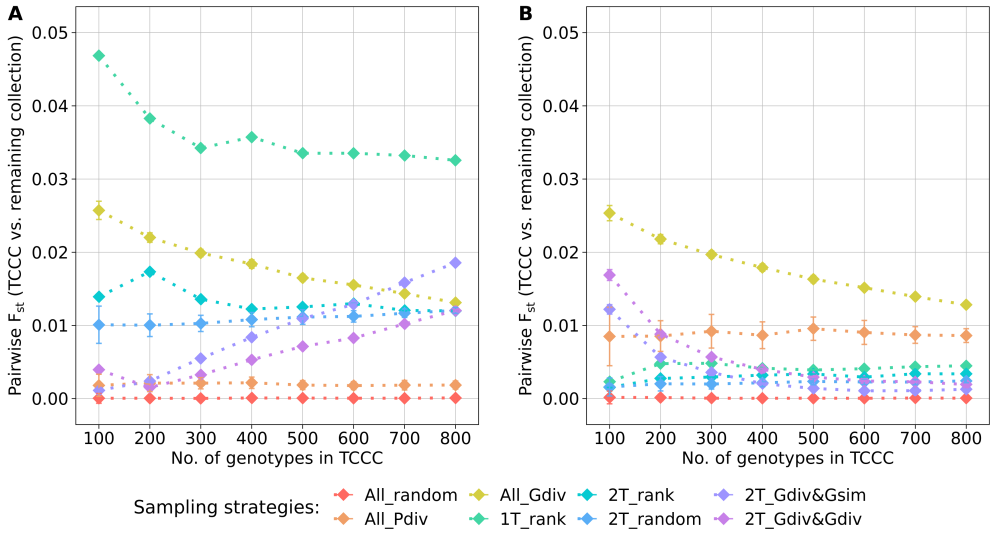


FIGURE 5
Pairwise F_{st} calculated between the trait-customized core collections (TCCC) and the remaining part of the genebank collection depending on the number of accession samples included in trait-customized core collections for yellow rust susceptibility **(A)** and stem lodging **(B)**, respectively. Values are depicted separately for eight different sampling strategies representing mean values of 50 independent replications; whiskers display the standard deviations.

this case constantly low for YR. For TCCCs of smaller size, low values of power might also originate from the fact that the monomorphic state of a marker results in a power of zero and thus, conclusions based on single markers can be diverging from the general trend. Additionally, the proportion of the explained phenotypic variance could not be calculated for some markers within the TCCC for factors such as complete collinearity and thus, power estimates were excluded in this specific case.

Investigation of statistical power separately per marker was dominated by strong fluctuation. As a general trend, the findings described for the averaged power (Figure 6) were also found for the single marker of the Top10_MTAs for SL (Supplementary Figure S3) and for YR (Supplementary Figure S4). Contradicting to the general assumption, decreasing power estimates were found for increasing TCCC sizes for the strategies 2T_rank and T_random in combination with several markers. For SL, the marker-specific MAF of six markers

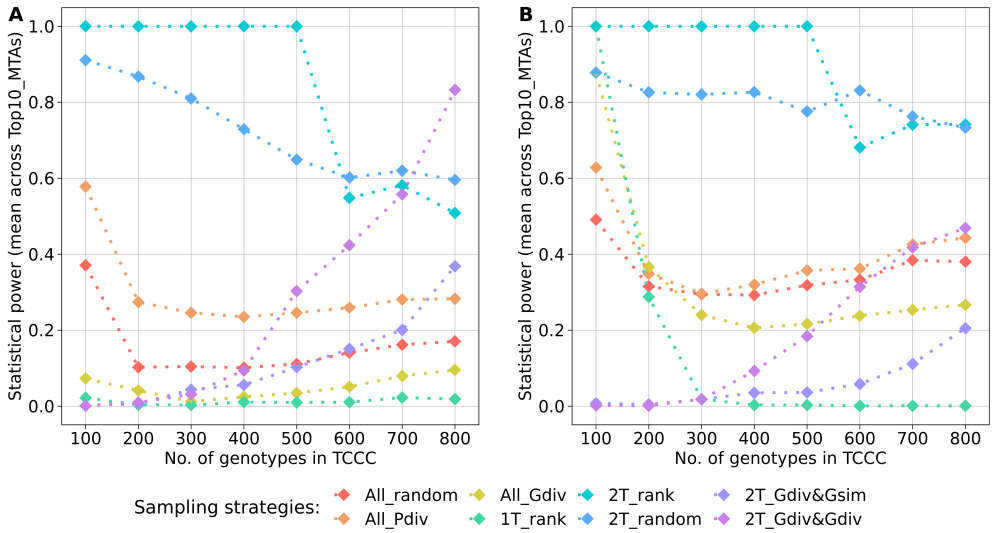


FIGURE 6
Arithmetic mean of the statistical power for the identification of marker-trait-associations calculated based on 10 markers of the Top10_MTAs panel. Power estimates were performed within trait-customized core collections for yellow rust susceptibility **(A)** and stem lodging **(B)**, respectively. Values are depicted for eight different sampling strategies representing mean values of 50 independent replications. For the calculation of the mean, the estimated power of monomorphic markers was considered with a value of zero. The estimated power of a marker was excluded if the proportion of the explained phenotypic variance could not be estimated within a specific trait-customized core collection.

among the Top10_MTAs was relatively high regardless of the sampling which indicate that they are not associated with rare alleles *per se* (Supplementary Figure S5). For YR, low MAF could be observed for multiple markers among the Top10_MTAs regardless of the sampling strategy as well as the size of the TCCC (Supplementary Figure S6). One marker even showed a maximum of MAF below 5% for all combination of sampling strategy and sizes of the TCCC (4A_725504657). These rare variants risk to fall below a MAF-threshold which is typically applied in GWAS studies and thus, might not be detectable.

4 Discussion

The presented comparison of eight strategies to sample a TCCC aimed to provide guidance for improved scrutinization of genebank collections by engaging GWAS. The results suggest that not a single approach can be optimally suited for all traits; however, some strategies, such as 1T_rank, showed generalized disadvantages. In consequence, the choice for a selection strategy is guided by a trade-off between the presented six evaluated criteria. Undoubtedly, the estimated statistical power, as defined by Wang and Xu (2019), is a corner stone of evaluating selection strategies which should later be analyzed engaging GWAS. Nevertheless, a mere focus on statistical power might allow for genetic redundancies in the TCCC and fortify the impact of population stratification. Thus, a case-dependent weighing of criteria is needed.

4.1 Phenotypic contrast boosts power in association studies

Sampling extreme phenotypes for association studies has been reported to leverage detection power (Van Gestel et al., 2000; Xing and Xing, 2009) and similarly, the presented results show an increase in estimated power for TCCCs by incorporating phenotypic extremes (Figure 6). As demonstrated by the comparison of 1T_rank with 2T_rank for YR, the contrast between positive and negative tails is in particular important for GWAS; thus, a mere accumulation of possible desired donor genotypes is hardly effective in this respect. In addition, the variance components used for the calculation had a strong impact: particularly for SL, 2T_rank and 2T_random combined high genetic and low residual variances (Supplementary Figures S7, S8), while 2T_Gdiv&Gsim and 2T_Gdiv&Gdiv resulted in the opposite. Due to the minimization of Rogers' distances between the positive and negative tails in the latter two sampling strategies, variation is assigned as residual variance even though the genetic diversity was maximized within the positive tail. This finding highlights the importance of the relative magnitude of polygenic variance compared with the residual variance. The ratio of polygenic variance to the residual variance is given by λ in Equation 2. The larger the value for λ gets, the greater the non-centrality parameter δ can become and as a consequence, a high statistical power could be expected. Representing phenotypic extremes could result in an enlarged genetic variance and thus in an elevated polygenic contribution to the phenotypic variance. Therefore, one can conclude that data with high genomic heritability has the potential for high statistical power in GWAS.

4.2 Limitations for the identification of rare variants within TCCCs

The identification of rare variants is a well-known problem in GWAS (Myles et al., 2009; Kosmicki et al., 2016) and grouping of genotypes was proposed as a method to increase the frequency of rare variants (Kosmicki et al., 2016). In the present study, sampling extreme phenotypes was implemented following the assumption that accessions with advantageous phenotypes have genotypes that are enriched with rare beneficial variants. In particular for SL, trends confirmed this assumption because the average MAF of the markers in the Top10_MTAs panel were augmented by the strategy 1T_rank, but also mildly augmented by the strategies 2T_rank and 2T_random, compared to All_random (Supplementary Figure S9). This trend was however not similarly present for YR and may even revert to the opposite at the level of individual Top10_MTAs markers, like for instance in case of marker 7A_367972613 (Supplementary Figure S5). A possible explanation could be the contrasting underlying genetic architectures assumed from the Gaussian-like (YR) and L-shaped (SL) distributions. Based on two disease traits in humans, Xing and Xing (2009) investigated the effect of extreme phenotype sampling on GWAS. The authors reported that common variants benefit in particular from this strategy and less prominent effect was reported on rare variants. This could arguably be the case for some of the markers in the Top10_MTAs panel in the present study.

Assuming that a targeted enrichment for rare variants would be difficult for specific traits, a selection strategy increasing the overall MAF could have its merits to avoid problems joint with rare associated variants. For both traits, sampling a TCCC with All_Gdiv clearly increased the overall mean MAF compared to the other sampling strategies (Supplementary Figure S10). The selected accession samples would not be specifically optimized for one trait but rather suits for all traits moderately. On the other hand, such a general approach would demand extremely large TCCC sizes, arguably larger than the sizes tested in the present report, in order to allow for a suitable statistical power and therefore, remains rather theoretical given the limitation in funding and capacities often-faced by genebank institutions.

In addition to the MAF, the effect sizes of a variant determine the success of the identification in GWAS (Kosmicki et al., 2016; Wang and Xu, 2019). Uncommon variants of large or at least medium effect size in the total collection are the only reasonable target of a GWAS in order to justify the high costs of creating and deeply characterizing a TCCC. When focusing on such markers across the Top10_MTAs panel (e.g. YR: 2B_57683037, 6A_135235117; SL: 2D_186808096), the power estimates per sampling strategy revealed that sampling from the phenotypic extremes based on the rank (2T_rank) outcompetes the

other sampling strategies (Supplementary Figures S3, S4). The differences between 2T_rank and 2T_random were however small in these cases. It should be mentioned that a drop in estimated power was observed for the sampling strategy 2T_rank in combination with larger TCCCs for some markers with low effect sizes (e.g. YR: 4A_725504657; SL: 7A_367972613) (Supplementary Figures S3, S4). While the sampling strategy 2T_rank showed higher power only for some particular TCCC sizes for variants with a low effect size, strategies which did not rely on the phenotypic extremes (All_random, All_Pdiv and All_Gdiv) showed an overall low performance in these cases. Similarly, Wang and Xu (2019) found an asymptotic increase of estimated power when increasing the effect size of a simulated QTL in a hybrid population as well as in a set of recombinant inbred lines of rice. These findings highlight that rather the effect size of a variant determines the success of GWAS than the composition of the TCCC. Identification of small effect variants cannot be a predictable aim of creating a TCCC; they will probably remain unnoticed.

4.3 Reduction of population stratification by maximizing genetic diversity

Sampling a TCCC exclusively from the phenotypic extremes increases the magnitude of population stratification exceeding the base level present in the genebank collection. For GWAS in general, this concern was already raised by Guey and collaborators (Guey et al., 2011) and later proven based on simulations by Panarella and Burkett (2019) who reported on a substantial increase in false-discovery rate compared with random sampling. The presented real data scenario is in accordance with these previous findings. Sampling from the two phenotypic extremes, specifically based on the rank (2T_rank), forces a boost in the correlation between phenotype and genotype when compared to the random sampling which is a stable representation of the characteristics of the entire population (Figure 3). Two options of accounting for population stratification were presented here; either minimize the phenotypic diversity (1T_rank) or maximize the genetic diversity (2T_Gdiv&Gsim, 2T_Gdiv&Gdiv and All_Gdiv). While the first option leads to a low averaged power estimate for the Top10_MTAs panel (Figure 6), the latter option seems to be vital to subdue population stratification in genebank genomics.

4.4 Maximizing genetic diversity avoids less informative duplicates

Maximizing the genetic diversity within the TCCC is negatively associated with the number of duplicate genotypes accumulated (Figure 4). As mentioned earlier, duplicate genotypes occur widespread across genebank collections proven for many institutes and crops (FAO, 2010). Without any doubts, the identification of completely identical duplicate genotypes is difficult due to the simplifying nature of marker data and therefore, the definition of a threshold value for similarity will always be disputable. Moreover, the threshold value in the present study might not only include identical but also highly similar genotypes. In theory, including completely identical genotypes within a collection is costly with respect to maintaining an additional accession in a pure and safe manner without gaining additional information for breeding and research. Therefore, discussions on removing duplicates from the collection could be a long-term perspective for easier exploitation of genebanks, expressly including the development of TCCCs. However, including duplicates in TCCCs is in particular an even more substantial waste of resources. In the first years after establishment of the TCCC, costs will be high due to more thorough genotyping, such as whole-genome-sequencing, phenotyping with more replications or at more locations, and the creation of crosses with elite germplasm. Duplicates will however be in every respect less informative for pre-breeding purposes. Additionally, TCCCs are only a meaningful contribution to the exploration of genebank collections if preserved on at least a medium-term and made publicly available to breeders without considerable burdens. For both duplicates and unique accessions, these costs incur continuously year after year and are probably often underestimated in the initial funding. Moreover, duplicate genotypes could increase the MAF of rare associated variants in GWAS without improving the resolution for the specific QTL due to strong linkage within haplotypes. Finally, having approximately or exactly duplicated columns and rows in the matrix used for kinship correction in GWAS could increase collinearity issues during mixed model computation. Already at the early times of creating CCs, Brown (1989b) stated that the sampling of core collections should not tolerate redundant entries. Nowadays, the identification of such accessions is more accurate having genomic data at hand and genebank genomics should take advantage of this.

4.5 Impact and considerations of a suitable size of a TCCC

The size of a TCCC is the key variable affecting all parameters discussed above and no size was determined leading to an adequate optimization of all the parameters simultaneously. However, the benefits of sampling strategies were associated with certain ranges of TCCC sizes. As a general trend, sampling strategies which do not rely on any genotypic information do not benefit of larger sizes of TCCC. Especially, large-effect variants can be identified within small TCCCs of just 100 accession samples if the phenotypic contrast is maximized. With a TCCC size of about ≥8% of the entire collection (corresponding to ≥500 accession samples), duplicated genotypes are getting accumulated in the TCCC (Figure 4) while the statistical power is stagnating or even decreasing (Figure 6). In contrast, sampling strategies engaging an algorithm for genetic maximization allow for the incorporation of more accession samples into the TCCC. Up to 11% of all accession samples can be selected without accumulating many duplicated genotypes within the TCCC for both traits (Figure 4). These additional accessions could allow to charge the TCCC with more beneficial variants at higher frequency. While increasing the TCCC size to the aforementioned maximum is favorable with respect to the constantly increasing statistical power found for 2T_Gdiv&Gsim and

2T_Gdiv&Gdiv, smaller TCCC sizes should preferably be chosen to account for population stratification (Figure 3).

Utilizing simulated data in research allows for a targeted alteration of a single parameter. As a consequence, dependency of statistical power on the number of genotypes was found to be very clear and only increasing trends were detected by Wang and Xu (2019). In contrast, the present real-data scenarios demonstrated that additional genotypes can severely change the characteristics of a GWAS population, such as the variance component for the genotype (Supplementary Figure S7), and may even lead to reverting trends for the statistical power (Figure 6). Considering the complexity faced in genebank genomics with broad genetic diversity, these unexpected deviations from assumed trends should be considered in future.

The presented study demonstrates opportunities and obstacles in selecting a TCCC specifically for the identification of donor genotypes based on the example of a relatively large wheat germplasm collections at the IPK Genebank. In practice, much smaller or genetically narrower genebank collections may require different approaches. In general, the selection strategy will largely depend on the availability of phenotypic and genotypic data as well as the present resources to gather more data in advance for the sampling of a TCCC. As a clear guidance for genebank curators we can summarize that knowing the phenotypic distribution is most important and will indicate the possible scope of a later GWAS. TCCC should be interpreted as a top-down strategy: a large-scale phenotypic evaluation using cheaper, however mostly less accurate, methods would be a suitable start before creating a TCCC. For most genebanks, this might rely on the curation of historical records originating from past field trials or taken during seed multiplication. If a TCCC of moderate size is intended, sampling accessions with extremely contrasting phenotypes might be most promising. In most genebanks, genotypic data might not be at hand for entire collections and relying on genotypic information for the sampling strategy will increase the costs tremendously. However, TCCCs of large size will especially profit of strategies such as 2T_Gdiv&Gdiv in order to avoid genetic redundancies. Once a TCCC has been established for one trait, a shrewd strategy for the enlargement considering additional traits is the following step. Depending on the selection strategy, the ratio of accessions samples overlapping between TCCCs could help to reduce costs for the leveraging of genebank collections for many traits in future.

Data availability statement

The marker profiles of the analyzed accession samples were previously published (Schulthess et al., 2022a, b) and can be found online. Phenotypic records for YR originate from recent work by Schulthess and collaborators (Schulthess et al., 2022a, b).

Author contributions

MB: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. YJ: Methodology, Supervision, Writing – review & editing. JR: Conceptualization, Funding acquisition, Supervision, Writing – review & editing. AS: Data curation, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The present research received financial support by the German Federal Ministry of Education and Research as part of the Project GeneBank3.0 (grant no. FKZ031B1300 to AS) and by the AGENT project which is financed by the European Union's Horizon 2020 research and innovation program (grant agreement no. 862613 to MB).

Acknowledgments

The authors are truly grateful for the support provided by Moritz Lell, who gave valuable advice on solving software interference encountered during executing of software code. Authors are much obliged to Dr. Norman Philipp, who did early phenotypic data curation and supervised first field experiments. Last but not least, authors would like to also thank very much all field technicians of the 'Quantitative Genetics' group at the IPK for their valuable technical support in field activities.

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.2024.1451749/full#supplementary-material

References

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., et al. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome* 8, 1-13. doi: 10.3835/plantgenome2015.04.0024

Brown, A. H. D. (1989a). "The case for core collections," in *The Use of Plant Genetic Resources*. Eds. A. H. D. Brown, D. R. Marshall, O. H. Frankel and J. T. Williams (Cambridge University Press, Cambridge, England), 136–156.

Brown, A. H. D. (1989b). Core collections: A practical approach to genetic resources management. *Genome* 31, 818–824. doi: 10.1139/g89-144

Bundessortenamt, (2020). *Richtlinien für die Durchführung von landwirtschaftlichen Wertprüfungen und Sortenversuchen* (Hannover, Germany: Landbuch Verlagsgesellschaft mbH). Available at: https://www.bundessortenamt.de/bsa/media/Files/Richtlinie_LW2000.pdf.

Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., and Thompson, R. (2009). *ASReml User Guide Release 3.0* (Hemel Hempstead: VSN International Ltd).

De Beukelaer, H., and Davenport, G. F. (2018). *Corehunter: Multi-purpose core subset selection. R package version 3.2.1*. Available online at: https://cran.r-project.org/package=corehunter (Accessed May 13, 2024).

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Esnault, F., Pellé, R., Dantec, J.-P., Bérard, A., Le Paslier, M.-C., and Chauvin, J.-E. (2016). Development of a potato cultivar (Solanum tuberosum L.) core collection, a valuable tool to prospect genetic variation for novel traits. *Potato Res.* 59, 329–343. doi: 10.1007/s11540-016-9332-x

FAO (2010). "Chapter 3 The state of *ex situ* conservation," in *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (FAO, Rome), 53–90.

Frankel, O. H. (1984). "Genetic perspectives of germplasm conservation," in *Genetic manipulation: impact on man and society*. Eds. W. Arber, K. Illemensee, W. J. Peacock and P. Starlinger (Cambridge University Press, Cambridge, England), 161–170.

Gao, X., Starmer, J., and Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32, 361–369. doi: 10.1002/gepi.20310

Goudet, J., and Jombart, T. (2022). *Hierfstat: Estimation and tests of hierarchical F-statistics. R package version 0.5-11*. Available online at: https://CRAN.R-project.org/package=hierfstat (Accessed May 13, 2024).

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53, 325–338. doi: 10.2307/2333639

Gu, R., Fan, S., Wei, S., Li, J., Zheng, S., and Liu, G. (2023). Developments on core collections of plant genetic resources: Do we know enough? *Forests* 14, 926. doi: 10.3390/f14050926

Guey, L. T., Kravic, J., Melander, O., Burtt, N. P., Laramie, J. M., Lyssenko, V., et al. (2011). Power in the phenotypic extremes: A simulation study of power in discovery and replication of rare variants. *Genet. Epidemiol.* 35, 236–246. doi: 10.1002/gepi.20572

Guo, Y., Kuang, L., Xu, Y., Yan, T., Jiang, L., Dong, J., et al. (2022). Construction of a worldwide core collection of rapeseed and association analysis for waterlogging tolerance. *Plant Growth Regul.* 98, 321–328. doi: 10.1007/s10725-022-00862-5

Haupt, M., and Schmid, K. (2020). Combining focused identification of germplasm and core collection strategies to identify genebank accessions for central European soybean breeding. *Plant Cell Environ.* 43, 1421–1436. doi: 10.1111/pce.13761

IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, 1–13. doi: 10.1126/science.aar7191

Kosmicki, J. A., Churchhouse, C. L., Rivas, M. A., and Neale, B. M. (2016). Discovery of rare variants for complex phenotypes. *Hum. Genet.* 135, 625–634. doi: 10.1007/s00439-016-1679-1

Lee, H.-Y., Ro, N.-Y., Jeong, H.-J., Kwon, J.-K., Jo, J., Ha, Y., et al. (2016). Genetic diversity and population structure analysis to construct a core collection from a large Capsicum germplasm. *BMC Genet.* 17, 142. doi: 10.1186/s12863-016-0452-8

Marita, J. M., Rodriguez, J. M., and Nienhuis, J. (2000). Development of an algorithm identifying maximally diverse core collections. *Genet. Resour. Crop Evol.* 47, 515–526. doi: 10.1023/A:1008784610962

Milner, S. G., Jost, M., Taketa, S., Mazón, E. R., Himmelbach, A., Oppermann, M., et al. (2019). Genebank genomics highlights the diversity of a global barley collection. *Nat. Genet.* 51, 319–326. doi: 10.1038/s41588-018-0266-x

Muñoz-Amatriaín, M., Cuesta-Marcos, A., Endelman, J. B., Comadran, J., Bonman, J. M., Bockelman, H. E., et al. (2014). The USDA barley core collection: Genetic diversity, population structure, and potential for genome-wide association studies. *PloS One* 9, 1–13. doi: 10.1371/journal.pone.0094688

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., et al. (2009). Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell* 21, 2194–2202. doi: 10.1105/tpc.109.068437

Panarella, M., and Burkett, K. M. (2019). A cautionary note on the effects of population stratification under an extreme phenotype sampling design. *Front. Genet.* 10. doi: 10.3389/fgene.2019.00398

Pascual, L., Fernández, M., Aparicio, N., López-Fernández, M., Fité, R., Giraldo, P., et al. (2020). Development of a multipurpose core collection of bread wheat based on high-throughput genotyping data. *Agronomy* 10, 534. doi: 10.3390/agronomy10040534

Phogat, B. S., Kumar, S., Kumari, J., Kumar, N., Pandey, A. C., Singh, T. P., et al. (2021). Characterization of wheat germplasm conserved in the Indian National Genebank and establishment of a composite core collection. *Crop Sci.* 61, 604–620. doi: 10.1002/csc2.20285

Ramirez-Villegas, J., Khoury, C. K., Achicanoy, H. A., Diaz, M. V., Mendez, A. C., Sosa, C. C., et al. (2022). State of *ex situ* conservation of landrace groups of 25 major crops. *Nat. Plants* 8, 491–499. doi: 10.1038/s41477-022-01144-8

R Core Team (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing* (Vienna, Austria: R Foundation for statistical computing). Available at: https://www.R-project.org/.

Rogers, J. S. (1972). "Measures of genetic similarity and genetic distance," in *Studies in Genetics VII*. Ed. M. Wheleer (University of Texas, Austin, TX), 145–153.

Schulthess, A. W., Kale, S. M., Liu, F., Zhao, Y., Philipp, N., Rembe, M., et al. (2022a). Genomics-informed prebreeding unlocks the diversity in genebanks for wheat improvement. *Nat. Genet.* 54, 1544–1552. doi: 10.1038/s41588-022-01189-7

Schulthess, A. W., Kale, S. M., Zhao, Y., Gogna, A., Rembe, M., Philipp, N., et al. (2022b). Large-scale genotyping and phenotyping of a worldwide winter wheat genebank for its use in pre-breeding. *Sci. Data* 9, 784. doi: 10.1038/s41597-022-01891-5

Singh, N., Wu, S., Raupp, W. J., Sehgal, S., Arora, S., Tiwari, V., et al. (2019). Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Sci. Rep.* 9, 650. doi: 10.1038/s41598-018-37269-0

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

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Software* 3, 731. doi: 10.21105/joss.00731

Upadhyaya, H. D., Ortiz, R., Bramel, P. J., and Singh, S. (2003). Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. *Genet. Resour. Crop Evol.* 50, 139–148. doi: 10.1023/A:1022945715628

Urbanek, S. (2021). *rJava: Low-level R to Java interface. R package version 1.0-6*. Available online at: https://CRAN.R-project.org/package=rJava (Accessed May 13, 2024).

Van Gestel, S., Houwing-Duistermaat, J. J., Adolfsson, R., Van Duijn, C. M., and Van Broeckhoven, C. (2000). Power of selective genotyping in genetic association analyses of quantitative traits. *Behav. Genet.* 30, 141–146. doi: 10.1023/A:1001907321955

Van Hintum, T. J., Brown, A. H. D., Spillane, C., and Hodgkin, T. (2000). *Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3* (Rome, Italy: International Plant Genetic Resources Institute).

Wang, M., and Xu, S. (2019). Statistical power in genome-wide association studies and quantitative trait locus mapping. *Heredity* 123, 287–306. doi: 10.1038/s41437-019-0205-3

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.* 7, 167–184. doi: 10.1007/s10592-005-9100-y

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (New York, NY: Springer-Verlag).

Xing, C., and Xing, G. (2009). Power of selective genotyping in genome-wide association studies of quantitative traits. *BMC Proc.* 3, 23. doi: 10.1186/1753-6561-3-s7-s23

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 December 2024
DOI 10.3389/fpls.2024.1501533

OPEN ACCESS

EDITED BY
Ming Luo,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

REVIEWED BY
Keisuke Nagai,
Nagoya University, Japan
Changhui Sun,
Sichuan Agricultural University, China

*CORRESPONDENCE
Siwaret Arikit
siwaret.a@ku.th
Samart Wanchana
samart.wan@biotec.or.th

RECEIVED 25 September 2024
ACCEPTED 27 November 2024
PUBLISHED 18 December 2024

CITATION
Klinsawang S, Aesomnuk W, Mangkalasane P, Ruanjaichon V, Siangliw JL, Pandey BK, Bennett MJ, Wanchana S and Arikit S (2024) Genome-wide association study identifies key F-box genes linked to ethylene responsiveness and root growth in rice (*Oryza sativa* L.). *Front. Plant Sci.* 15:1501533. doi: 10.3389/fpls.2024.1501533

Genome-wide association study identifies key F-box genes linked to ethylene responsiveness and root growth in rice (*Oryza sativa* L.)

Suparad Klinsawang[1], Wanchana Aesomnuk[2], Piyamongkol Mangkalasane[2], Vinitchan Ruanjaichon[3], Jonaliza L. Siangliw[3], Bipin K. Pandey[4], Malcolm J. Bennett[4], Samart Wanchana[3]* and Siwaret Arikit[1,2]*

[1]Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Nakhon Pathom, Thailand, [2]Rice Science Center, Kasetsart University, Nakhon Pathom, Thailand, [3]National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand, [4]School of Biosciences, University of Nottingham, Leicestershire, United Kingdom

Rice (*Oryza sativa* L.) is a staple food for more than half of the world's population, but its yields are increasingly threatened by environmental problems, including soil compaction. This problem limits root growth which limits water and nutrient foraging capacity thus reduces productivity due to, restricted diffusion of ethylene, a key plant hormone playing an important role in exacerbating these effects. Elevated ethylene levels in compacted soils can further inhibit root development. However, rice varieties that are less sensitive to ethylene may have an advantage as they exhibit better root growth and resource utilization under such conditions. In this study, 220 diverse rice accessions were analyzed to uncover the genetic factors that influence root length reduction (RLR) in response to ethylene. Genome-wide association studies (GWAS) identified a significant QTL on chromosome 10, named *qRLR10*, associated with ethylene response. Within this region, 20 candidate genes were identified, with three F-box genes namely *Os10g0124700*, *Os10g0126600* and *Os10g0128200* showing a strong correlation with RLR variations. These genes are involved in protein degradation, root development and hormone signaling, indicating their possible role in regulating ethylene sensitivity. The results suggest that rice varieties with lower ethylene sensitivity may have better root growth in compacted soils, making them ideal targets for breeding programs aimed at improving resilience to harsh environmental conditions. These results underscore the critical role of ethylene in rice root development and provide valuable insights for future rice improvement strategies aimed at mitigating the effects of soil compaction.

KEYWORDS

rice, root, ethylene sensitivity, GWAS, F-box

1 Introduction

Rice (*Oryza sativa* L.) is one of the most important staple crops that feeds more than half of the world's population. Thailand, a major player in global rice production, ranks as the sixth-largest rice producer, following China, India, Indonesia, Bangladesh, and Vietnam (FAO, 2017). Farming practices in Thailand and other countries have significantly evolved over the years, with mechanization becoming increasingly central. The use of heavy machinery, such as power tillers, tractors, and harvesters, has become widespread in agricultural operations, including soil preparation, sowing, and harvesting (FAO, 2017). While these advancements have boosted agricultural productivity, they have also led to unintended consequences, particularly soil compaction, which poses a significant challenge to crop growth and yield (Hassan et al., 2007). Soil compaction, a condition where soil particles are pressed together, reducing macropore spaces, is a major constraint in modern agriculture. It negatively impacts crop production by restricting root growth, limiting soil water and nutrient availability, and decreasing soil porosity (Arocena, 2000; Khan et al., 2012). Additionally, soil compaction hinders gas exchange between plant roots and the surrounding environment, further exacerbating plant stress. In this context, plants naturally produce various volatile phytohormones, such as ethylene, which can diffuse through air spaces between soil aggregates in noncompacted soil. However, soil compaction restricts ethylene diffusion in soil micropores, leading to inhibited root elongation and reduced overall root growth, which negatively impacts the plant's ability to access water and nutrients (Pandey et al., 2021; Arocena, 2000). Elevated ethylene levels in compacted soils can trigger stress responses that further restrict root development, thereby compromising plant growth and yield (Lin et al., 2010).

Ethylene, a gaseous phytohormone synthesized from methionine in the presence of oxygen, is integral to plant stress responses. It is produced in most plant tissues and serves as a key regulator of various physiological processes, particularly under stress conditions (Lin et al., 2010). Ethylene Response Factors (ERFs) are central to ethylene's role in stress response, acting as regulatory hubs that interact with other stress-related signals, such as abscisic acid and (ABA) and jasmonate (Müller and Munné-Bosch, 2015). These interactions are critical in modulating plant responses to diverse abiotic stresses, including salinity, drought, low temperature, heat, and variations in light availability. The ethylene signaling pathway is a complex network involving several key components, including receptors, *EIN2*, and transcription factors like *EIN3* (*Ethylene insensitive 3*)/*EIL1 (Ethylene insensitive-Like1)*. These components activate ERFs, which, in turn, regulate the expression of genes involved in stress and downstream hormone signaling pathway. Notably, rice mutant lines with alterations in EIN3/EIL1 maintain normal primary root growth even under compacted soil conditions, similar to wild-type plants in non-compacted soils (Pandey et al., 2021). This resilience is likely due to the disrupted ethylene signaling in these mutants, which prevents the typical ethylene-induced inhibition of root elongation in response to soil compaction.

The regulation of root growth and development in rice is a complex process influenced by multiple phytohormones, including auxin, ABA, cytokinin, and ethylene. These hormones are involved in various aspects of root growth and development, from biosynthesis to signal transduction, and play a crucial role in how plants respond and acclimate to environmental stresses (Dello Ioio et al., 2007; De Smet and Jürgens, 2007; McSteen, 2010). Ethylene, in particular, interacts with auxin by modulating its biosynthesis and transport, highlighting its role as a mediator in the regulation of root development (Le et al., 2001). Given the critical role of ethylene in plant stress response, manipulating its pathway offers a promising strategy for improving crop production, especially under adverse conditions. For instance, in maize, mutations in the ACC synthase gene, a key enzyme in ethylene biosynthesis, have been shown to delay leaf senescence and maintain photosynthesis during drought, leading to increased grain yield under stress (Habben et al., 2014). However, breeding for ethylene-responsive traits in crops remains challenging, particularly in non-GMO contexts. Nevertheless, utilizing diverse plant collections and innovative screening protocols can provide valuable genetic resources for breeding programs aimed at enhancing tolerance to unfavorable environments. In rice, developing ethylene-insensitive varieties could be a potential strategy to improve growth and production, particularly in compacted soils where water and nutrient extraction are critical for survival. This approach is especially relevant for countries like Thailand, where traditional breeding methods are preferred over genetic modification. By leveraging the genetic diversity found in local or landrace rice accessions, which often contain adaptive traits to harsh environments, breeders can identify and incorporate valuable alleles into commercial cultivars.

In this study, we aimed to investigate the root length response of 220 rice accessions under controlled and ethylene-treated conditions. By employing a genome-wide association study (GWAS), we sought to identify the genomic regions and key genes regulating root length and ethylene sensitivity in rice. The findings from this study are expected to contribute to the development of new rice varieties with enhanced root penetration ability and improved adaptation to soil compaction and other abiotic stresses.

2 Materials and methods

2.1 Plant materials and phenotypic screening experiment

A rice diversity panel consisting of 220 rice accessions, including Thai landraces, Thai improved rice varieties, and rice varieties from other countries, were used to represent the diverse geographical regions and rice ecosystems such as rainfed lowland, lowland, upland, deepwater and irrigated system which are found in Thailand (Supplementary Table S1). All rice accessions were evaluated ethylene sensitivity using an ethylene chamber to observe root characteristics in response to ethylene. A total of 10-

20 seeds per accession were used to screen the ethylene sensitivity. Seeds were first sterilized with 10% sodium hypochlorite and rinsed with distilled water. The seeds were pre-germinated by soaking with distilled water at 30°C in the dark for 2 days then placed on stainless net, each accession was separated with a well. The samples were then put in an air-tight plastic box (27.5 x 39.5 x 25 cm) with a fitted net. The water (200 ml) was added into the box to provide enough moisture and humidity to allow the seedlings to grow in a rice growth room (25°C). The experiments were set into two conditions, ethylene treatment and control. The experimental design was a complete randomized design (CRD) with three replications. For ethylene treatment, the ethylene gas was injected into the plastic box by a calibrated syringe to get 20 ppm concentration. The root length phenotype was observed at 2 days after ethylene treatment following the modified protocols (Ma et al., 2013 and Pandey et al., 2021). For ethephon treatment, the rice seeds were sterilized with 10% sodium hypochlorite for 10 minutes, then rinsed with deionized (DI) water until clean. Next, they were transferred to plates with 20 mL of DI water and incubated in the dark at 30°C for 2 days. After germination, seedlings were placed on stainless steel mesh channels in two chambers: one with 3 L of DI water (control) and one with 100 μM ethephon in 3 L of DI water (treatment). Seedlings were treated for 48 hours, and root lengths were measured using ImageJ software to calculate root length reduction. The percentage of root length reduction is calculated by subtracting the root length in ethylene from the root length in control and multiplied by 100, then divided by root length in control. R version 4.4.1 (R Core Team, 2024) was used to perform data distribution and statistical analysis with R packages in Rstudio (Wickham, 2016; RStudio Team, 2020).

2.2 Population study and genome wide association study analysis

A total of 2,340,775 SNPs with a minor allele frequency (MAF) greater than 5%, missing data less than 10%, and pruned with a 1-kb window were used for population studies and GWAS (Supplementary Figure S1). These SNPs were derived from a whole-genome resequencing project conducted by the National Center for Genetic Engineering and Biotechnology, Thailand (Theerawitaya et al., 2022) using Nipponbare IRGSP 1.0 as the reference rice genome. A SNP density heatmap was generated using the CMplot package in RStudio (Yang et al., 2022). Population structure and individual relatedness were determined through principal component analysis (PCA) and kinship analysis using the GAPIT software (Wang and Zhang, 2021). GWAS analysis was performed using the FarmCPU model in GAPIT, with significant SNP associations defined by a uniform threshold of $-\log_{10}(\text{p-value}) > 6$.

2.3 Linkage disequilibrium decay, haplotype analysis and candidate gene identification

Genome-wide LD decay was estimated by PopLDdecay (Zhang et al., 2019) and the LD structure and haplotypes within the QTL region were inferred using LDBlockshow (Zhang et al., 2019). The identification of genes located within the QTL region was based on the annotations from RAP-DB (https://rapdb.dna.affrc.go.jp) (Kawahara et al., 2013; Sakai et al., 2013). The effects of SNP variation in the genes were identified using the Variant Effect Predictor (VEP - https://asia.ensembl.org/info/docs/tools/vep/index.html). The gene-based haplotyped were analyzed using GeneHapR (Zhang et al., 2023) based on the SNPs with a moderate or high effect present in the genes. Candidate genes were identified by conducting an analysis of variance (ANOVA) to test for significant differences between the major haplotypes (each consisting of at least 7 samples) within the QTL.

3 Results

3.1 Phenotypic variation in root length reduction percentage under ethylene gas exposure in a panel of 220 rice accessions

Root responsiveness to ethylene was assessed by calculating the percentage of reduction in root length (%RLR) for each rice accession by comparing the root lengths under ethylene treatment (20 ppm) with control conditions (Figure 1A). KDML105 served as a sensitive control to validate ethylene response and ensure reproducibility (Figure 1B). The %RLR in this diversity panel ranged from 52.06% to 82.98%, with an average of 67.76% (Figure 1A). Rice accessions with a %RLR of 60.67% or lower were classified as low ethylene sensitivity, while those with a %RLR of 74.9% or higher were considered highly sensitive. Naew Ubon 2 exhibited the lowest %RLR (52.06%), and Phitsanulok 80 had the highest %RLR (82.98%). The %RLR distribution across the panel displayed a slightly skew towards low values, forming a continuous curve (Figure 1C). The RLR phenotype in rice accessions with low and high sensitivity was confirmed using an ethephon solution treatment (Supplementary Figure S2).

3.2 Population study and linkage disequilibrium analysis

The population structure and cryptic relationship among the 220 rice accessions were analyzed using Principal Component Analysis (PCA) and a kinship matrix (Figure 2). All rice accessions in this study belong to the indica ecotype (Supplementary Table S1). The PCA and VanRaden kinship matrix results indicated the absence of distinct subpopulations within the diversity panel. The kinship coefficient is in the range of 0 to 0.5 which is often used for relatedness in GWAS. To estimate genome-wide linkage disequilibrium (LD) decay, we calculated the pairwise LD index (r^2) and plotted the LD decay across the rice diversity panel. The average LD decay for the entire genome of the 220 rice accessions was approximately 190 kb ($r^2 < 0.2$). Chromosome-wise, the shortest LD decay distance was around 100 kb on chromosome 11, while the longest was approximately 280 kb on chromosome 12 (Figure 3).

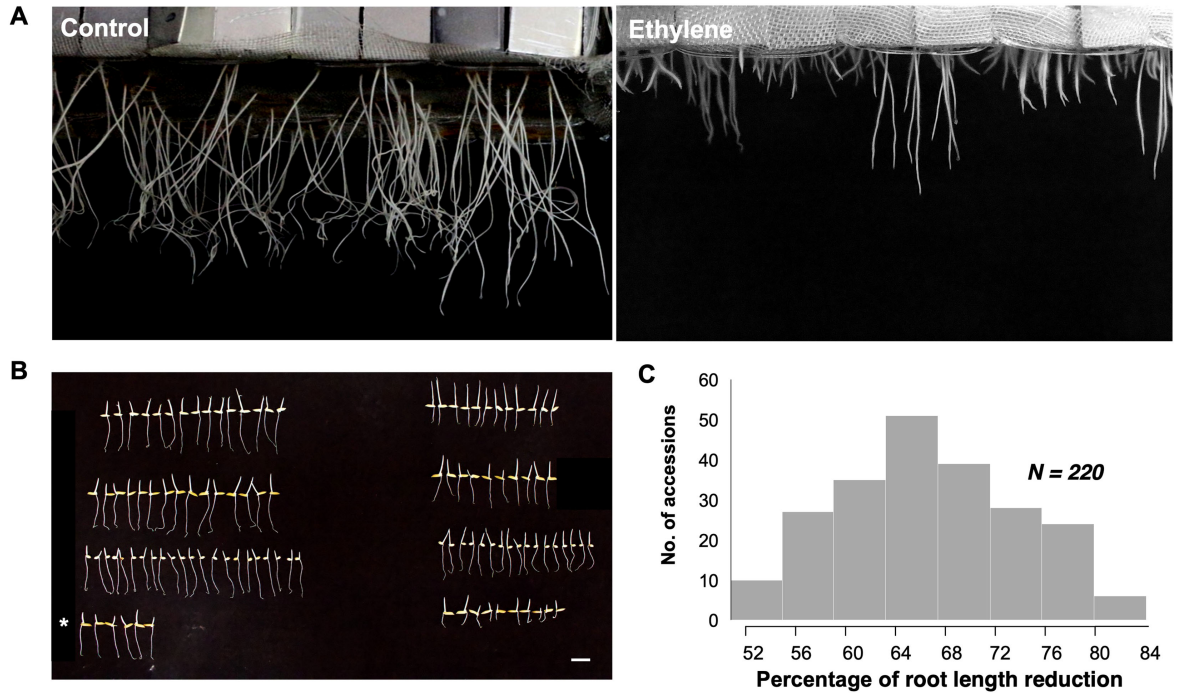


FIGURE 1
(A) Root length reduction under control and ethylene treatments, showing contrasting phenotypes of root length among different accessions. **(B)** Comparison of root length between the two treatments for selected accessions, with KDML105 as the reference check (denoted by *). Left: control; right: ethylene treatment. Scale bar: 2 cm. **(C)** Distribution of data showing the percentage of root length reduction across accessions.

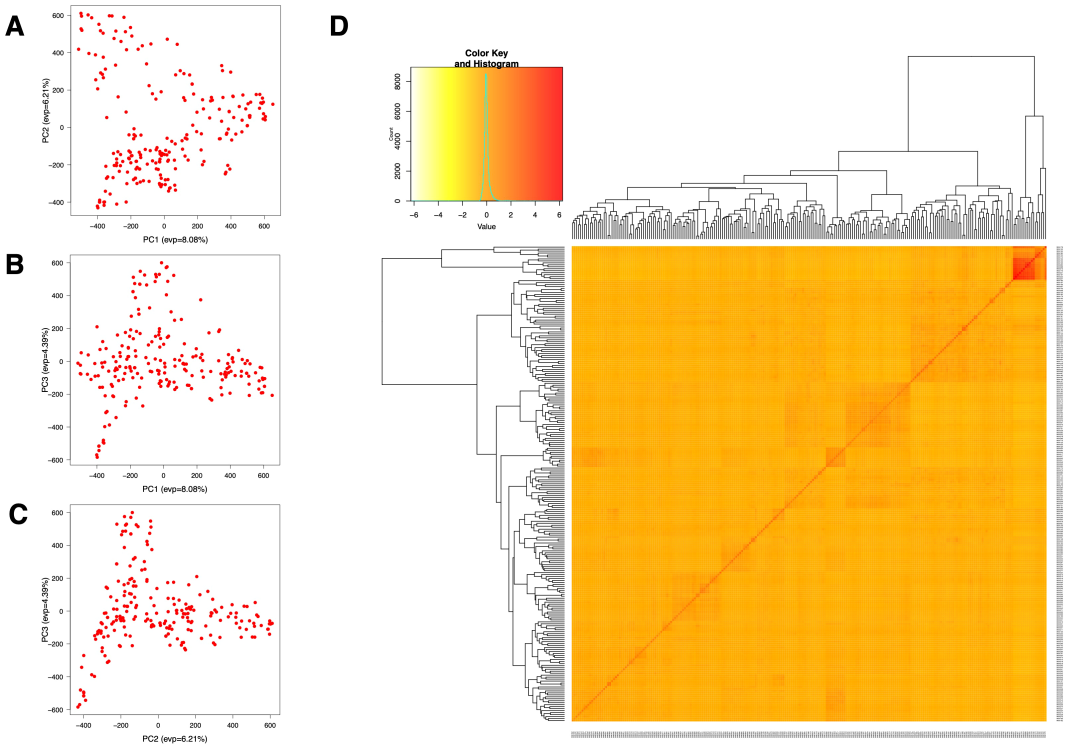


FIGURE 2
Population structure for GWAS analysis of 220 rice accessions. **(A–C)** Variation captured by the first three principal components (PCs): **(A)** PC1 vs. PC2, **(B)** PC1 vs. PC3, and **(C)** PC2 vs. PC3. **(D)** Kinship matrix of the 220 individuals, presented as a heatmap.

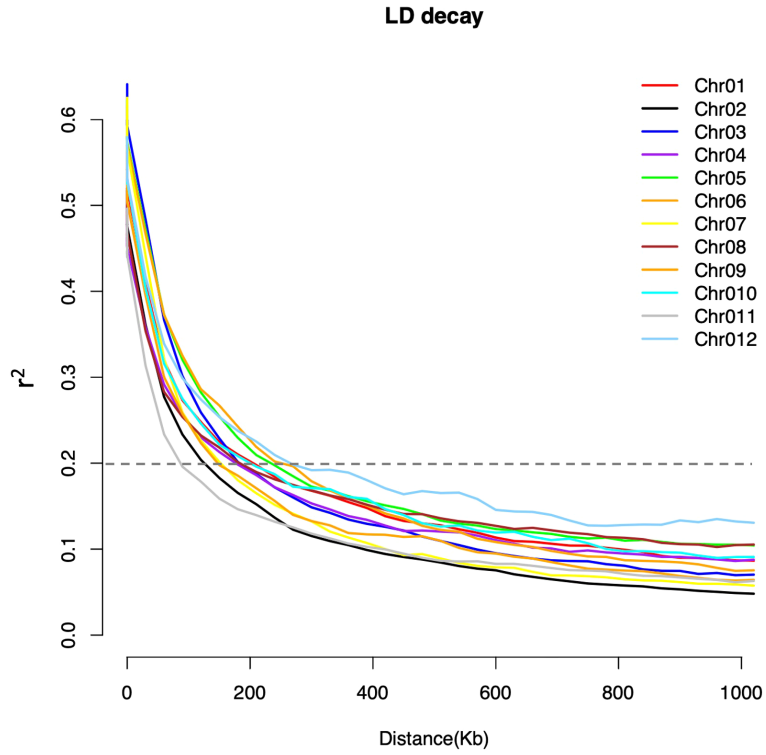


FIGURE 3
Chromosome-wide linkage disequilibrium (LD) decay estimated from SNP genotypes of 220 rice accessions. Each line represents the smoothed r^2 values for all marker pairs on each chromosome as a function of the distance between marker pairs. The LD decay threshold (r^2 = 0.2) is indicated by the dotted line.

3.3 Genome-wide association analysis and candidate gene identification

To identify genomic regions associated with root responsiveness to ethylene, we conducted a genome-wide association study (GWAS) on 220 rice accessions, utilizing 2,340,775 SNP as genotype data. All SNPs had a minor allele frequency (MAF) greater than 0.05. We employed the FarmCPU model, incorporating the first three principal components (PCs) and the kinship matrix, to detect quantitative trait loci (QTLs). As a result, a quantitative trait nucleotide (QTN) was detected with the highest significance (-log10 p-value of 6.86) on chromosome 10 at position 1,562,158 bp (Figure 4). Given the average linkage disequilibrium (LD) decay on chromosome 10, approximately 200 kb (r^2 = 0.2), we defined the region between 1.36 and 1.76 Mb as the QTL region (named *qRLR10*) associated with root responsiveness to ethylene (Table 1).

We then identified candidate genes within the *qRLR10* region by analyzing genes annotated within the LD block containing the lead

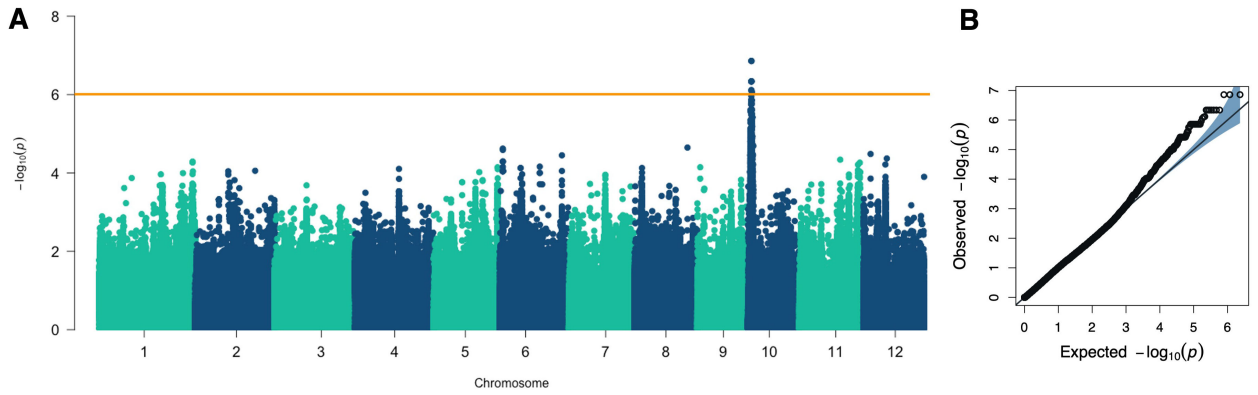


FIGURE 4
GWAS results for the percentage of root length reduction. **(A)** Manhattan plot from the FarmCPU model displaying the association signals across the genome. **(B)** Quantile-Quantile (Q-Q) plot comparing the distribution of observed versus expected p-values from the GWAS analysis.

TABLE 1 The QTLs for percentage of root length reduction based on significant SNPs.

QTL	Chromosome	QTL region (Mb)	Lead SNP	P-value	MAF
qRLR10	10	1.36 – 1.76	1562158	1.39 × 10-07	0.44

SNP. A total of 20 genes with annotated functions were identified (Table 2). Among these, several genes are related to root development and plant hormone responses, such as *Os10g0122600* (nodulin domain-containing protein, *OsSNDP1*), which is associates with root hair elongation (Huang et al., 2013). In addition, 12 genes encode F-box domain-containing proteins, including *Os10g0123200, Os10g0124500 (OsFbox507, OsFBX343), Os10g0124700 (OsFbox508, OsFBX344), Os10g0125300, Os10g0126000 (OsFbox510, OsFBX346), Os10g0126500 (OsFbox511, OsFBX347), Os10g0126600 (OsFbox512, OsFBX348), Os10g0126700 (OsFbox513, OsFBX346), Os10g0126800 (OsFbox514, OsFBX350), Os10g0127000 (OsFbox515, OsFBX351), Os10g0127900 (OsFbox516, OsFBX352)*, and *Os10g0128200* (*OsFbox517, OsFBX353*). Furthermore, genes encoding CCR4-associated factor 1 (*CAF1*) proteins, *Os10g0123900* (*OsCAF1-14*) and *Os10g0124200* (*OsCAF1-15*), which are important enzymes for catalysis of mRNA de-adenylation in eukaryotes, were also identified. CAF1 proteins have been reported to response to abiotic stresses in plants. A search of Rice Expression Profile Database (http://ricexpro.dna.affrc.go.jp/) showed that 16 gene expression profiles are in database, and they have different expression patterns, referring their different biological functions. Particularly, *Os10g0122000*, *Os10g0122600*, *Os10g0124700*, *Os10g0125700*, and *Os10g0126000* are expressed in various rice organs throughout rice development, including high expression in root. According to root gene expression profile, *Os10g0122000* is highly expressed in maturation zone, *Os10g0124700* is highly expressed in division zone, *Os10g0122600*, *Os10g0125700*, and *Os10g0126000* are expressed in elongation zone, and maturation I.

3.4 Haplotype analysis

To prioritize potential candidate genes, we performed a gene-based haplotype analysis on each of the 20 genes to identify those with haplotypes associated with variations in root length reduction (RLR) within the diversity panel. The analysis focused on SNPs with

TABLE 2 The 20 associated genes within the QTL region containing nonsynonymous SNPs causing missense and nonsense effect.

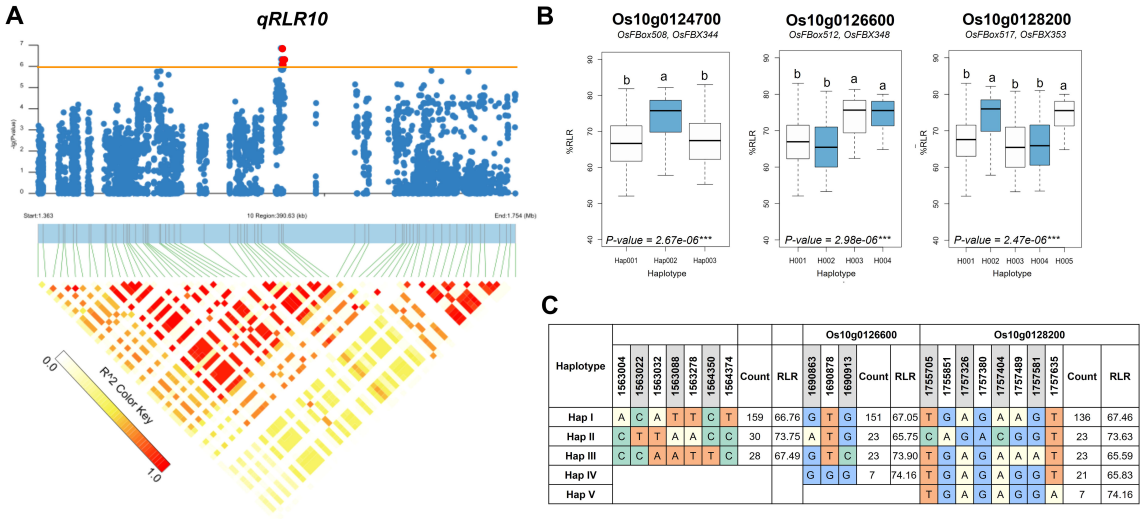
QTL	Chr.	Start	Stop	Locus ID	Description
qRLR10	10	1405692	1407206	*Os10g0122000*	UDP-glucuronosyl/UDP-glucosyltransferase family protein. (Os10t0122000-00)
		1431142	1432407	*Os10g0122300*	Transferase domain containing protein. (Os10t0122300-00)
		1447351	1451372	*Os10g0122600*	Sec14 and nodulin domain-containing protein, Root hair elongation (Os10t0122600-01)
		1479530	1481991	*Os10g0123200*	Cyclin-like F-box domain containing protein. (Os10t0123200-01)
		1525512	1526367	*Os10g0123900*	Similar to CAF1 family ribonuclease containing protein. (Os10t0123900-00)
		1530563	1532176	*Os10g0124000*	Ribosomal protein L7A/RS6 family domain containing protein. (Os10t0124000-01)
		1539462	1540349	*Os10g0124200*	Similar to CAF1 family ribonuclease containing protein. (Os10t0124200-00)
		1556778	1559803	*Os10g0124500*	Cyclin-like F-box domain containing protein. (Os10t0124500-01)
		1562608	1565335	*Os10g0124700*	Cyclin-like F-box domain containing protein. (Os10t0124700-01)
		1622222	1635958	*Os10g0125300*	F-box domain, cyclin-like domain containing protein. (Os10t0125300-01);F-box protein, Tapetum cell development, Pollen formation, Control of anther development (Os10t0125300-02)
		1642565	1646770	*Os10g0125700*	NB-ARC domain containing protein. (Os10t0125700-00)
		1656688	1659249	*Os10g0126000*	Similar to F-box domain containing protein, expressed. (Os10t0126000-01)
		1681092	1684225	*Os10g0126500*	Cyclin-like F-box domain containing protein. (Os10t0126500-01)
		1688331	1691224	*Os10g0126600*	Cyclin-like F-box domain containing protein. (Os10t0126600-01)
		1694920	1697854	*Os10g0126700*	Similar to F-box domain containing protein, expressed. (Os10t0126700-01)
		1700332	1702953	*Os10g0126800*	Cyclin-like F-box domain containing protein. (Os10t0126800-01)
		1711433	1713322	*Os10g0127000*	Similar to F-box domain containing protein, expressed. (Os10t0127000-01)
		1726495	1727783	*Os10g0127350*	Glycosyl transferase, family 14 domain containing protein. (Os10t0127350-00)
		1743579	1752159	*Os10g0127900*	Similar to F-box domain containing protein, expressed. (Os10t0127900-01; Regulation of glucose-induced delay of seed germination (Os10t0127900-03)
		1755462	1757886	*Os10g0128200*	Similar to F-box domain containing protein, expressed. (Os10t0128200-01)

moderate to high effects, such as nonsynonymous and nonsense SNPs, within each gene. Among the 20 genes analyzed, three—*Os10g0124700*, *Os10g0126600*, and *Os10g0128200*—were found to contain haplotypes significantly associated with the phenotype differences. For *Os10g0124700*, three distinct major haplotypes were identified. Haplotype H002 (n = 30) was significantly associated with a higher %RLR (73.75), whereas haplotypes H001 (n = 159) and H003 (n = 28) were associated with lower %RLR values (66.76 and 67.49, respectively: Figure 5). In the case of Os10g0126600, four major haplotypes were found. Haplotypes H001 (n = 151) and H002 (n = 23) were associated with lower % RLR values (67.05 and 65.75, respectively), while the haplotypes H003 (n = 23) and H004 (n = 7) were associated with higher %RLR values (73.90 and 74.16, respectively). For *Os10g0128200*, five major haplotypes were identified. Among these, the haplotypes H002 (n = 23) and H005 (n = 7) were associated with higher %RLR values (73.63 and 74.16, respectively), whereas haplotypes H001 (n = 136), H003 (n = 23) and H004 (n = 21) were associated with lower %RLR values (67.46, 65.59 and 65.83, respectively).

4 Discussion

Root plays a vital role in rice growth and performance under both normal and stress conditions, serving as the primary interface between the plant and its environment. The growth, development, and adaptive responses of rice are regulated by a complex network of internal and external factors, including phytohormones such as ethylene, auxin and cytokinin. The hormones can act synergistically or antagonistically to modulate plant growth, particularly under stress conditions (He et al., 2005; Ueda et al., 2020; Yamauchi et al., 2020). Ethylene, often referred to as the stress hormones, is a key mediator in plant adaptive responses, interacting with other hormones to regulate growth under adverse conditions (Huang et al., 2020; Ma et al., 2013; Yang et al., 2015). Roots are particularly sensitive to changes in ethylene concentration, which can lead to reduced root length and increased radial cell expansion (Vaseva et al., 2018). Additionally, ethylene has been shown to influence root hair growth and development (Feng et al., 2017) and plays a role in the plant's response to soil compaction by reducing root length and increasing root thickness in rice, a process mediated by abscisic acid (ABA) and auxin (Huang et al., 2022). Transgenic rice plants overproducing ethylene exhibit lower 1000-grain weight and faster leaf senescence compared to those with reduced ethylene production, indicating the significant impact of ethylene on overall plant development and yield (Ma et al., 2013). In the ethylene signaling pathway, the ethylene gas triggers the translocation of *CONSTITUTIVE TRIPLE RESPONSE1 (CTR1)* from the endoplasmic reticulum to the nucleus. This process involves the phosphorylation of *ETHYLENE-INSENSITIVE2 (EIN2)* by CTR1, which inhibits EIN2 from signaling in the absence of ethylene (Ju et al., 2012; Jun et al., 2004). In Arabidopsis, plants with enhanced nuclear-localized CTR1 have shown to improve tolerance to drought and salinity stress (Park et al., 2023). These findings suggest that manipulating ethylene sensitivity could be a valuable strategy for improving crop resilience under unfavorable environmental conditions, such as those exacerbated by climate change, while maintaining productivity.

In rice, ethylene's role in root response to soil compaction has been further elucidated through studies involving rice mutants with varying levels of ethylene sensitivity. For instance, ethylene-insensitive mutants exhibited significantly longer root than those with high ethylene sensitivity under soil compaction, highlighting the genetic basis of ethylene responsiveness (Pandey et al., 2021). In this study, 220 rice accessions were collected from many areas in Thailand in order to have a diverse genetic source to study GWAS. According



Haplotype	1563004	1563022	1563032	1563088	1563278	1564350	1564374	Count	RLR	1690863	1690878	1690913	Count	RLR	1755705	1755851	1757326	1757380	1757404	1757489	1757581	1757635	Count	RLR
										Os10g0126600					Os10g0128200									
Hap I	A	C	A	T	T	C	T	159	66.76	G	T	G	151	67.05	T	G	A	G	A	A	G	T	136	67.46
Hap II	C	T	T	A	A	C	C	30	73.75	A	T	G	23	65.75	C	A	G	A	C	G	G	T	23	73.63
Hap III	C	C	A	A	T	T	C	28	67.49	G	T	C	23	73.90	T	G	A	G	A	A	A	T	23	65.59
Hap IV										G	G	G	7	74.16	T	G	A	G	A	G	G	T	21	65.83
Hap V															T	G	A	G	A	G	G	A	7	74.16

FIGURE 5
Associated regions and haplotype analysis of *qRLR10*. **(A)** Regional Manhattan plots and LD heatmap across the 390.6-kb region surrounding the significant SNP. Red dots above the red line indicate significant SNPs. Triangles in the LD heatmap represent haploblocks with varying R^2 values. **(B)** Boxplots displaying the distribution of root length reduction (RLR) across haplotype groups of *Os10g0124700, Os10g0126600,* and *Os10g0128200*. The middle line in each boxplot represents the median. Different letters above each haplotype group show significant difference at $p \leq 0.05$. **(C)** Haplotype analysis of the three genes in the region.

to Kinship and PCA analysis, our germplasm showed the low levels of relatedness within the rice indica panel (Aesomnuk et al., 2021). Thus, this diversity panel can be used in GWAS analysis. Importantly, the higher genetic diversity of germplasm, the more genetic variation in target trait, leading to successfully identify the crucial gene underlying the target trait (Quero et al., 2018). This will help plant breeder to enhance breeding programs to genetically improve target traits (Mourad et al., 2020). Our genome-wide association study (GWAS) using 220 indica rice accessions identified a QTL, *qRLR10*, associated with difference in root length reduction (RLR) in response to ethylene. Within this QTL region, 20 candidate genes were identified. The subsequent gene-based haplotype analysis revealed that specific haplotypes within three of these genes—*Os10g0124700*, *Os10g0126600*, and *Os10g0128200*—are significantly associated with variations in RLR. This association underscores the potential functional importance of these genes in the regulation of ethylene responsiveness.

The gene *Os10g0124700* (*OsFbox508, OsFBX344*), which encodes an F-box protein, showed the highest expression in rice roots before the flowering stage according to the strand-specific RNA-Seq database in RAP-DB (Wang et al., 2015). F-box proteins are components of the SCF (SKP1-CUL1-F-box protein) complex, which functions as an E3 ubiquitin ligase, tagging target proteins for degradation. This pathway is crucial for regulating various physiological processes, including growth, development, and stress responses (Gagne et al., 2002; Lechner et al., 2006; Dreher and Callis, 2007; Zhang et al., 2008; Guo et al., 2018; Xu et al., 2021). F-box proteins are likely involved in the fine-tuning of ethylene responses by regulation the stability of key component signaling pathway. For example, in Arabidopsis, the F-box protein EBF1/EBF2 targets the ethylene-responsive transcription factor EIN3 for degradation, thus modulating the ethylene response. Given the association of *Os10g0124700* with variations in RLR, it is plausible that this gene plays a similar role in rice. *Os10g0126600* (*OsFbox512, OsFBX348*) is another F-box domain-containing protein, which has been linked to drought resistance in Shanlan upland rice through its involvement with long non-coding RNA (lncRNA) (Wang et al., 2015). F-box proteins are known to integrate with signaling pathways, including those mediated by phytohormones like ethylene, auxin, and ABA (Koops et al., 2011; Sasaki et al., 2003; Wang et al., 2009; Xu et al., 2021; Yan et al., 2011). The association of *Os10g126600* with RLR suggests its involvement in the regulation of root architecture in response to ethylene, potentially through interactions with other hormones and stress-related pathways. In drought conditions, ethylene production is often elevated, and the ability of roots to adapt to this stress is critical for plant survival. The role of *Os10g0126600* in drought resistance suggests that it may be crucial for modulating root growth under conditions where ethylene levels are altered, helping the plant to balance growth and survival (Yang et al., 2022). This gene's involvement in RLR further supports its importance in the adaptive response of rice roots to environmental stresses, likely through the modulation of hormone signaling networks. *Os10g0128200* (*OsFbox517, OsFBX353*) has also been identified as a gene associated with RLR, though its specific role in ethylene signaling and root development is less characterized compared to the other two genes. However, its significant association with root length reduction suggest that it may be involved in key regulatory process that control root growth in response to ethylene.

Although not all identified candidate genes showed significant differences in RLR among haplotypes, they may still play roles in root growth and ethylene response. Evidence suggests crosstalk among plant hormones, with ethylene inhibiting rice root growth in compacted soil via ABA and auxin mechanisms (Huang et al., 2022). For example, *OsSNDP1* (Os10g0122600), involved in root hair elongation, might contribute to root responses under compacted soil conditions (Huang et al., 2013; Kong et al., 2024). *OsFbx352* (Os10g0127900) influences ABA biosynthesis and catabolism, affecting root architecture under stress (Song et al., 2012). *OsCAF1-15* (Os10g0124200) is involved in stress hormone responses, suggesting a role in ethylene signaling under compaction (Liang et al., 2009). Additional functional studies on these genes, as well as research on root growth in compacted soil conditions, will be necessary to clearly understand the roles of these genes in ethylene signaling and root development. Understanding the specific functions of these genes could provide valuable insights into the genetic basis of stress resilience in rice, with potential applications in breeding programs aimed at improving crop performance under adverse environmental conditions.

5 Conclusion

In this study, we identified a QTL on chromosome 10, *qRLR10*, associated with root length reduction (RLR) in response to ethylene using GWAS in 220 indica rice accessions. Within this QTL region, a cluster of genes encoding F-box proteins were identified as candidate genes, with three of these F-box genes considered particularly promising. These findings highlight the crucial role of ethylene in root development and stress adaptation, and the identification of these candidate genes lays the groundwork for future research and crop improvement strategies.

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

SK: Formal analysis, Writing - original draft, Writing – review & editing. WA: Formal analysis, Writing – review & editing. PM: Formal analysis, Writing – review & editing. VR: Formal analysis, Writing – review & editing. JS: Conceptualization, Supervision, Writing – review & editing. BP: Formal analysis, Writing – review & editing. MB: Conceptualization, Supervision, Writing – review & editing. SW: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. SA:

Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the Office of the Ministry of Higher Education, Science, Research and Innovation; and the Thailand Science Research and Innovation through the Kasetsart University Reinventing University Program 2021. This work was also partially financially supported by the Thailand Rice Science Research Hub of Knowledge (NRCT Grant Number: N34E670027), and the National Science, Research and Innovation Fund, Thailand Science Research and innovation (TSRI) (Grant No.: FFB670076/0337). SK was supported by Thailand Graduate Institute of Science and Technology (TGIST) Scholarships (Grant No. SCA-CO-2563-12042-TH), NSTDA, Thailand, and Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Thailand.

Acknowledgments

The authors are grateful to Assist. Prof. Dr. Wachiraya Imsabai from Postharvest Technology Center, Department of Horticulture, Faculty of Agriculture at Kasetsart University, Kamphaeng Sean campus for supporting the ethylene gas in this study.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1501533/full#supplementary-material

References

Aesomnuk, W., Ruengphayak, S., Ruanjaichon, V., Sreewongchai, T., Malumpong, C., Vanavichit, A., et al. (2021). Estimation of the genetic diversity and population structure of. *Agronomy* 11 (5), 995. doi: 10.3390/agronomy11050995

Arocena, J. M. (2000). Cations in solution from forest soils subjected to forest floor removal and compaction treatments. *For. Ecol. Manage.* 133, 71–80. doi: 10.1016/S0378-1127(99)00299-6

Dello Ioio, R., Linhares, F. S., Scacchi, E., Casamitjana-Martinez, E., Heidstra, R., Costantino, P., et al. (2007). Cytokinins determine arabidopsis root-meristem size by controlling cell differentiation. *Curr. Biol.* 17 (8), 678–682 doi: 10.1016/j.cub.2007.02.047

De Smet, I., and Jürgens, G. (2007). Patterning the axis in plants – auxin in control. *Curr. Opin. Genet. Dev.* 17, 337–343. doi: 10.1016/j.gde.2007.04.012

Dreher, K., and Callis, J. (2007). Ubiquitin, hormones and biotic stress in plants. *Ann. Bot.* 99, 787–822. doi: 10.1093/aob/mcl255

FAO. (2017). *Rice technical manual for extension officers* (Apia: Food and Agriculture Organization of the United Nations).

Feng, Y., Xu, P., Li, B., Li, P., Wen, X., An, F., et al. (2017). Ethylene promotes root hair growth through coordinated *EIN3/EIL1* and *RHD6/RSL1* activity in *Arabidopsis.* Proc. *Natl. Acad. Sci. U. S. A.* 114, 13834–13839. doi: 10.1073/pnas.1711723115

Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M., and Vierstra, R. D. (2002). The F-box subunit of the *SCF E3* complex is encoded by a diverse superfamily of genes in *Arabidopsis.* Proc. *Natl. Acad. Sci. U. S. A.* 99, 11519–11524. doi: 10.1073/pnas.162339999

Guo, X., Zhang, Y., Tu, Y., Wang, Y., Cheng, W., and Yang, Y. (2018). Overexpression of an EIN3-binding F-box protein2-like gene caused elongated fruit shape and delayed fruit development and ripening in tomato. *Plant Sci.* 272, 131–141. doi: 10.1016/j.plantsci.2018.04.016

Habben, J. E., Bao, X., Bate, N. J., Debruin, J. L., Dolan, D., Hasegawa, D., et al. (2014). Transgenic alteration of ethylene biosynthesis increases grain yield in maize under field drought-stress conditions. *Plant Biotechnol. J.* 12, 685–693. doi: 10.1111/pbi.12172

Hassan, F. U., Ahmad, M., Ahmad, N., and Abbasi, M. K. (2007). Effects of subsoil compaction on yield and yield attributes of wheat in the sub-humid region of Pakistan. *Soil Tillage Res.* 96, 361–366. doi: 10.1016/j.still.2007.06.005

He, X. J., Mu, R. L., Cao, W. H., Zhang, Z. G., Zhang, J. S., and Chen, S. Y. (2005). *AtNAC2*, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant J.* 44, 903–916. doi: 10.1111/j.1365-313X.2005.02575.x

Huang, L., Jiang, Q., Wu, J., An, L., Zhou, Z., Wong, C. E., et al. (2020). Zinc finger protein 5 (*ZFP5*) associates with ethylene signaling to regulate the phosphate and potassium deficiency-induced root hair development in Arabidopsis. *Plant Mol. Biol.* 102, 143–158. doi: 10.1007/s11103-019-00937-4

Huang, G., Kilic, A., Karady, M., Zhang, J., Mehra, P., Hu, H., et al. (2022). Ethylene inhibits rice root growth in compacted soil via ABA and auxin mediated mechanisms. *PNAS* 119 (30), e2201072119. doi: 10.1073/pnas.2201072119

Huang, J., Kim, C. M., Xuan, Y. H., Park, S. J., Piao, H. L., Je, B., et al. (2013). *OsSNDP1*, a Sec14-nodulin domain-containing protein, plays a critical role in root hair elongation in rice. *Plant Mol. Biol.* 82, 39–50. doi: 10.1007/s11103-013-0033-4

Ju, C., Yoon, G. M., Shemansky, J. M., Lin, D. Y., Ying, Z. I., Chang, J., et al. (2012). *CTR1* phosphorylates the central regulator *EIN2* to control ethylene hormone signaling from the ER membrane to the nucleus in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19486–19491. doi: 10.1073/pnas.1214848109

Jun, S. H., Han, M. J., Lee, S., Seo, Y. S., Kim, W. T., and An, G. (2004). *OsEIN2* is a positive component in ethylene signaling in rice. *Plant Cell Physiol.* 45, 281–289. doi: 10.1093/pcp/pch033

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., Mccombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 3–10. doi: 10.1186/1939-8433-6-4

Khan, S. R., Abbasi, M. K., and Hussan, A. U. L. (2012). Effect of induced soil compaction on changes in soil properties and wheat productivity under sandy loam and sandy clay loam soils : A greenhouse experiment. *Commun. Soil Sci. Plant Anal.* 43, 2550–2563. doi: 10.1080/00103624.2012.711877

Kong, X., Yu, S., Xiong, Y., Song, X., Nevescanin-Moreno, L., Wei, X., et al. (2024). Root hairs facilitate rice root penetration into compacted layers. *Curr. Biol.* 34, 2039–2048.e3. doi: 10.1016/j.cub.2024.03.064

Koops, P., Pelser, S., Ignatz, M., Klose, C., Marrocco-Selden, K., and Kretsch, T. (2011). *EDL3* is an F-box protein involved in the regulation of abscisic acid signalling in *Arabidopsis thaliana.* J. *Exp. Bot.* 62, 5547–5560. doi: 10.1093/jxb/err236

Le, J., Vandenbussche, F., van der Straeten, D., and Verbelen, J.-P. (2001). In the early response of arabidopsis roots to ethylene, cell elongation is up- and down-regulated and uncoupled from differentiation1. *Plant Physiol.* 125, 519–522. doi: 10.1104/pp.125.2.519

Lechner, E., Achard, P., Vansiri, A., Potuschak, T., and Genschik, P. (2006). F-box proteins everywhere. *Curr. Opin. Plant Biol.* 9, 631–638. doi: 10.1016/j.pbi.2006.09.003

Liang, W., Li, C., Liu, F., Jiang, H., Li, S., Sun, J., et al. (2009). The Arabidopsis homologs of CCR4-associated factor 1 show mRNA deadenylation activity and play a role in plant defence responses. *Cell Res.* 19, 307–316. doi: 10.1038/cr.2008.317

Lin, L. C., Hsu, J. H., and Wang, L. C. (2010). Identification of novel inhibitors of 1-aminocyclopropane-1-carboxylic acid synthase by chemical screening in *arabidopsis thaliana* J. *Biol. Chem.* 285, 33445–33456. doi: 10.1074/jbc.M110.132498

Ma, B., He, S. J., Duan, K. X., Yin, C. C., Chen, H., Yang, C., et al. (2013). Identification of rice ethylene-response mutants and characterization of *MHZ7/OsEIN2* in distinct ethylene response and yield trait regulation. *Mol. Plant* 6, 1830–1848. doi: 10.1093/mp/sst087

McSteen, P. (2010). Auxin and monocot development. *Cold Spring Harb. Perspect. Biol.* 2, a001479. doi: 10.1101/cshperspect.a001479

Mourad, A. M. I., Belamkar, V., and Baenziger, P. S. (2020). Molecular genetic analysis of spring wheat core collection using genetic diversity, population structure, and linkage disequilibrium. *BMC Genomics* 21, 434. doi: 10.1186/s12864-020-06835-0

Müller, M., and Munné-Bosch, S. (2015). Ethylene response factors: A key regulatory hub in hormone and stress signaling. *Plant Physiol.* 169, 32–41. doi: 10.1104/pp.15.00677

Pandey, B. K., Huang, G., Bhosale, R., Hartman, S., Sturrock, C. J., Jose, L., et al. (2021). Plant roots sense soil compaction through restricted ethylene diffusion. *Plant Sci.* 371, 276–280. doi: 10.1126/science.abf3013

Park, H. L., Seo, D. H., Lee, H. Y., Bakshi, A., Park, C., Chien, Y. C., et al. (2023). Ethylene-triggered subcellular trafficking of *CTR1* enhances the response to ethylene gas. *Nat. Commun.* 14, 1–15. doi: 10.1038/s41467-023-35975-6

Quero, G., Gutiérrez, L., Monteverde, E., Blanco, P., Pérez de Vida, F., Rosas, J., et al. (2018). Genome-wide association study using historical breeding populations discovers genomic regions involved in high-quality rice. *Plant Genome* 11, 1–12. doi: 10.3835/plantgenome2017.08.0076

R Core Team. (2024). *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.R-project.org/ (accessed September 1, 2024).

RStudio Team (2020). *RStudio: Integrated Development for R* (Boston, MA: RStudio, PBC). Available at: http://www.rstudio.com/.

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54. doi: 10.1093/pcp/pcs183

Sasaki, A., Itoh, H., Gomi, K., Ueguchi-Tanaka, M., Ishiyama, K., Kobayashi, M., et al. (2003). Accumulation of phosphorylated repressor for gibberellin signaling in an F-box mutant. *Science* 299 (5614), 1896–1898. doi: 10.1126/science.1081077

Song, S., Dai, X., and Zhang, W. H. (2012). A rice F-box gene, *OsFbx352*, is involved in glucose-delayed in rice. *J. Exp. Bot.* 63, 695–709. doi: 10.1093/jxb/err313

Theerawitaya, C., Wanchana, S., Ruanjaichon, V., Tisaram, R., Samphumphuang, T., Sotesaritkul, T., et al. (2022). Determination of traits responding to iron toxicity stress at different stages and genome-wide association analysis for iron toxicity tolerance in rice (*Oryza sativa* L.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.994560

Ueda, H., Ito, T., Inoue, R., Masuda, Y., Nagashima, Y., Kozuka, T., et al. (2020). Genetic interaction among phytochrome, ethylene and abscisic acid signaling during dark-induced senescence in arabidopsis thaliana. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00564

Vaseva, I. I., Qudeimat, E., Potuschak, T., Du, Y., Genschik, P., Vandenbussche, F., et al. (2018). The plant hormone ethylene restricts Arabidopsis growth via the epidermis. *Proc. Natl. Acad. Sci. U. S. A.* 115, E4130–E4139. doi: 10.1073/pnas.1717649115

Wang, X., Kong, H., and Hong, M. (2009). F-box proteins regulate ethylene signaling and more. *Genes Dev.* 23, 391–396. doi: 10.1101/gad.1781609

Wang, H., Niu, Q. W., Wu, H. W., Liu, J., Ye, J., Yu, N., et al. (2015). Analysis of non-coding transcriptome in rice and maize uncovers roles of conserved lncRNAs associated with agriculture traits. *Plant J.* 84, 404–416. doi: 10.1111/tpj.13018

Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinf.* 19 (4), 629-640. doi: 10.1016/j.gpb.2021.08.005

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York). Available at: https://ggplot2.tidyverse.org (accessed September 1, 2024).

Xu, K., Wu, N., Yao, W., Li, X., Zhou, Y., and Li, H. (2021). The biological function and roles in phytohormone signaling of the F-box protein in plants. *Agronomy* 11, 1–17. doi: 10.3390/agronomy11112360

Yamauchi, T., Tanaka, A., Tsutsumi, N., Inukai, Y., and Nakazono, M. (2020). A role for auxin in ethylene-dependent inducible. *Plants* 9, 610. doi: 10.3390/plants9050610

Yan, Y.-S., Chen, X.-Y., Yang, K., Sun, Z.-X., Fu, Y.-P., Zhang, Y.-M., et al. (2011). Overexpression of an F-box protein gene reduces abiotic stress tolerance and promotes root growth in rice. *Mol. Plant* 4, 190–197. doi: 10.1093/mp/ssq066

Yang, X., Liu, C., Niu, X., Wang, L., Li, L., Yuan, Q., et al. (2022). Research on lncRNA related to drought resistance of Shanlan upland rice. *BMC Genomics* 23, 336. doi: 10.1186/s12864-022-08546-0

Yang, C., Ma, B., He, S. J., Xiong, Q., Duan, K. X., Yin, C. C., et al. (2015). *MAOHUZI6/ETHYLENE INSENSITIVE3-LIKE1* and *ETHYLENE INSENSITIVE3-LIKE2* regulate ethylene response of roots and coleoptiles and negatively affect salt tolerance in rice. *Plant Physiol.* 169, 148–165. doi: 10.1104/pp.15.00353

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhang, R., Jia, G., and Diao, X. (2023). geneHapR: an R package for gene haplotypic statistics and visualization. *BMC Bioinf.* 24, 199. doi: 10.1186/s12859-023-05318-9

Zhang, Y., Xu, W., Li, Z., Deng, X. W., Wu, W., and Xue, Y. (2008). F-box protein DOR functions as a novel inhibitory factor for abscisic acid-induced stomatal closure under drought stress in arabidopsis. *Plant Physiol.* 148, 2121–2133. doi: 10.1104/pp.108.126912

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 February 2025
DOI 10.3389/fpls.2025.1539068

OPEN ACCESS

EDITED BY
Hao Zhou,
Sichuan Agricultural University, China

REVIEWED BY
Xiangzheng Fu,
Hunan University, China
Lei Zheng,
The University of Chicago, United States

*CORRESPONDENCE
Jingchao Fan
fanjingchao@caas.cn
Shen Yan
yanshen@caas.cn
Longyu Huang
huanglongyu1@163.com
Jianhua Zhang
zhangjianhua@caas.cn

RECEIVED 03 December 2024
ACCEPTED 20 January 2025
PUBLISHED 18 February 2025

CITATION
Chen X, Fan J, Yan S, Huang L, Zhou G and Zhang J (2025) TAL-SRX: an intelligent typing evaluation method for KASP primers based on multi-model fusion. *Front. Plant Sci.* 16:1539068. doi: 10.3389/fpls.2025.1539068

TAL-SRX: an intelligent typing evaluation method for KASP primers based on multi-model fusion

Xiaojing Chen[1,2], Jingchao Fan[1,2*], Shen Yan[1*], Longyu Huang[2,3,4*], Guomin Zhou[1,2] and Jianhua Zhang[1,2*]

[1]National Agriculture Science Data Center, Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China, [2]National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya, China, [3]Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang, China, [4]Hainan Yazhou Bay Seed Laboratory, Sanya, China

Intelligent and accurate evaluation of KASP primer typing effect is crucial for large-scale screening of excellent markers in molecular marker-assisted breeding. However, the efficiency of both manual discrimination methods and existing algorithms is limited and cannot match the development speed of molecular markers. To address the above problems, we proposed a typing evaluation method for KASP primers by integrating deep learning and traditional machine learning algorithms, called TAL-SRX. First, three algorithms are used to optimize the performance of each model in the Stacking framework respectively, and five-fold cross-validation is used to enhance stability. Then, a hybrid neural network is constructed by combining ANN and LSTM to capture nonlinear relationships and extract complex features, while the Transformer algorithm is introduced to capture global dependencies in high-dimensional feature space. Finally, the two machine learning algorithms are fused through a soft voting integration strategy to output the KASP marker typing effect scores. In this paper, the performance of the model was tested using the KASP test results of 3399 groups of cotton variety resource materials, with an accuracy of 92.83% and an AUC value of 0.9905, indicating that the method has high accuracy, consistency and stability, and the overall performance is better than that of a single model. The performance of the TAL-SRX method is the best when compared with the different integrated combinations of methods. In summary, the TAL-SRX model has good evaluation performance and is very suitable for providing technical support for molecular marker-assisted breeding and other work.

KEYWORDS

KASP fractal evaluation, multi-model fusion, stacking integration, deep learning, hyperparameter tuning

1 Introduction

The kompetitive allele-specific PCR (KASP) technique is capable of realizing the precise identification of site-specific SNP (single nucleotide polymorphism) double allele genotypes in different species genome sample types (Wang et al., 2020), and is widely used in molecular marker-assisted selective breeding, quality testing, variety identification, and stress assessment, etc., because of its unique advantages of flexibility, high efficiency, and low cost (Tang et al., 2022). However, population genotype amplification and segregation are complex and variable, and the evaluation of typing results directly affects the efficiency of KASP marker development (Zhi et al., 2024). Therefore, it is necessary to realize the intelligent and accurate evaluation of the relative independence of population genotyping results, in order to scale up the screening of excellent KASP markers and to improve the efficiency of marker development.

Up to now, there are three main methods for evaluating the typing results in studies utilizing competitive allele-specific PCR technology. The most widely used method is the manual visual judgment of genotyping, which is mainly observed and recognized by agricultural experts or technicians. Due to the flexible and changeable performance of the typing results, the application of this method in breeding practice requires that professionals must have long-term and rich experience in reading typing diagrams, and spend a great deal of time in order to select well-typed KASP markers, so the evaluation process is accompanied by the problems of time-consuming, subjective, and large-scale material validation. For example, Yu et al. (2023) directly observed the fluorescence typing status of different colored dots in the KASP fluorescence detector in the screening and validation of candidate core markers for genotype identification of tobacco varieties, and used the subjectively evaluated well-typed SNP sites as molecular markers in their subsequent studies; Schoonmaker et al. (2023) in the validation of cotton leafroll virus resistance gene association markers, combined with the observation results, proposed that the pure and heterozygous group separation, the cluster within the tight pattern can be proved that the markers are good; Zhao et al. (2023) in the development of Pi2 KASP marker for rice blast resistance gene, because the visual typing results were not clear enough, they further set up negative and positive controls, and initially judged that the marker was feasible after observation. However, because the typing results were scattered, four additional cycles were added to the original testing program to obtain more intuitive and clearer KASP genotyping results; Kalendar et al. (2022) used grid lines with parallel horizontal and vertical axes to partition the KASP genotyping map in an experiment to identify alleles of barley varieties with known genotypes, and observed whether pure and heterozygous genotypes were located in different partitioned regions within the map, respectively, in order to evaluate the accurate validity of the markers. The second category is the method of quantitative assessment of indicator values, although with the development of new SNP genotyping techniques, some scholars have proposed the use of ANOVA to quantitatively assess the differences in indicator values in comparison tests between KASP and TaqMan and other techniques. However, the criteria proposed by this method to use the relative height of the index value to judge the good or bad typing effect are vague, and the application scope in KASP test is limited (Broccanello et al., 2018). The third category is the traditional machine learning approach, usually the SNP genotyping results data sample size is large, the data dimension is high and has the complexity of non-linear relationships, compared with the statistical analysis using a small number of indicators, machine learning as a powerful data-driven framework is more suitable for providing accurate solutions to the complex relationships between a large number of variables in the KASP test results (Kok et al., 2021), such as Chen et al. (2024) proposed an intelligent typing evaluation model for KASP marker primers, which is based on the typing effect level evaluation criteria, introduces K-Means clustering algorithm in the design module to fit the gene population aggregation and classification effects, and finally realizes the intelligent typing result evaluation by logical decision tree algorithm.

However, there are relatively few relevant studies on intelligent typing evaluation of KASP, and the existing evaluation criteria for typing effect levels lack more detailed classification hierarchies, which may lead to a low identification rate of good markers for large-scale screening, thus resulting in a large amount of wasted financial and material resources, which collectively limit the application of KASP technology in assisted breeding work. Second, shallow learning models still have bottlenecks in handling big data, and given that deep learning, as one of the most popular data-driven methods, it may be a useful exploration to apply it to automatically extract and learn the intrinsic features of KASP trial result data (Du et al., 2019; Ahmed et al., 2023).

Based on the above considerations, we propose TAL-SRX, an intelligent typing evaluation method for KASP priming based on multi-model fusion. We utilize the Stacking integrated learning framework to synthesize and apply multiple heterogeneous base learners, and construct a two-layer structure to combine and train with the eXtreme Gradient Boosting (XGBoost) model to drive the data while improving the prediction accuracy of the model. In addition, two deep learning models are selected to be given weights and then introduced into the integrated learning framework to further enhance the ability of the model to master multidimensional features under complex task conditions. The performance of the model is tested by the results of KASP marker typing test of 3399 sets of cotton variety resource materials to verify the effectiveness of the proposed method. Our results not only provide new insights for evaluating the distribution patterns of KASP marker amplification products, but also lay an important foundation for accelerating breeding efforts to precisely localize and select target traits at the molecular level in a variety of crops.

2 Materials and methods

2.1 Experimental data

2.1.1 Raw data

In this study, we selected the KASP marker typing report statistics of the resource materials of cotton varieties produced by SNPline, a high-throughput genotyping detection platform of LGC (Figures 1A, B), and the resource materials included resource varieties, line materials, validated varieties and genetically segregated population materials. The source of the genotyping

report statistics is the Cotton Quality Supervision, Inspection and Testing Center of the Ministry of Agriculture and Rural Affairs, Cotton Research Institute, Chinese Academy of Agricultural Sciences, which contains 319 test results from different SNPs and different DNA samples from 2019 to 2023. Due to the different sample arrangements on the motherboards of each test, we extracted statistics in the format of 94 DNA samples and 2 NTC (negative control reaction without adding DNA samples in the PCR assay) assay data set to build the original dataset, and obtained a total of 3399 sets of statistics.

2.1.2 Criteria for evaluating the typing effect of KASP primers

Based on a large amount of experimental data, we made a detailed division of the evaluation criteria of KASP primer typing effect, as shown in Table 1, and we set a scoring range from 100 to 0 to indicate the primer combination morphology from the best case to the worst case. Specifically, when the competitive primer combination morphology exhibits independent aggregation of pure and heterozygous genotypes and the pure genotypes are located at the maximum of the two axes respectively, and the heterozygous genotypes are located in the center of the line connecting the two pure genotypes, the score is 100. And the score gradually decreases as the degree of the independent aggregation of pure and heterozygous genotypes exhibited by the competitive primer combination morphology decreases, and the segregation of the genotypes decreases, and the score gradually decreases Until complete relative dispersion and inability to accurately typify, the score is 0. This criterion provides a quantitative method for evaluating the combinatorial morphology of competitive primers, and provides a powerful tool for analyzing the amplification efficiency, specificity, and combinatorial competitiveness of genotypes.

2.1.3 Data set construction

The SNP site number, HEX fluorescence signal magnitude relative value X, FAM fluorescence signal magnitude relative value Y and the sample number in the original data set were taken to apply the KASP-IEva model for typing (Figure 1C), and the results of the expert's scoring of the typing effect were used as the criteria. Finally, using the SNP locus number as the number of each data set, the HEX fluorescence signal magnitude relative value X and FAM fluorescence signal magnitude relative value Y of each group were combined and reconstructed into 192 high-dimensional feature variables, and the expert scores were used as the labeling categories, which together comprise the model dataset (Figures 1D, E), of which 84% was used as the training set and 16% as the test set to evaluate the model validity. Table 2 is the sample of the model data set. Fusion method in the Stacking model will be cross-validated using five folds of the training set from which the validation subset will be further divided.

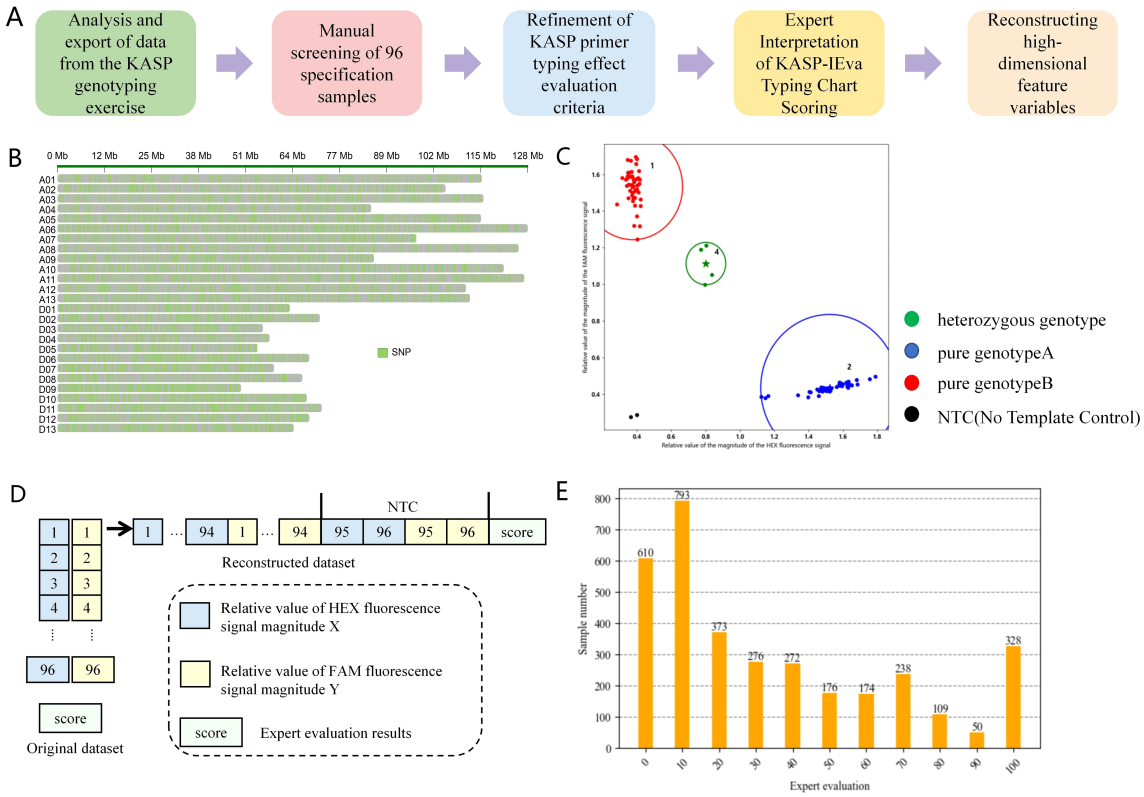


FIGURE 1
Flowchart of dataset construction. **(A)** Steps of dataset integration construction. **(B)** Schematic diagram of genome-wide distribution of SNP variants. **(C)** Example of KASP-IEva model typing diagram. **(D)** KASP marker typing report statistical data structure and sample structure of high-dimensional dataset. **(E)** Distribution of the amount of data in each score of the dataset after scored by the experts.

TABLE 1 Criteria for evaluating the typing effect of KASP primers.

Score	Primer combination morphology
100	NTC has no obvious specificity, pure and heterozygous genotypes are clustered independently, pure genotypes are located at the maximum of each of the two axes, and heterozygous genotypes are located in the center of the line connecting the two pure genotypes
90	NTC has no obvious specificity, pure and heterozygous genotypes are clustered independently, pure genotypes are located at the maximum of each of the two axes, and heterozygous genotypes deviate from the line connecting the two pure genotypes
80	NTC has no obvious specificity, pure and heterozygous genotypes are clustered independently, pure genotypes are located at the maximum of each of the two axes, and heterozygous genotypes are shifted along the line connecting the two pure genotypes
70	NTC has no obvious specificity, pure and heterozygous genotypes are independent, one pure genotype is shifted or trailing along the coordinate axis, and the heterozygous genotype is located in the center of the line connecting the two pure genotypes
60	NTC has no significant specificity, pure and heterozygous genotypes are independent, pure genotypes are shifted along the coordinate axis, and heterozygous genotypes are located in the center of the line connecting two pure genotypes
50	NTC has no obvious specificity, pure and heterozygous genotypes are independent, pure genotypes have a trailing tail, and heterozygous genotypes are located in the center of the line connecting the two pure genotypes
40	NTC is not clearly specific, pure and heterozygous genotypes are partially diffuse or shifted along the axes, but can be typed
30	NTC is not clearly specific, all pure genotypes are relatively diffuse and heterozygous genotypes are shifted, but can be barely differentiated
20	NTC has no obvious specificity, all pure and heterozygous genotypes are relatively diffuse, some genotypic loci are crossed and cannot be accurately typed
10	NTC shows marked specificity
0	Pure and heterozygous genotypes are all relatively diffuse, genotypic loci are crossed and cannot be typed

2.2 TAL-SRX multi-model fusion evaluation methodology

2.2.1 TAL-SRX architecture

In order to improve the accuracy of good markers screening recognition at scale, this paper selects six machine learning algorithms, namely Support Vector Machine (SVM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN), Long Short-Term Memory (LSTM) and Transformer. Two integrated strategies of Stacking and Soft Voting are used for algorithm combination to construct a multi-model hybrid learning method driven by statistical data for KASP marker typing report, which overcomes the defects of a single learning model, enhances the functional nature of feature parsing of high-dimensional data, and possesses a powerful, stable, and comprehensive learning capability. Figure 2 illustrates the flow of the KASP primer typing effect evaluation method.

The TAL-SRX multi-model fusion evaluation method makes full use of the advantages of traditional machine learning algorithms and deep learning algorithms. Firstly, the "base model-metamodel" approach is used to connect the three traditional machine learning algorithms. In order to accelerate the parameter search process and maximize the performance of each learner, different optimization algorithms are used to meet the unique tuning requirements of

TABLE 2 Model data set samples based on KASP marker typing results.

SNP site number SNPID	Relativevalue of HEX fluorescence signal magnitude X and relativevalue of FAM fluorescence signal magnitude Y							Label
	1	2	3	-	190	191	192	
CS01	0.34152	1.23918	0.31628	-	0.36266	0.29725	0.31885	50
CS02	1.23677	1.20181	0.32964	-	0.35707	0.29542	0.29948	30
CS03	0.35668	1.18205	1.20185	-	0.51436	0.96719	0.86575	10
CS04	0.36522	0.31756	1.09486	-	0.36946	0.28544	0.29809	40
CS05	1.42639	0.32153	1.42668	-	0.3707	0.29418	0.28624	60
CS06	1.40921	0.46171	1.41027	-	0.3593	0.30589	0.32609	100
CS07	0.36557	0.32076	0.32602	-	0.39087	0.30391	0.31229	30
CS08	0.71165	0.32425	0.31644	-	0.35485	0.2853	0.2939	50
CS09	1.20593	0.31494	0.31246	-	0.34687	0.28328	0.29112	100
CS10	0.41205	0.38226	0.37761	-	0.35003	0.28565	0.30938	40

different learners. In the Stacking integrated learning framework based on hyperparameter optimization algorithms, the heterogeneous learner parses features of different dimensions within the dataset to generate new feature variables of lesser dimensions, and the meta-learner integrates the data and thus predicts the probabilities. Then the deep learning algorithm is introduced, and the data output processed by ANN is passed to the LSTM layer to flexibly capture complex features while enhancing the model's ability to adapt to heterogeneous samples, and then the Transformer model based on the self-attention mechanism is added to capture the global dependencies between the input variables and enhance the model robustness. Finally, the voting integration model extracts the shared features of the output data of each basic algorithm and obtains the evaluation results of KASP primer typing effect.

2.2.2 Stacking integration based on hyperparameter optimization algorithm

Stacked integration is a widely used integration learning technique, the basic principle of which is to make predictions through a two-layer nested structure (He et al., 2024; Shi et al., 2023). In this paper, the stacked model trains two base learners and uses their prediction results as inputs to the meta-learner to fully exploit the original dataset features to improve the prediction accuracy. For each sub-model, an optimization algorithm is used to improve the model structure, and RS-RF, PSO-SVM and BO-XGBoost are constructed respectively to find the hyper-parameter combinations with the best performance, Figure 3 illustrates the overall framework of Stacking, and the specific implementation process is (Fu et al., 2020):

(1) Data preprocessing is performed on the training set S = {(y_n, x_n), n = 1,…, N} by using the five-fold cross-validation method, which randomly divides the data set S into five equal-sized and disjoint subsets S1, S2,…, S5; (2) one subset is selected as the validation set and the remaining four subsets are used as the training set, and the PSO-SVM model is trained on the training set, and the prediction is performed on the validation set and the prediction result is saved. And then select the remaining four subsets sequentially as validation sets, respectively, and finally obtain five prediction results, which are combined into the set Z1; (3) For the RS-RF model, repeat the operation of step (2) to obtain the set Z2; (4) Combine the prediction results of the base learner in the five-fold cross-validation, and horizontally splice them into the new feature variables Z = {Z1, Z2}, thus realizing the feature

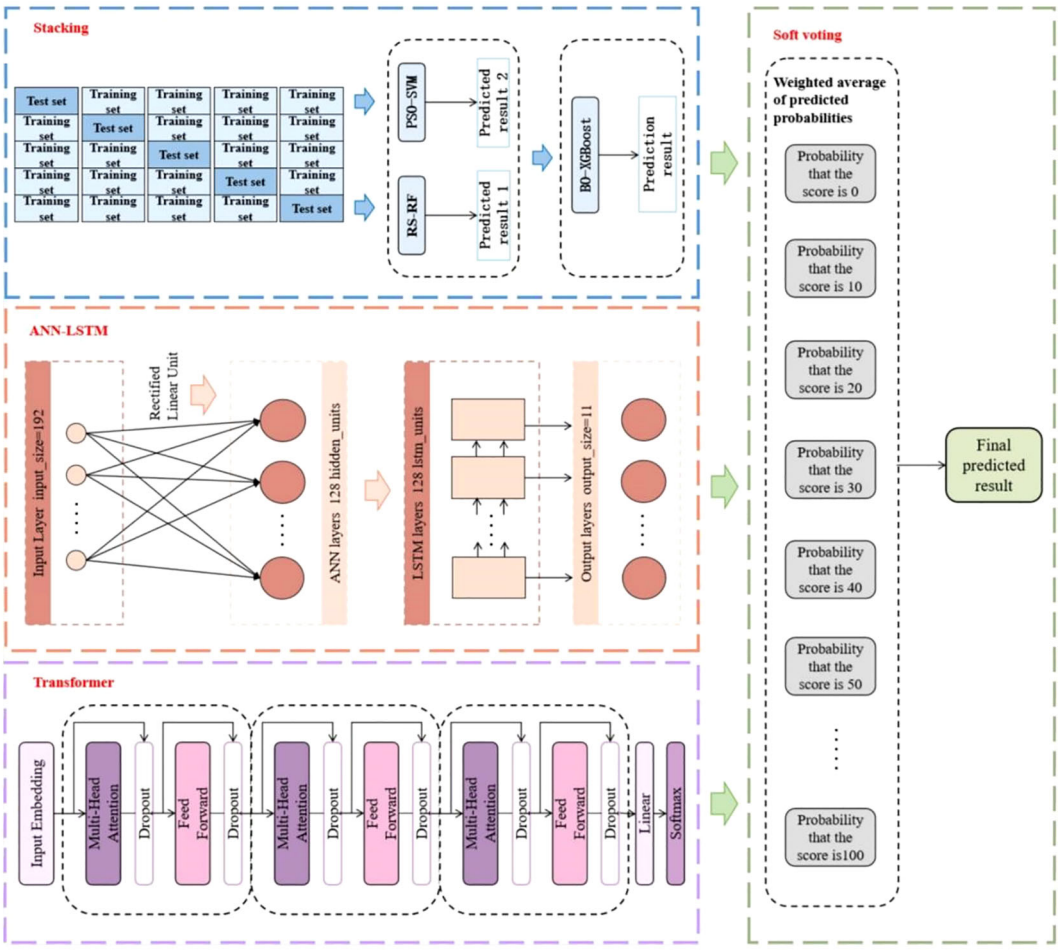


FIGURE 2
Overall architecture of TAL-SRX multi-model fusion approach.

conversion from the base learner to the meta-model, and train the meta-model by using these new features and the original labels y_n; (5) After the model training is completed, use the base model to predict the test set, and the prediction results are input to the meta-model to obtain the final KASP primer typing effect evaluation results (Liu et al., 2024). The base and meta learners are described in detail in the subsequent subsections.

2.2.2.1 RF model based on RS optimization

Random Search (RS) finds the best model parameters by randomly sampling multiple parameter combinations in a predefined parameter space and evaluating the performance of each combination (Shakya et al., 2024). It eliminates the need for gradient information and can explore globally to avoid local optimization.

RF is an algorithm that integrates multiple decision trees based on the idea of bagging (Breiman, 2001). In RF, many trees are constructed using a randomly selected training dataset and a random subset of predictor variables, and the results of each tree are aggregated using the absolute majority voting method to obtain the prediction categories (Speiser et al., 2019). In this paper, we optimize the modeling process of RF based on RS by extracting a sample subset through Bootstrap method, randomly selecting a subset from the sample features to find the optimal splitting point when each node splits, and constructing multiple unpruned decision trees using these sample and feature subsets, and evaluating the performance of the model in each iteration to select the optimal hyper-parameter configurations. Where the absolute majority voting strategy is formulated as:

$$H(x) = \begin{cases} c_j, & if \sum_{i=1}^{T} h_i^j(x) > 0.5 \sum_{k=1}^{N} \sum_{i=1}^{T} h_i^k(x); \\ reject, & otherwise\,. \end{cases} \quad (1)$$

Where h_i is the base learner, i.e., the decision tree, c_j is the category labeling, T is the total number of base learners, $h_i^j(x)$ is the output of h_i on the category labeling c_j, and N is the $h_i^j(x)$ total number of predicted outputs of h_i on sample x.

Each tree in RF has different segmentation features and segmentation points, which has stronger nonlinear data processing ability and overfitting resistance compared with a single decision tree model, and meanwhile, stochastic search can help RF model better adapt to the dataset and improve its robustness and generalization ability.

2.2.2.2 SVM model based on PSO optimization

Particle Swarm Optimization (PSO) originates from the phenomenon of bird flock foraging (Kennedy and Eberhart, 1995), and its essence is to simulate the intelligent behavior of the group, through the information sharing and mutual learning between individuals and groups, each particle adjusts its own position and speed in the search space (Tang et al., 2023), and gradually finds the optimal solution of the problem.

The core idea of SVM is to classify data by solving the maximum margin hyperplane in the feature space, a method known as maximum interval classification. When confronted with linearly indivisible data, SVM maps the data to a higher dimensional space by a kernel method so that it becomes linearly divisible in this new space (Utkin et al., 2016). In this paper, we use the PSO algorithm to optimize the parameters of the SVM modeling process, and determine the optimal parameter model by evaluating the fitness of each particle and updating the optimal positions of the individuals and populations in each iteration, where the decision function of the SVM can be expressed as:

$$f(x) = w \cdot \phi(x) + b \quad (2)$$

Where x is the input eigenvector, $\phi(x)$ is the feature map (defined by the kernel function) mapping the input eigenvector x to the high-dimensional space, w is the normal vector (weight vector) of the hyperplane found in the eigenspace, b is the bias term, and the inner product $w \cdot \phi(x)$ denotes the projection into the eigenspace.

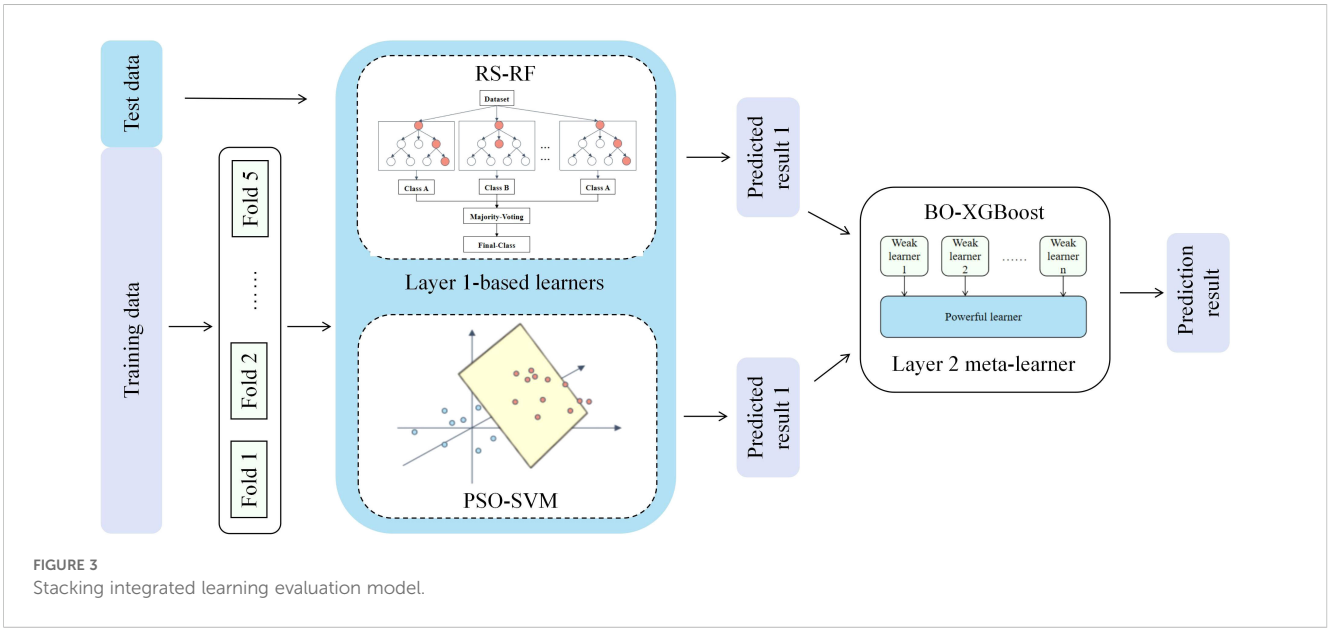


FIGURE 3
Stacking integrated learning evaluation model.

SVM has excellent adaptability to class imbalance problems with large feature differences, and the PSO algorithm, with its outstanding performance of simple implementation, high accuracy and fast convergence, can help SVM to improve its ability to handle high-dimensional complex datasets (Luo et al., 2023).

2.2.2.3 XGBoost model based on BO optimization

Bayesian Optimization (BO) emerges at the forefront of black-box optimization methods due to its high efficiency in finding the global optimal solution with fewer times, and its advantage lies in its ability to infer the posterior distribution of the objective function based on the *a priori* information and the results that have already been observed, achieving a good balance between exploration and exploitation in the search process 0 (Wang et al., 2023).

XGBoost is a scalable end-to-end tree boosting technique in Boosting integrated learning model (Dong et al., 2022; Chen and Guestrin, 2016). Its objective function consists of two parts: the loss function and the regular term, and the core idea of the model is to measure the deviation through the loss function, and use the regular term to control the complexity of the model to avoid overfitting while reducing the deviation. In this paper, we use BO to optimize the modeling process of XGBoost by evaluating the performance of sampling points and updating the Gaussian process agent model in each iteration to select the optimal parameter combinations, so as to gradually approximate the optimal solution of the unknown objective function, where the objective function is defined as follows (Talukder et al., 2024; Cai et al., 2021):

$$L(\phi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{3}$$

Where l is the loss function, Ω is the regularity term, y_i is the true value, $\hat{y}_i$ is the predicted value, and f_k denotes each tree.

XGBoost integrates multiple weak learners into a single strong learner with higher computational speed and better model performance, and BO optimization improves the stability of the XGBoost model on the dataset and reduces the risk of overfitting or underfitting.

2.2.3 Voting integration incorporating deep learning algorithms

Since the KASP marker typing report statistics are produced from different batches, in order to avoid the threat of genetic data heterogeneity to the model robustness, this paper proposes an integration method that introduces the deep learning models ANN-LSTM and Transformer on top of the integration of traditional machine learning algorithms, which will be described in detail in the following two deep learning frameworks:

1. ANN-LSTM: ANN is a class of computational models inspired by biological neural networks, which are widely used in tasks such as data classification (Yang and Wang, 2020; Affonso et al., 2015). Combined with LSTM, we construct a hybrid neural network that uses ANN as a feed-forward neural network layer, integrating the nonlinear relationship capturing ability of ANN and the complex data modeling ability of LSTM (Greff et al., 2016).In the forward propagation process, the input data first passes through the ANN layer, which uses ReLU as the activation function for nonlinear transformation to increase the expressive ability of the model, and the data passes through the 128 hidden units of the layer to adjust the shape of the output after the initial feature extraction, and the LSTM layer is also set up with 128 units, which is able to efficiently capture the intrinsic complex structure of the data by means of the mechanism of the memory unit and the forgetting gate. Finally, the output of LSTM is mapped to the target category space. In the backpropagation process, the cross-entropy loss function is calculated, and the model parameters are updated in training rounds (epochs) by the Adam optimizer. After training, the model is predicted on the test set and the probability distribution of each score is calculated.
2. Transformer: the Transformer is a neural network based mostly on self-attentive mechanisms for capturing global dependencies between input features and focusing on key details of the data (Han et al., 2021). In this architecture, an encoder with feature extraction capability is first defined and the input data is processed through three encoder layers to increase the model depth. During forward propagation, a self-attention mechanism and a feed-forward neural network process the data to capture higher level relationships and extract more complex features. To prevent the model from overfitting, a Dropout layer is added to randomly discard some neurons, and finally the processed features are mapped to the target category space by a linear layer. During the training process, the model calculates the cross-entropy loss by backpropagation and updates the model parameters using the Adam optimizer, and during the testing phase, the model calculates the score category probability for each sample by softmax function and outputs it.

The prediction probabilities of the three models, ANN-LSTM, Transformer and Stacking, are weighted and averaged to obtain the results of the KASP primer typing effect evaluation. With this approach, we obtain a powerful and stable system that contains multiple heterogeneous base learners, and TAL-SRX has the ability to adapt to different scenarios compared to a single prediction model, thus obtaining better prediction results (Ribeiro and dos Santos Coelho, 2020).

3 Test results and analysis

3.1 Test environment

The operating system environment was Windows 11 with a 12th Gen Intel(R) Core(TM) i5-12500 3.00 GHz CPU, 32.0 GB of RAM on board, and a GPU of NVDIA GeForce RTX 3080, 10 GB of RAM. The training environment was created by Anaconda3 and the environment was configured with Python 3.10.13 and PyTorch2.0.0 deep learning framework.

3.2 Indicators for model assessment

In order to verify the effectiveness of the proposed modeling method, the model performance is evaluated using accuracy, precision, recall, F1 score and Cohen's Kappa coefficient. The value range of the first four indicators is 0-1, and each scoring category is calculated separately, and then the macro-averaging (Macro-averaging) method is used to obtain the average indicator values, and the evaluation formulas for each category are as follows:

$$Accuracy = \frac{TP}{TP + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \tag{7}$$

Where *TP*, *FP* and *FN* denote i.e., the number of samples correctly predicted to be that label, the number of samples incorrectly predicted to be that label and the number of samples incorrectly predicted to be other labels, respectively.

The Kappa coefficient provides a more reliable measure of consistency than simple accuracy by taking into account the contingency factor of classification, with values ranging from -1 to 1 on the following scale: 0.81 - 1.00 indicates almost perfect agreement; 0.61 - 0.80 indicates significant agreement; 0.41 - 0.60 indicates moderate agreement; 0.21 - 0.40 indicates fair agreement; 0.00 - 0.20 indicates very low agreement; less than 0 indicates no agreement or very poor agreement. The formula for Kappa coefficient is given below:

$$k = \frac{p_o - p_e}{1 - p_e} \tag{8}$$

Where p_o denotes the proportion of predicted labels that are consistent with actual labels, and *pe* denotes the proportion of consistency under the assumption that predicted and actual labels are stochastically independent.

3.3 Model hyperparameter selection and model performance analysis

The selection of the hyperparameters of the base model is crucial for the improvement of the prediction performance of the integrated model (Yang and Shami, 2020), in order to ensure that the performance of TAL-SRX tends to be optimal, RS is used to optimize the RF model, PSO is used to optimize the SVM model, BO is used to optimize the XGBoost model, the hyperparameter optimization of ANN-LSTM and Transformer are both using trial and error method, and the determination of the voting weights is using the grid search method, and the optimization of the hyperparameters is performed by the grid search method. The parameter optimization of each model is shown in Table 3.

The prediction performance of the base model is improved to some extent after optimization, and the TAL-SRX integrated strategy obtains a more comprehensive learning prediction result for the dataset by integrating diverse base algorithms and observing the data space and structure from different perspectives. Comparative analysis of test set prediction results between the base model and the TAL-SRX integrated algorithm, Figure 4 shows the error size between the predicted value and the actual value of each model, which intuitively demonstrates which samples have been incorrectly evaluated and classified in the prediction process of the model, and the color bar on the right side of the heatmap indicates the correspondence between the color and the size of the error, with the topmost displaying the color of the largest error, and the bottommost being the color of the smallest error. As can be seen from the figure, the TAL-SRX algorithm has fewer blue bars and generally lighter colors, which indicates that the TAL-SRX algorithm has a lower evaluation error rate than the other base models, and the error values are generally smaller than those of a single model.

Table 4 compares the prediction performance of the base model with the TAL-SRX algorithm, and the accuracy of the five single models ranges from 84.93% to 89.15%, which verifies the feasibility of the five algorithms, including RS-RF, SO-SVM, and BO-XGBoost, as the base learner. The accuracy of our proposed TAL-SRX integration algorithm is 92.83%, which is 3.68% higher than the PSO-SVM model with the highest accuracy among the single models, and the precision, recall, and F1 score of this method are 93.82%, 87.65%, and 89.88%, respectively, which are higher than that of the five single models, indicating that the TAL-SRX integration method has a better performance than the base model The integrated performance is significantly improved and has high prediction accuracy.

In order to reflect the model's ability to distinguish between different labels, the ROC curve is drawn to calculate the average AUC value of each model in the case of multiple categories. As shown in Figure 5, the dashed line is the baseline, and the AUC represents the area below the ROC curve. The farther the ROC curve is from the baseline, the larger the AUC is, indicating that the model has a stronger ability to distinguish between samples. Among the six evaluation algorithms, the base model BO-XGBoost has the

TABLE 3 Core hyperparameters of the base model.

Model name	Hyperparametric configuration
RF	n_estimators=202, max_depth=34, min_samples_split=5, min_samples_leaf=1
SVM	C=695.65, kernel="rbf", Gamma=0.0089
XGBoost	colsample_bytree=0.87, learning_rate=0.29, max_depth=7, n_estimators=139, subsample=0.83
ANN-LSTM	lr=0.01, epochs=500
Transformer	nhead=6, dropout=0.01, lr=0.001, epochs=500
TAL-SRX Soft Voting	weight_transformer=0.25, weight_stacking=0.95, weight_ANN-LSTM=0.55

10.3389/fpls.2025.1539068

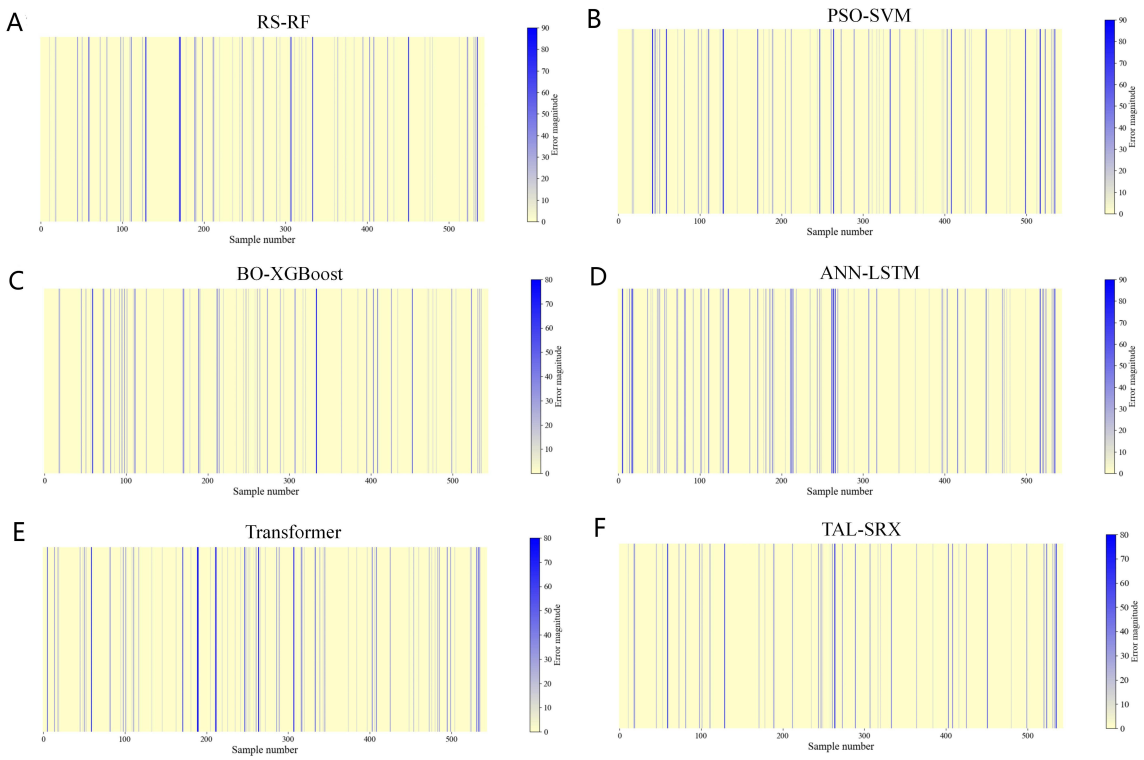


FIGURE 4
Error heatmap of the base model and the TAL-SRX approach. **(A)** Error heatmap of RS-RF. **(B)** Error heatmap of PSO-SVM. **(C)** Error heatmap of BO-XGBoost. **(D)** Error heatmap of ANN-LSTM. **(E)** Error heatmap of Transformer. **(F)** Error heatmap of TAL-SRX.

second highest AUC value of 0.9891, and the TAL-SRX strategy has the highest AUC value, which is 0.0014-0.0276 higher than that of the single learner, and on the whole, the TAL-SRX can effectively recognize and evaluate the KASP typing samples with different labels.

The models are further analyzed for different label classification consistency, and the box violin plot of Kappa coefficient for each model is shown in Figure 6. It can be seen that the median Kappa coefficient of each single model is distributed between 0.7 and 0.9, and the distribution is denser near 0.8. After applying the integrated method, the median Kappa coefficient of TAL-SRX is higher than 0.9 and densely distributed near 0.9, and the overall distribution pattern is tighter, which is a very significant improvement, indicating that the predicted labels of the TAL-SRX method are more consistent with the actual labels are more consistent and the model is more stable. Among them, the anomalous value of Kappa coefficient for 90-point labels is affected by the uneven distribution of samples in the dataset.

In summary, the experimental results prove that compared with the base model, the TAL-SRX method has higher evaluation accuracy, consistency and stability. Analyzing from the theoretical point of view, on the one hand, TAL-SRX makes full use of the differences of the heterogeneous models and takes advantage of the powerful nonlinear modeling ability of the base model, so as to be able to comprehensively capture the detailed characteristics of the data, and, on the other hand, the base model takes advantage of the differences. For example, BO-XGBoost is based on the gradient boosting framework, which enhances its generalization ability and

TABLE 4 Comparison of base model and TAL-SRX test set performance.

Model	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)
RS-RF	88.97	93.2	84.51	87.71
PSO-SVM	89.15	92.8	84.34	87.41
BO-XGBoost	88.97	90.91	83.89	86.38
ANN-LSTM	84.93	86.09	82.2	83.01
Transformer	86.58	84.95	82.71	83.48
TAL-SRX	92.83	93.82	87.65	89.88

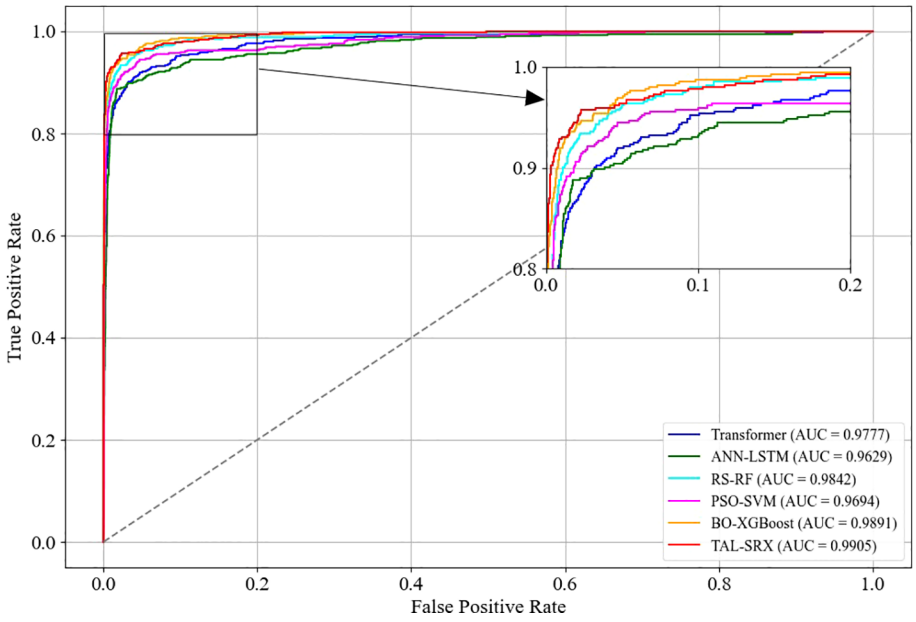


FIGURE 5
ROC curves for the base model and the TAL-SRX approach.

robustness to noisy data by controlling the depth of the tree, and the characteristics of each internal model enhance the integrated expression ability of TAL-SRX, so that TAL-SRX has obvious advantages over a single model, and is able to achieve effective prediction of the typing effect of KASP primers.

3.4 Comparative analysis of integrated combined approach ablation

In order to demonstrate the effectiveness of the two integration strategies in TAL-SRX, ablation comparison tests are performed

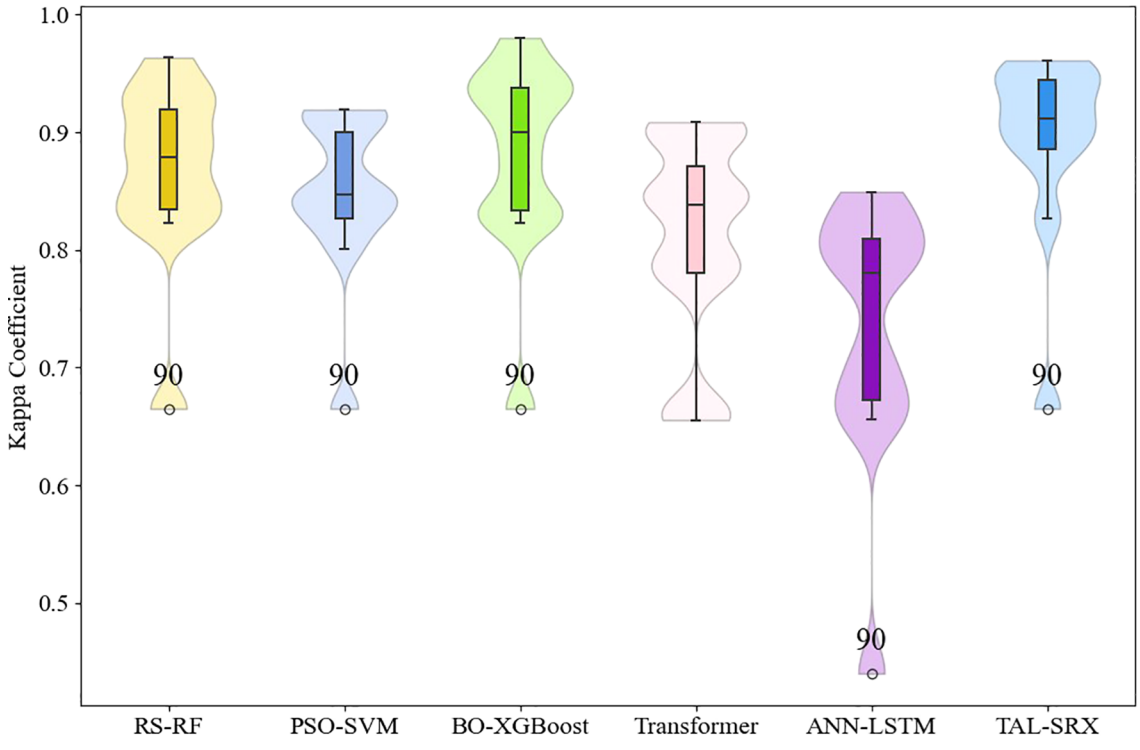


FIGURE 6
Kappa coefficients for the base model and the TAL-SRX approach.

using different integration combination approaches. Firstly, Stacking integration and deep learning model integration using traditional machine learning algorithms alone are used for prediction, in which the integration combinations of ANN-LSTM and Transformer are used to determine the best weight combinations using grid search, which are 0.45 and 0.4, respectively, and, secondly, the ANN-LSTM integration is added on top of Stacking, and the best weight combinations are determined using grid search weight combinations were 0.25 and 0.05, while the Transformer integration was added on top of Stacking and grid search was used to determine the optimal weight combinations of 0.35 and 0.1.

The test of each integration combination method is shown in Figure 7, which shows that compared to the Stacking integration using traditional machine learning algorithms alone, the accuracy, precision, recall and F1 score of TAL-SRX are 2.02%, 1.00%, 1.86% and 1.58% higher, respectively, which indicates that the deep learning algorithms have the ability to extract high-level features and key details when processing complex data is stronger. Compared to the integrated combination of deep learning alone, the accuracy, precision, recall, and F1 score of TAL-SRX are 4.78%, 8.07%, 3.60%, and 5.23% higher, respectively, which is attributed to the fact that Stacking integrates the advantages of more learners and reduces the bias due to data segmentation through five-fold cross-validation in the model training stage, thus improving the prediction Accuracy. In addition, compared to the simple Stacking model, the accuracy of adding two deep learning algorithms respectively is improved, and the TAL-SRX method, which introduces the two together, has the highest accuracy, which indicates that the Stacking, ANN-LSTM, and Transformer algorithms have their own strengths and positive effects on the specific types of data on the dataset, and that the three algorithms scoring the right samples do not completely cover each other, thus effectively improving the model scoring performance.

3.5 Comparative analysis of TAL-SRX methodology and expert scoring prediction results

Figure 8 demonstrates the scoring and sample size relationship curves of expert scoring and TAL-SRX method, which shows that the trajectories of the orange and blue curves are basically the same, and the sample size distribution curves of expert evaluation and algorithmic evaluation have similar trends in each score band, indicating that there is a high degree of consistency between the TAL-SRX evaluation and the expert evaluation, and that the evaluation results of the multi-model fusion method can reflect well the success rate of the development of KASP markers, and then effectively screen out the well-typed markers. However, it is worth noting that there are obvious deviations in the curve trajectories of the two score regions of 0 and 10, which is due to the fact that the Transformer, which has the strongest global feature capturing ability, has the smallest weight in the voting and contributes less to the final prediction results, which makes the performance of the integrated model performance on the samples with dispersed distribution of allele genotypes weaker.

4 Discussion

The demand for molecular marker-assisted breeding has driven the development of KASP markers, but the complexity and diversity of allele genotype distribution patterns in the typing results have seriously hindered the large-scale screening of markers, and the existing manual discrimination and algorithms do not have the ability to analyze the data of the typing results with high accuracy and efficiency, making it challenging to develop good molecular markers in bulk. Deep learning, as an important branch of machine learning, has been widely used in various fields, however, there has

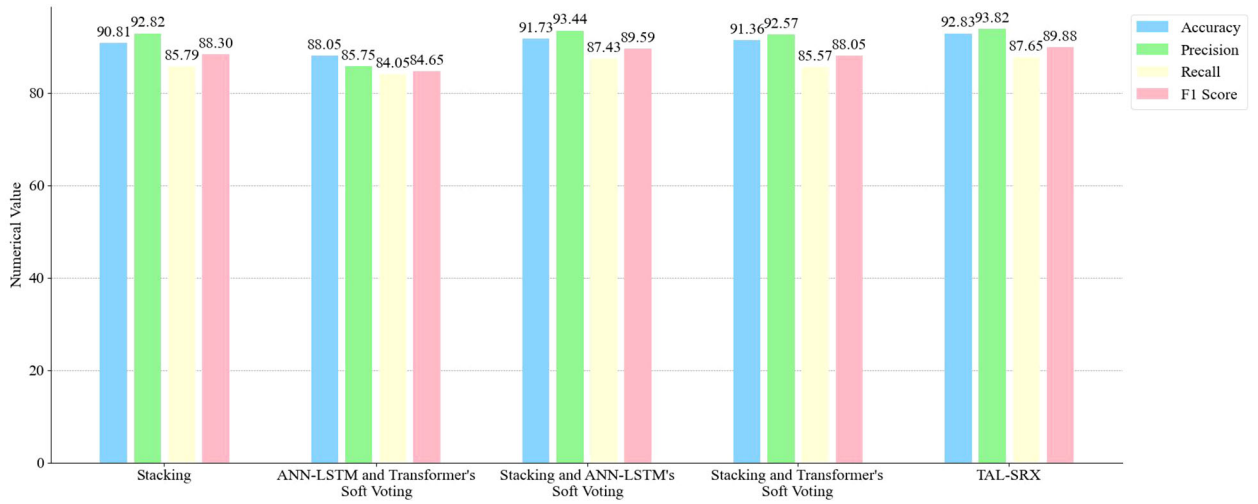


FIGURE 7
Performance comparison of different integration combinations.

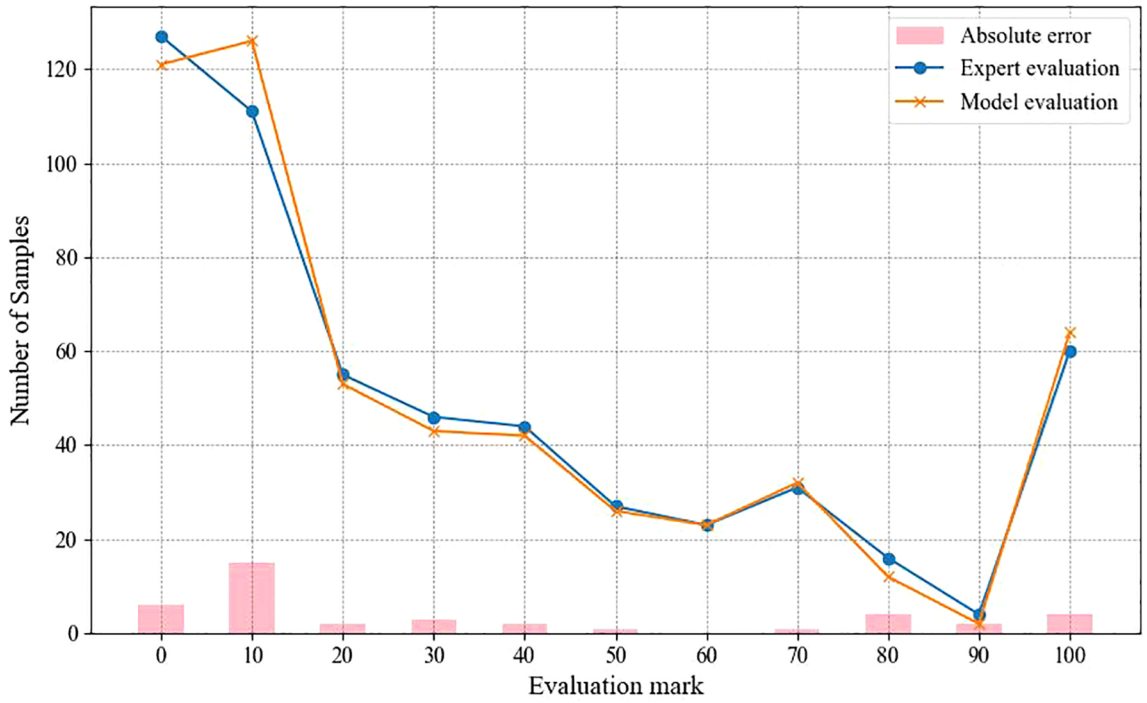


FIGURE 8
Comparison of TAL-SRX methodology and expert scoring results.

not been any study using deep learning algorithms to evaluate population genotype amplification and segregation, so it is of great significance to use an integrated strategy to introduce a deep learning model to evaluate the effect of KASP primer typing. In this study, two integration strategies, stacking and soft voting, were used for algorithm combination to construct a multi-model hybrid learning method driven by statistics of KASP marker typing report, and the following conclusions were drawn from the experimental study:

1. Refined the evaluation criteria of KASP primer typing effect, and constructed a competitive allele distribution pattern evaluation system from 0-100 points. This evaluation criterion provides a powerful quantitative tool for analyzing the amplification efficiency, specificity and combinatorial competitiveness of genotypes, which is not only applicable to the results of the KASP test in cotton, but also provides a reference to the resources of other varieties, and can be used to screen for the superior KASP markers that meet the characteristics of different crops.
2. In the multi-model fusion method proposed in this paper, the Stacking integrated learning model is connected using the "base model-metamodel" approach, which gives full play to the respective advantages of heterogeneous learners, and different optimization algorithms are adopted for different learners to accelerate the parameter search process and maximize the performance of each model, and the training process adopts five-fold cross-validation to enhance the model stability. The ANN-LSTM hybrid neural network model combines the ability of nonlinear relationship capture and complex feature extraction, which enhances the adaptability of the model to heterogeneous samples, and the Transformer model is based on the self-attention mechanism, which captures the global dependencies in the high-dimensional feature space, and improves the robustness of TAL-SRX.
3. The proposed model was trained and tested using 3399 sets of KASP marker typing report statistics of cotton varietal resource materials, and the performance comparison between the TAL-SRX method and the base model was carried out firstly, and the TAL-SRX algorithm had a lower error rate and error value than that of the base model, with an accuracy of 92.83% and an AUC value of 0.9905, which was of high evaluative accuracy, consistency and stability, and the performance is significantly better than the single model. Secondly, the impact of each algorithm on the model prediction performance was investigated by integrating and combining ways of ablation comparison, and the experimental results show that the two integration strategies can enhance the model performance, and different sub-algorithms have positive effects on the model. Finally, a comparison between the TAL-SRX method and expert scoring was carried out, and the evaluation results of this method have a high consistency between the number of samples on each score band and the expert evaluation. Therefore, the TAL-SRX method has good evaluation performance.
4. In this study, we explored an intelligent KASP primer typing effect evaluation method using deep learning algorithms and stacking integration, and verified its effectiveness through experiments, which provided the

possibility of applying the algorithm to KASP typing data of various crop variety resources.

5 Conclusion

In this study, an intelligent typing evaluation method for KASP primers with higher accuracy, called TAL-SRX, is proposed, which can be used to provide more accurate data support for improving the success rate of KASP marker development. The method first introduces an optimization algorithm to construct the Stacking integrated learning framework through three improved models, RS-RF, PSO-SVM and BO-XGBoost, which improves the model prediction performance to some extent and enhances the stability and prediction accuracy of the model. Then, a hybrid neural network is constructed using ANN and LSTM to capture nonlinear relationships and extract complex features to strengthen the model's ability to adapt to discrepant samples, and the Transformer algorithm with a multi-attention mechanism is introduced to capture the global dependencies in the high-dimensional feature space, which effectively enhances the model's robustness. Comparison tests were conducted on the test set, and the distribution of the number of samples on each score band was in high agreement between the TAL-SRX method and the expert evaluation results, with an evaluation accuracy of 92.83%, an AUC value of 0.9905, and a lower error rate and error value than that of the base model, and the overall performance was significantly better than that of each single model. In the integrated combination approach ablation comparison test, the TAL-SRX integrated method showed better accuracy and was suitable for KASP primer intelligent typing evaluation. However, it is worth noting that the model proposed in this paper mainly addresses the problem of data evaluation for medium-sized experiments, and is only applicable to the experimental dataset where the mother plate is a 96-well plate, and in the actual operation of KASP, the researchers will also obtain the experimental data of different well plates on a larger scale and a large scale, so it is necessary to further adjust the structure of the model in the subsequent experiments to enhance the generalization ability of the model for more datasets.

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

XC: Conceptualization, Data curation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JF: Writing – review & editing. SY: Supervision, Writing – review & editing. LH: Data curation, Writing – review & editing. GZ: Writing – review & editing. JZ: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key R&D Program (2022YFF0711805, 2022YFF0711801); National Natural Science Foundation of China (31971792); Special Program for Southern Propagation of the National Institute of Southern Propagation of the Chinese Academy of Agricultural Sciences in Sanya (YBXM2409, YBXM2410, YBXM2312, ZDXM2311); Special Program for the Fundamental Scientific Research Operating Expenses Special Project (JBYW-AII-2024-05, JBYW-AII-2023-06); Scientific and Technological Innovation Project of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2024-AII, CAAS-ASTIP-2023-AII); Scientific and Technological Special Funding for Sanya Yazhou Bay Science and Technology City (SCKJ-JYRC-2023-45).

Acknowledgments

We thank all the participants in this study.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Affonso, C., Sassi, R. J., and Barreiros, R. M. (2015). Biological image classification using rough-fuzzy artificial neural network. *Expert Syst. Appl.* 42, 9482–9488. doi: 10.1016/j.eswa.2015.07.075

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., et al. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artif. Intell. Rev.* 56, 13521–13617. doi: 10.1007/s10462-023-10466-8

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Broccanello, C., Chiodi, C., Funk, A., McGrath, J. M., Panella, L., and Stevanato, P. (2018). Comparison of three PCR-based assays for SNP genotyping in plants. *Plant Methods* 14, 1–8. doi: 10.1186/s13007-018-0295-6

Cai, L., Ren, X., Fu, X., Peng, L., Gao, M., and Zeng, X. (2021). iEnhancer-XG: interpretable sequence-based enhancers and their strength predictor. *Bioinformatics* 37, 1060–1067. doi: 10.1093/bioinformatics/btaa914

Chen, T., and Guestrin, C. (2016). "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (San Francisco, CA, USA: ACM), 785–794. doi: 10.1145/2939672.2939785

Chen, X., Huang, L., Fan, J., Yan, S., Zhou, G., and Zhang, J. (2024). KASP-IEva: an intelligent typing evaluation model for KASP primers. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1293599

Dong, J., Chen, Y., Yao, B., Zhang, X., and Zeng, N. (2022). A neural network boosting regression model based on XGBoost. *Appl. Soft Computing* 125, 109067. doi: 10.1016/j.asoc.2022.109067

Du, S., Li, T., Yang, Y., and Horng, S. J. (2019). Deep air quality forecasting using hybrid deep learning framework. *IEEE Trans. Knowledge Data Eng.* 33, 2412–2424. doi: 10.1109/TKDE.2019.2954510

Fu, X., Cai, L., Zeng, X., and Zou, Q. (2020). StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency. *Bioinformatics* 36, 3028–3034. doi: 10.1093/bioinformatics/btaa131

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Trans. Neural Networks Learn. Syst.* 28, 2222–2232. doi: 10.1109/TNNLS.2016.2582924

Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., and Wang, Y. (2021). Transformer in transformer. *Adv. Neural Inf. Process. Syst.* 34, 15908–15919. doi: 10.48550/arXiv.2103.00112

He, Y., Zhang, H., Dong, Y., Wang, C., and Ma, P. (2024). Residential net load interval prediction based on stacking ensemble learning. *Energy* 296, 131134. doi: 10.1016/j.energy.2024.131134

Kalendar, R., Baidyussen, A., Serikbay, D., Zotova, L., Khassanova, G., Kuzbakova, M., et al. (2022). Modified "Allele-specific qPCR" Method for SNP genotyping based on FRET. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.747886

Kennedy, J., and Eberhart, R. (1995). "November. Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4. (Piscataway, NJ, USA: ieee), 1942–1948. doi: 10.1109/ICNN.1995.488968

Kok, Z. H., Shariff, A. R. M., Alfatni, M. S. M., and Khairunniza-Bejo, S. (2021). Support vector machine in precision agriculture: a review. *Comput. Electron. Agric.* 191, 106546. doi: 10.1016/j.compag.2021.106546

Liu, T., Zhu, H., Yuan, Q., Wang, Y., Zhang, D., and Ding, X. (2024). Prediction of photosynthetic rate of greenhouse tomatoes based on multi-model fusion strategy. *Trans. Chin. Soc. Agric. Machinery* 55, 337–345. doi: 10.6041/j.issn.1000-1298.2024.04.033

Luo, H., Fang, Y., Wang, J., Wang, Y., Liao, H., Yu, T., et al. (2023). Combined prediction of rockburst based on multiple factors and stacking ensemble algorithm. *Underground Space* 13, 241–261. doi: 10.1016/j.undsp.2023.05.003

Ribeiro, M. H. D. M., and dos Santos Coelho, L. (2020). Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl. soft computing* 86, 105837. doi: 10.1016/j.asoc.2019.105837

Schoonmaker, A. N., Hulse-Kemp, A. M., Youngblood, R. C., Rahmat, Z., Atif Iqbal, M., Rahman, M. U., et al. (2023). Detecting Cotton Leaf Curl Virus Resistance Quantitative Trait Loci in Gossypium hirsutum and iCottonQTL a New R/Shiny App to Streamline Genetic Mapping. *Plants* 12, 1153. doi: 10.3390/plants12051153

Shakya, D., Deshpande, V., Safari, M. J. S., and Agarwal, M. (2024). Performance evaluation of machine learning algorithms for the prediction of particle Froude number (Frn) using hyper-parameter optimizations techniques. *Expert Syst. Appl.* 256, 124960. doi: 10.1016/j.eswa.2024.124960

Shi, J., Li, C., and Yan, X. (2023). Artificial intelligence for load forecasting: A stacking learning approach based on ensemble diversity regularization. *Energy* 262, 125295. doi: 10.1016/j.energy.2022.125295

Speiser, J. L., Miller, M. E., Tooze, J., and Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* 134, 93–101. doi: 10.1016/j.eswa.2019.05.028

Talukder, M. A., Islam, M. M., Uddin, M. A., Hasan, K. F., Sharmin, S., Alyami, S. A., et al. (2024). Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. *J. big Data* 11, 33. doi: 10.1186/s40537-024-00886-w

Tang, J., Duan, H., and Lao, S. (2023). Swarm intelligence algorithms for multiple unmanned aerial vehicles collaboration: A comprehensive review. *Artif. Intell. Rev.* 56, 4295–4327. doi: 10.1007/s10462-022-10281-7

Tang, W., Lin, J., Wang, Y., An, H., Chen, H., Pan, G., et al. (2022). Selection and validation of 48 KASP markers for variety identification and breeding guidance in conventional and hybrid rice (*Oryza sativa* L.). *Rice* 15, 48. doi: 10.1186/s12284-022-00594-0

Utkin, L. V., Chekh, A. I., and Zhuk, Y. A. (2016). Binary classification SVM-based algorithms with interval-valued training data using triangular and Epanechnikov kernels. *Neural Networks* 80, 53–66. doi: 10.1016/j.neunet.2016.04.005

Wang, D., Yang, T., Liu, R., Li, N., Wang, X., Sarker, A., et al. (2020). RNA-Seq analysis and development of SSR and KASP markers in lentil (*Lens culinaris* Medikus subsp. *culinaris*). *Crop J.* 8, 953–965. doi: 10.1016/j.cj.2020.04.007

Wang, X., Jin, Y., Schmitt, S., and Olhofer, M. (2023). Recent advances in Bayesian optimization. *ACM Computing Surveys* 55, 1–36. doi: 10.1145/3582078

Yang, L., and Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415, 295–316. doi: 10.1016/j.neucom.2020.07.061

Yang, G. R., and Wang, X. J. (2020). Artificial neural networks for neuroscientists: a primer. *Neuron* 107, 1048–1070. doi: 10.1016/j.neuron.2020.09.005

Yu, S., Cao, L., Wang, S., Liu, Y., Bian, W., and Ren, X. (2023). Development core SNP markers for tobacco germplasm genotypin. *Biotechnol. Bull.* 39, 89–100. doi: 10.13560/j.cnki.biotech.bull.1985.2022-0810

Zhao, C., Cong, S., Liang, S., Xie, H., Lu, W., Xiao, W., et al. (2023). Development and application of KASP marker for rice blast resistance gene Pi. *J. South China Agric. Univ.* 44, 725–734. doi: 10.7671/j.issn.1001-411X.202304026

Zhi, L., Gong, X., Zhang, H., Liu, J., Cao, S., Zhang, Y., et al. (2024). Identification of QTL for alkylresorcinols in wheat and development of KASP markers for marker-assisted selection of health-promoting varieties. *J. Agric. Food Chem.* 72 (5), 1234–1245. doi: 10.1021/acs.jafc.4c04674

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 February 2025
DOI 10.3389/fpls.2025.1537605

OPEN ACCESS

EDITED BY
Hao Zhou,
Sichuan Agricultural University, China

REVIEWED BY
Shenghui Zhou,
Chinese Academy of Agricultural Sciences (CAAS), China
Yongming Chen,
Peking University, China

*CORRESPONDENCE
Xinkun Hu
huxinkun123@163.com

RECEIVED 01 December 2024
ACCEPTED 13 February 2025
PUBLISHED 28 February 2025

CITATION
Wu L, Wang J, Shen S, Yang Z and Hu X (2025) Transcriptomic analysis of two Chinese wheat landraces with contrasting Fusarium head blight resistance reveals miRNA-mediated defense mechanisms.
Front. Plant Sci. 16:1537605.
doi: 10.3389/fpls.2025.1537605

Transcriptomic analysis of two Chinese wheat landraces with contrasting Fusarium head blight resistance reveals miRNA-mediated defense mechanisms

Lijuan Wu[1,2], Junqiang Wang[1], Shian Shen[1], Zaijun Yang[3] and Xinkun Hu[1*]

[1]Institute of Ecology, China West Normal University, Nanchong, Sichuan, China, [2]College of Agronomy, Sichuan Agricultural University, Chengdu, Sichuan, China, [3]College of Life Science, China West Normal University, Nanchong, Sichuan, China

Introduction: Fusarium head blight (FHB), caused primarily by *Fusarium graminearum* (*Fg*), poses a significant threat to wheat production. It is necessary to deeply understand the molecular mechanisms underlying FHB resistance in wheat breeding.

Methods: In this study, the transcriptomic responses of two Chinese wheat landraces—Wuyangmai (WY, resistant) and Chinese Spring (CS, susceptible)—to *F. graminearum* infection were examined using RNA sequencing (RNA-seq). Differential expression of mRNAs, long non-coding RNAs (lncRNAs), circular RNAs (circRNAs), and microRNAs (miRNAs) was analyzed at 3 and 5 days post-*Fg* inoculation (dpi).

Results: The results showed that WY exhibited a targeted miRNA response, primarily modulating defense-related pathways such as glutathione metabolism and phenylpropanoid biosynthesis, which are crucial for oxidative stress regulation and pathogen defense response. In contrast, CS displayed a broader transcriptional response, largely linked to general metabolic processes rather than immune activation. Notably, the up-regulation of genes involved in oxidative stress and immune defense in WY confirmed its enhanced resistance to FHB. The integrated analysis of miRNA-mRNA interactions highlighted miRNAs as central regulators of defense mechanisms in WY, particularly at later stages of infection. These miRNAs targeted genes involved in immune responses, while lncRNAs and circRNAs played a more limited role in the regulation of defense responses. The GO and KEGG pathway enrichment analyses further revealed that WY enriched for plant-pathogen interaction and secondary metabolite biosynthesis pathways, which are crucial for pathogen resistance. In contrast, CS prioritized metabolic homeostasis, suggesting a less effective defense strategy.

Discussion: Overall, this study underscores the critical role of miRNA-mediated regulation in FHB resistance in WY. These insights into miRNA-mediated regulatory mechanisms provide a molecular basis for breeding FHB-resistant wheat varieties and highlight miRNA-mRNA interactions as promising targets for enhancing disease resilience.

KEYWORDS

Sichuan wheat landraces, Fusarium head blight, lncRNAs, circRNAs, miRNAs, RNA sequencing

1 Introduction

As one of the vital food crops, wheat (*Triticum aestivum* L.) supplies approximately 20% of the caloric intake for the global population (Ma et al., 2020). Wheat Fusarium head blight (FHB), destructive fungal disease primarily caused by *F. graminearum* (*Fg*), leads to a significant yield losses and quality deteriorates during epidemic years (Zhu et al., 2019). Additionally, FHB contaminates grain with harmful mycotoxins, including nivalenol and deoxynivalenol (DON), which seriously endanger the health of humans and livestock (Chen et al., 2019; Hu et al., 2023). Recent climate changes and wheat farming practices have increased the frequency and severity of FHB outbreaks (Ma et al., 2020). Identification and utilization of resistant germplasms in breeding programs are one of the most sustainable and economical approach to manage FHB.

FHB resistance is a complex quantitative trait governed by multiple genes and influenced by genotype-environment interactions (Zhang et al., 2021). Approximately 500 quantitative trait loci (QTLs) related to FHB resistance were identified from wheat and its relatives, and distributed across all 21 wheat chromosomes (Buerstmayr et al., 2009, Buerstmayr et al., 2020). However, most of these QTLs have minor effects and limited breeding value in addition to seven widely recognized major resistance genes (*Fhb1* to *Fhb7*). Recently, two additional major QTLs, *Fhb8* and *Fhb9*, were identified and associated with specific resistance types (Wang et al., 2024a; Zhang et al., 2024).

Despite these advances, only *Fhb1* and *Fhb7* have been cloned using map-based techniques and have shown significant resistance. *Fhb1* confers resistance through a histidine-rich calcium-binding protein (His or TaHRC), while the exact mechanism remains unclear (Su et al., 2019; Li et al., 2019a). *Fhb7*, derived from *Thinopyrum* species, encodes a glutathione-S-transferase (GST) that detoxifies trichothecene toxins (Wang et al., 2020a). However, wheat varieties incorporated with *Fhb7* are not yet widely available, and those with *Fhb1* displayed a moderate resistance, likely due to interactions with parent genotype (Li et al., 2019b). FHB resistance is generally inherited quantitatively and affected by multiple minor genes (Bai and Shaner, 2004). Other resistance genes, such as *TaFROG* and *TaABCC3*, have been identified and characterized for their roles in response to DON and *F. graminearum* (Perochon et al., 2015; Walter et al., 2015). To effectively combat FHB in wheat effectively, more effort is still needed to identify new resistant genes.

Transcriptome analysis is a powerful tool for studying FHB resistance mechanisms, offering insights into wheat-pathogen interactions. Previous studies utilized microarray techniques to analyze transcriptomic responses (Bernardo et al., 2007), while recent ones have applied RNA-seq to reveal new pathways and defense-related genes (Xiao et al., 2013; Brauer et al., 2019; Pan et al., 2018). Small RNAs, particularly miRNAs, have been found to regulate the response of wheat to *F. graminearum*, either by silencing fungal genes or modulating host defense pathways (MaChado et al., 2018; Jin et al., 2020; Biselli et al., 2018). However, little is known about the role of other non-coding RNAs, such as long non-coding RNAs (lncRNAs) and circular RNAs (circRNAs), in response to FHB resistance.

Chinese wheat landraces are gene pools for FHB-resistant. Some of the well-known resistant resources including Wangshuibai, Sanyuehuang, Shuilizhan, and Wuyangmai (WY), originated in China (Wan et al., 1997). WY, in particular, is an FHB-resistant landrace from Yibin city in Sichuan, China. However, the loci against FHB in WY have not yet been genetically identified, and the mechanisms underlying FHB resistance remain unknown. In this study, RNA-seq analysis was performed on WY (an FHB-resistant genotype) and Chinese Spring (CS, an FHB-susceptible genotype) in response to *F. graminearum* at 3 and 5 dpi to identify lncRNAs, circRNAs, miRNAs, and mRNAs involved in host-pathogen interactions. Differentially expressed miRNAs and mRNAs were used as central elements to target other differentially expressed RNAs, including lncRNAs and host genes associated with circRNAs. Additionally, differential expression analysis, along with Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses, was conducted on significantly up-regulated genes in CS and WY at different time points. Finally, the regulatory relationships between differentially expressed miRNAs and mRNAs were characterized, leading to the construction of a miRNA-mRNA targeting network map. The results provide valuable information

for gaining a deeper understanding the FHB resistance differences of the two Sichuan wheat landraces, WY and CS.

2 Materials and methods

2.1 Plant materials and cultivation conditions

The two Sichuan wheat landraces, WY and CS, which exhibit high resistance (HR) and high susceptibility (HS) to FHB, respectively (Wan et al., 1997), were used in this study. The two landraces were provided by the Triticeae research Institute of Sichuan Agricultural University of China, and showed similar plant heights, spikes, heading and flowering dates, minimizing the potential discrepancies due to variations in plant architecture and development periods. The experiment was conducted at the breeding field of China West Normal University (Nanchong, Sichuan Province, China; 30°48′N, 106°05′E) during the 2022-2023 growing season. A randomized block design was implemented, with each genotype sown in a plot containing ten rows. Each row was 1.5 meters long, with 20 seedlings spaced 0.3 meters apart. Field cultivation and fertilization practice were the same as local wheat cultivation standards, and no fungicide was applied.

2.2 Macroconidia preparation and *F. graminearum* inoculation

A highly virulent and 15-acetyldeoxynivalenol (15ADON) producing isolate of *F. graminearum*, F0609, supplied by the Jiangsu Academy of Agricultural Science (Nanjing, Jiangsu Province, China), was used to infect the wheat plants. To prepare the macroconidia, isolate F0609 was first cultured on potato dextrose agar (PDA) medium at room temperature for 10 days under fluorescent-UV lights. A small piece of mycelium was then transferred to a mung bean medium and shaken at 180 rpm at 28°C for 4 days. The mycelia were filtered with double-layered medical gauze, and the macroconidia concentration was measured by a blood cell counting plate under a microscope. The macroconidia suspension was then diluted to a concentration of 1×10^5 spores/mL.

At mid-anthesis, 10 μL of the *F. graminearum* macroconidial spore suspension (1×10^5 spores/mL) was point-inoculated into the two basal florets of fully developed spikelet between the lemma and palea using a micropipette (Hu et al., 2019). Each wheat genotype was inoculated with three biological replicates, each consisting of 18 spikes. The inoculated spikes were then covered with a transparent plastic bag to maintain humidity and sprayed with sterile water twice daily until harvest. Spikelets were harvested at 0 (sterile water mock inoculation control), 3, and 5 dpi, with three biological replicates for each time point. Those samples collected from CS and WY at 0, 3 and 5 dpi were designated as CS-0, CS-3, CS-5, WY-0, WY-3, and WY-5, respectively. A total of nine samples were collected for each genotype.

2.3 RNA extraction, library construction and sequencing

Total RNA of each sample was extracted with the TRIzol reagent (Invitrogen, CA, USA), and the residual DNA was removed using RNase-free DNase I (Invitrogen, CA, USA) following the manufacturer's instructions. The integrity and quality of the RNA were preliminarily assessed with 1.5% formaldehyde denaturing agarose gel electrophoresis. The concentration and purity of RNA were measured by a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, MA, USA). The RNA integrity was further determined by using the RNA 6000 Nano Assay Kit on the Agilent Bioanalyzer 2100 System (Agilent, CA, USA).

For cDNA library construction, 1.5 μg of RNA per sample was used for rRNA removal with the Ribo-Zero rRNA Removal Kit (Epicentre, Madison, WI, USA). Sequencing libraries were then prepared with the NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, MA, USA) following the manufacturer's guidelines, with index codes included for sample identification. For small RNA (sRNA) library construction, 2.5 μg of RNA per sample was used. Sequencing libraries were generated with the NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB, MA, USA) based on the manufacturer's instructions, and index codes were also added for sample identification (Zhang et al., 2020). The quality of all libraries was assessed using the Agilent Bioanalyzer 2100. Paired-end sequencing (PE150-bp) was performed on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) by Biomarker Technologies (Qingdao, China).

2.4 Data processing

The raw data (raw reads) were filtered with in-house Perl scripts to remove low-quality reads and sequences containing adapters or poly-N reads. Clean data (clean reads) were collected, and analyzed for the Q20, Q30, GC content, and sequence duplication levels. For sRNA-seq data, sequences ranging from 15 to 35 nucleotides (nt) in length were retained after trimming. All downstream analyses were conducted with high-quality clean data. The clean reads were then mapped to the CS reference genome IWGSC_RefSeq_v2.1 (IWGSC; Alaux et al., 2023) using HISAT2 software version 2.2.1 (Kim et al., 2019). Only reads with a perfect match or a single mismatch were further analyzed and annotated based on the reference genome.

2.5 Identification of mRNA, lncRNA, circRNA, and miRNA

For mRNA and lncRNA identification, the transcriptome was assembled using StringTie (v2.2.0) (Kovaka et al., 2019) based on the reads mapped to the wheat CS reference genome (IWGSC_RefSeq_v2.1). The GffCompare program (v0.12.6)

(Pertea and Pertea, 2020) was used to annotate the assembled transcripts. Unknown transcripts were screened as potential lncRNAs. Four computational tools—CPC, CNCI, Pfam, and CPAT—were combined to distinguish non-protein-coding RNA (ncRNA) candidates from the unknown transcripts. Putative protein-coding RNAs were filtered out using thresholds for minimum transcript length and exon number. LncRNA candidates were selected based on transcript lengths greater than 200 nucleotides (nt) and more than 2 exons, and further validated with the four computational tools mentioned above. Those tools effectively differentiate protein-coding genes from non-coding ones and classify various lncRNA types, such as lincRNAs, intronic lncRNAs, antisense, and sense lncRNAs (Zhang et al., 2020).

The circRNA identification tool "CIRI" (CircRNA Identifier) (Gao et al., 2015) was used to detect circRNAs from the transcriptome data. To collect sufficient information for circRNA identification and characterization, the SAM files were scanned twice with CIRI. The identified circRNAs were then output with annotation information. Target miRNAs of the circRNAs and the target genes of miRNAs were predicted using miRanda (Doron et al., 2008) and RNAhybrid for animals (Rehmsmeier et al., 2004), and TargetFinder for plants (Bo and Wang, 2005). During the predictions, FASTA sequences of these circRNAs and miRNAs were used as input files.

The clean reads were aligned against four bioinformatic databases, Silva, GtRNAdb, Rfam, and Repbase with Bowtie software (Langmead and Salzberg, 2012) to identify candidate miRNA. The alignments allowed for length variations at both the 5′ and 3′ ends, as well as one internal mismatch. Common RNA families, including rRNA, tRNA, snRNA, snoRNA, and other ncRNAs, along with repeats, were filtered out. The remaining reads were then subject to miRNA identification through comparison with the wheat reference genome. The known and novel miRNAs were detected with the miRBase v22 database (Kozomara et al., 2019) and miRDeep2 module (Friedländer et al., 2012), respectively.

2.6 Differential expression analysis

The expression levels of coding genes and lncRNAs in each sample were estimated with fragments per kilobase of transcript per million fragments mapped (FPKM) values (Trapnell et al., 2010). The expression of circRNAs was determined based on the number of junction reads identified by the CIRI tool. For miRNAs, expression levels of each sample were estimated through the following steps: (1) sRNAs were mapped back to the precursor sequence, and (2) the read count for each miRNA was obtained from the mapping results. The expression levels of miRNAs and circRNAs in each sample were calculated using transcripts per million (TPM) (Fahlgren et al., 2007). Differential expression analysis among three time points (0, 3, and 5 dpi) of two genotypes (WY and CS) was conducted with DESeq2 R package (v1.10.1) (Love et al., 2014). Differentially expressed genes (DEGs) were identified using an adjusted *p*-value (calculated using the Benjamini and Hochberg method) ≤ 0.01 to control the false discovery rate (FDR) and a $|\log_2$ (Fold change (FC))| > 1. Venny 2.1 (Oliveros, 2015) was then used for Venn diagram analysis, and SRplot web server (Tang et al., 2023; available at http://www.bioinformatics.com.cn/SRplot) was used for heatmap drawing.

2.7 Analysis of competing endogenous RNA

Candidate ceRNA relationship pairs were identified based on miRNA targeting relationships. Using the predicted miRNA-mRNA, lncRNA-miRNA, and circRNA-miRNA interaction pairs, groups of lncRNA-miRNA-mRNA or circRNA-miRNA-mRNA that shared the same miRNAs were collected. The screening criteria for ceRNAs were as follows: (1) The number of shared miRNAs among ceRNAs must be greater than or equal to 5; (2) The hypergeometric test *p*-value must be < 0.01, and the corrected FDR value must be < 0.01; (3) mRNAs and lncRNA/circRNA interaction pairs with a co-expression correlation (r > 0.9) were further selected for ceRNA network construction.

2.8 Gene functional annotation and GO and KEGG enrichment analysis

Gene function annotation was performed using multiple databases, including Nr (Non-redundant protein), Nt (Non-redundant nucleotide), Pfam (Protein family), KOG/COG (Clusters of Orthologous Groups of proteins), Swiss-Prot (a manually annotated and non-redundant protein sequence database), KEGG, and GO. GO enrichment analysis of DEGs was conducted with the GOseq R package, based on Wallenius' non-central hypergeometric distribution (Young et al., 2010), which adjusts for gene length bias in DEGs. The enrichment analysis of DEGs in KEGG pathways was performed using KOBAS software (version 2.0) (Xie et al., 2011).

2.9 Real-time quantitative PCR validation

For RT-qPCR analysis, cDNA of all RNA samples (those used for Illumina RNA-sequencing) was synthesized with the PrimeScript RT reagent Kit (TaKaRa, Dalian, China) according to the manufacturer's protocol. The reaction volume was 20 µL including random hexamers and 1 µg of each RNA sample. Two wheat genes—glyceraldehyde-3-phosphate dehydrogenase (GAPDH, TraesCS7A02G313100) and indole-3 acetaldehyde oxidase (IAAOx, TraesCS2A02G246300) (Hu et al., 2019)—were used as reference genes to normalize gene expression in the two wheat genotypes. Primers (Dataset S1) were designed using the PrimerQuest™ Tool from Integrated DNA Technologies (IDT) and synthesized by Sangon Biotech (Shanghai, China). For each reaction, 1 µL of 10×-diluted cDNA was added to a 10 µL reaction volume using the SYBR Green PCR kit (TaKaRa, Dalian, China). RT-qPCR was carried out in an ABI7500 Real-Time PCR System (Applied Biosystems, CA, USA). The cycling conditions were 5 minutes at 95°C, followed by 40 cycles of 30 seconds at 95°C,

30 seconds at the melting temperature (as indicated in Dataset S1), and 30 seconds at 72°C. The melting curve analysis was conducted from 55°C to 95°C, with readings taken every 1°C and held for 5 seconds. Four technical replicates were performed for each sample. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

3 Results

3.1 Sequencing and identification of RNAs

After sequencing of the 18 samples, the raw data were filtered to obtain clean data, resulting in a total of 1,035,915,752 reads and 309.95 Gb of clean data. Each sample yielded approximately 15.95 Gb of clean data, with Q20 and Q30 base percentages exceeding 97.02% and 92.05%, respectively. The GC content ranged from 47.39% to 50.51% (Dataset S2). These metrics indicate that the quality of RNA sequencing in this study was high and suitable for further analysis. The clean reads were mapped to the reference wheat genome sequences (IWGSC_RefSeq_v2.1) of Chinese Spring (Alaux et al., 2023). After identification of RNAs, a total of 15,765 lncRNAs, 286,397 genes, 621 circRNAs, and 3,063 miRNAs were detected.

3.2 Differential expression analysis of mRNA, lncRNA, circRNA, and miRNA

To identify differentially expressed RNAs, pairwise comparisons were conducted between time points for the two wheat genotypes. In CS, the numbers of differentially expressed mRNAs, lncRNAs, circRNAs, and miRNAs for 0 *vs.* 3-dpi, 0 *vs.* 5-dpi, and 3 *vs.* 5-dpi, were 5,713, 174, 8, and 29; 20,596, 331, 7, and 172; and 16,028, 229, 10, and 69, respectively. In WY, the numbers of differentially expressed mRNAs, lncRNAs, circRNAs, and miRNAs for 0 *vs.* 3-dpi, 0 *vs.* 5-dpi, and 3 *vs.* 5-dpi were 2,196, 150, 5, and 30; 9,793, 288, 11, and 179; and 6,769, 188, 5, and 239, respectively (Figure 1). The expression profiles of those differentially expressed RNAs were visualized using circos maps, with the height of the profiles representing significance ($-\log_{10}$ (FDR)). The results showed that the *F. graminearum* invasion had a significant impact on the expression of mRNA and miRNA, and had moderate to low effects on the expression of lncRNAs and circRNAs, respectively (Figure 2) (Supplementary Figure 1).

Analysis of the differentially expressed RNAs showed that the numbers of differentially expressed genes in CS were 2.60, 2.10, and 2.37 times greater than those in WY for the comparison time points of 0 *vs.* 3-dpi, 0 *vs.* 5-dpi, and 3 *vs.* 5-dpi, respectively. The

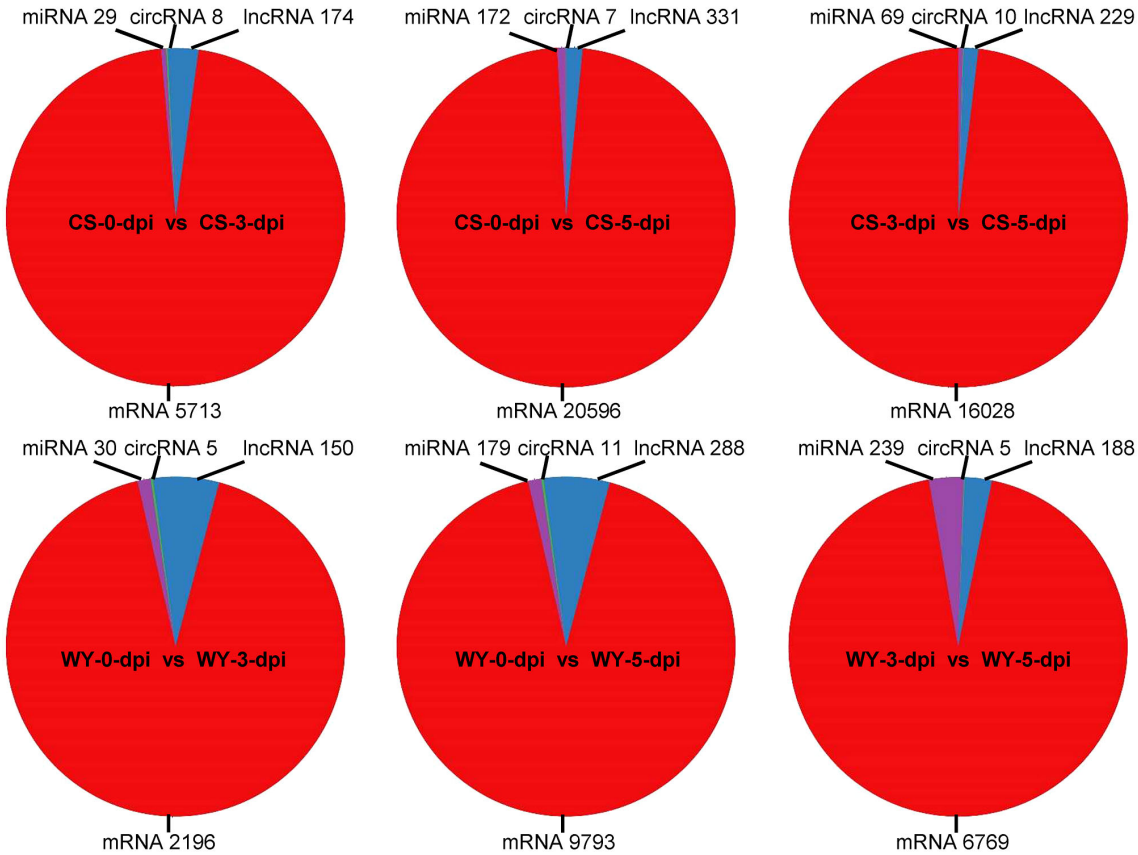


FIGURE 1
Statistics of differentially expressed RNAs across various comparisons. The 0, 3, and 5-dpi represent samples collected at 0, 3, and 5 days post *F. graminearum* inoculation (dpi), respectively. CS and WY refer to Chinese Spring and Wuyangmai, respectively. LncRNA, long non-coding RNA; miRNA, microRNA; circRNA, circular RNA.

10.3389/fpls.2025.1537605

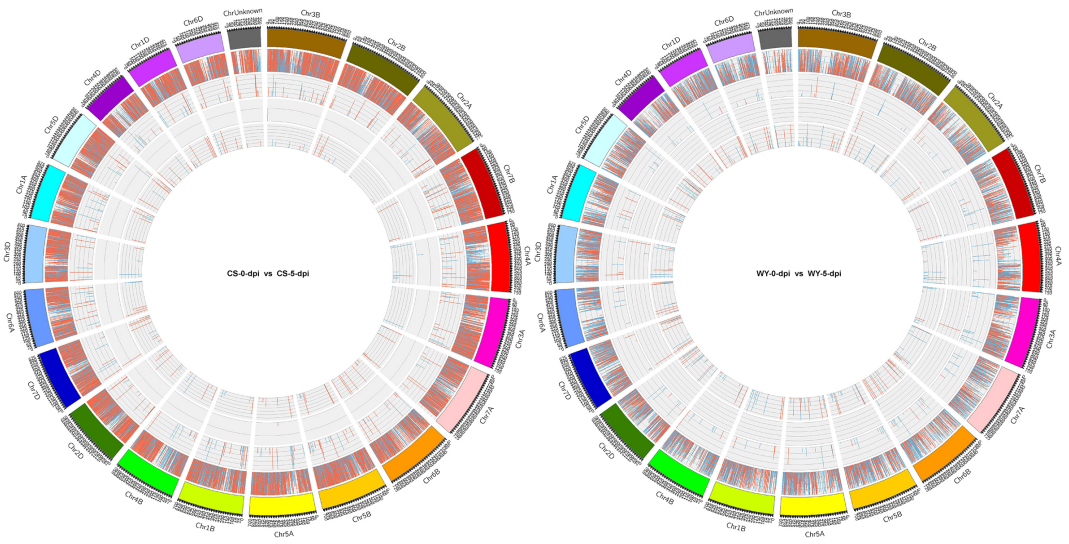


FIGURE 2
Expression profiles of differentially expressed RNAs at 5 dpi with *F. graminearum*. The outermost ring shows chromosome information, followed by mRNA (gene), lncRNA, circRNA, and miRNA. For each group of differentially expressed RNAs, red, blue, and the height indicate up-regulation, down-regulation, and significance [$-\log_{10}$ (FDR)], respectively.

differentially expressed lncRNAs in CS were slightly higher than those in WY, with fold changes of 1.16, 1.15, and 1.22 for the above-mentioned time points. However, the numbers of differentially expressed circRNAs were very low in both CS and WY, ranging from 5 to 11. Interestingly, the number of differentially expressed miRNAs in WY was slightly higher than that in CS for the comparison time points of 0 *vs.* 3 dpi and 0 *vs.* 5 dpi, and 3.46 times higher for the comparison time points of 3 *vs.* 5 dpi (Table 1).

3.3 Integrated analysis of differentially expressed RNA targeting relationships

Differentially expressed lncRNAs, circRNAs, miRNAs, and mRNAs were analyzed by taking each RNA type as the center and examining the targeted relationships involving other differentially expressed RNAs (including host genes corresponding to circRNAs). First, an integrated analysis was conducted using differentially expressed lncRNAs (DE_lncRNAs) as the center. In the comparisons of 0 *vs.* 3-dpi, 0 *vs.* 5-dpi, and 3 *vs.* 5-dpi of CS, the Venn diagram revealed that 0 (1), 11 (14), and 2 (4) lncRNAs, respectively, were located at the intersection of the three data sets: DE_lncRNA, DE_Cis.mRNA_Target lncRNA (or DE_Trans.mRNA_Target lncRNA), and DE_miRNA_Target lncRNA. Similarly, for the pairwise comparisons of the three time points in WY, the Venn diagram indicated that 0 (0), 1 (7), and 1 (4) lncRNAs were located at the intersection of the same three data sets, respectively (Supplementary Figures 2A, B). When circRNAs were taken as the center for the targeting relationship analysis, no circRNAs were found at the intersection of the three data sets: DE_circRNA, DE_Hostgene_circRNA, and DE_miRNA_Target circRNA, for any of the pairwise comparisons among the three time points in CS or WY (Supplementary Figure 3).

Subsequently, an integrated analysis of targeting relationships was conducted using differentially expressed miRNAs (DE_miRNA) as the center. In the pairwise comparisons of the three time points in CS, the Venn diagram indicated that 0 (3), 4 (9), and 1 (4) miRNA were present at the intersection of the three data sets: DE_miRNA, DE_circRNA_Target miRNA (or DE_lncRNA_Target miRNA), and DE_mRNA_Target miRNA, respectively. Similarly, for the pairwise comparisons of the three time points in WY, the Venn diagram showed that 0 (0), 1 (28), and 0 (28) miRNAs were located at the intersections of the same data sets, respectively (Figure 3). Additionally, the analysis revealed that the primary

TABLE 1 Comparative analysis of differentially expressed RNAs between Chinese Spring (CS) and Wuyangmai (WY).

RNA	0 *vs.* 3-dpi			0 *vs.* 5-dpi			3 *vs.* 5-dpi		
	CS	WY	CS/WY	CS	WY	CS/WY	CS	WY	CS/WY
mRNA	5713	2196	2.60	20596	9793	2.10	16028	6769	2.37
lncRNA	174	150	1.16	331	288	1.15	229	188	1.22
circRNA	8	5	1.6	7	11	0.66	10	5	2.00
miRNA	29	30	0.97	172	179	0.96	69	239	0.29

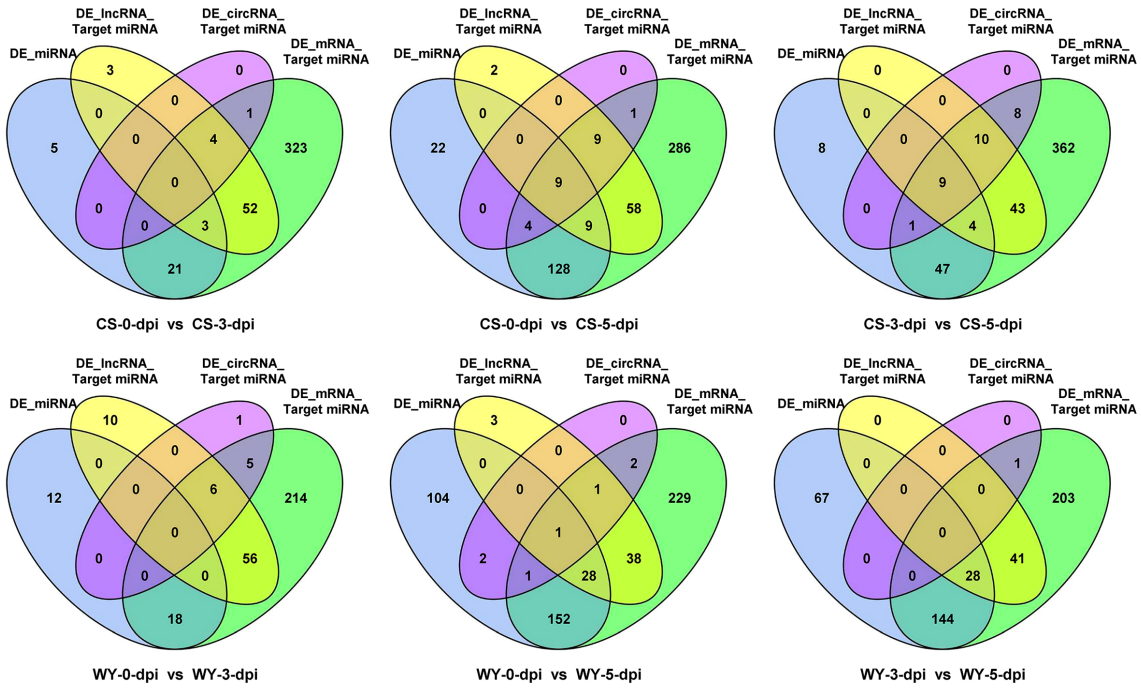


FIGURE 3
Interaction of differentially expressed miRNAs with all miRNAs targeted by differentially expressed mRNAs, lncRNAs, and circRNAs. Here, DE_miRNA represents differentially expressed miRNAs; DE_mRNA_Target miRNA, DE_lncRNA_Target miRNA, and DE_circRNA_Target miRNA represent miRNAs targeted by differentially expressed genes, lncRNAs, and circRNAs, respectively.

targeting relationships existed between the DE_miRNA and DE_mRNA_Target miRNA data sets. Specifically, 24 (18), 150 (182), and 61 (172) miRNAs were identified at their intersection in the pairwise comparisons of the three CS (or WY) time points, respectively (Figure 3).

Finally, differentially expressed mRNAs (DE_mRNA) were used as the center for analyzing targeting relationships. The Venn diagram showed that the primary targeting relationships occurred between the DE_mRNA data set and either DE_lncRNA_Target mRNA or DE_miRNA_Target mRNA. In the pairwise comparisons of the three CS time points, 39 (1,241), 283 (7,027), and 166 (5,608) mRNAs were detected at the intersection of DE_mRNA with DE_lncRNA_Target Cis.mRNA (or DE_lncRNA_Target Trans.mRNA), respectively. Similarly, for the pairwise comparisons of the three WY time points, 12 (167), 42 (1,551), and 36 (1,358) mRNAs were found at the corresponding intersections of the above two data sets (Supplementary Figures 4A, B). When considering the targeting relationships between DE_mRNA and DE_miRNA_Target mRNA, 72 (23), 2,152 (1,239), and 1,179 (783) mRNAs were identified at their intersections in the pairwise comparisons of the three CS (or WY) time points, respectively (Supplementary Figures 4A, B).

3.4 Identification of ceRNA

mRNAs, lncRNAs, and circRNAs can function as competing endogenous RNAs (ceRNAs) in a ceRNA network, where they regulate expression levels of each other by competitively binding to the same miRNA response elements (MREs) (Salmena et al., 2011). Based on the screening criteria for ceRNAs, co-expression analysis identified 98 circRNA-miRNA-mRNA relationship pairs and 49 lncRNA-miRNA-mRNA pairs (Dataset S3). However, no differentially expressed ceRNA relationship pairs were identified using a one-step nearest-neighbor network analysis, when each group of differentially expressed RNAs was extracted from the ceRNA relationship pairs.

3.5 Differential expression analysis for up-regulated genes

Differential gene expression analysis revealed that in CS, there were 3,989, 13,141, and 10,521 up-regulated DEGs, and 1,724, 7,455, and 5,507 down-regulated DEGs for 0 *vs.* 3-dpi, 0 *vs.* 5-dpi, and 3 *vs.* 5-dpi comparisons, respectively. The number of up-regulated DEGs was consistently much greater than that of down-regulated DEGs across all three comparisons. In WY, there were 1,060, 5,099, and 4,608 up-regulated DEGs, and 1,136, 4,694, and 2,161 down-regulated DEGs for the same comparisons. The number of up-regulated DEGs was higher than down-regulated DEGs only in the 3 *vs.* 5-dpi comparison, while for the 0 *vs.* 3-dpi and 0 *vs.* 5-dpi comparisons, the up-regulated DEGs were slightly fewer or greater than the down-regulated ones, respectively (Figure 4A). Additionally, the expression levels and significance of up-regulated DEGs in CS were consistently much higher than those in WY across all three comparisons (Figure 4B) (Supplementary Figures 5A, 6A).

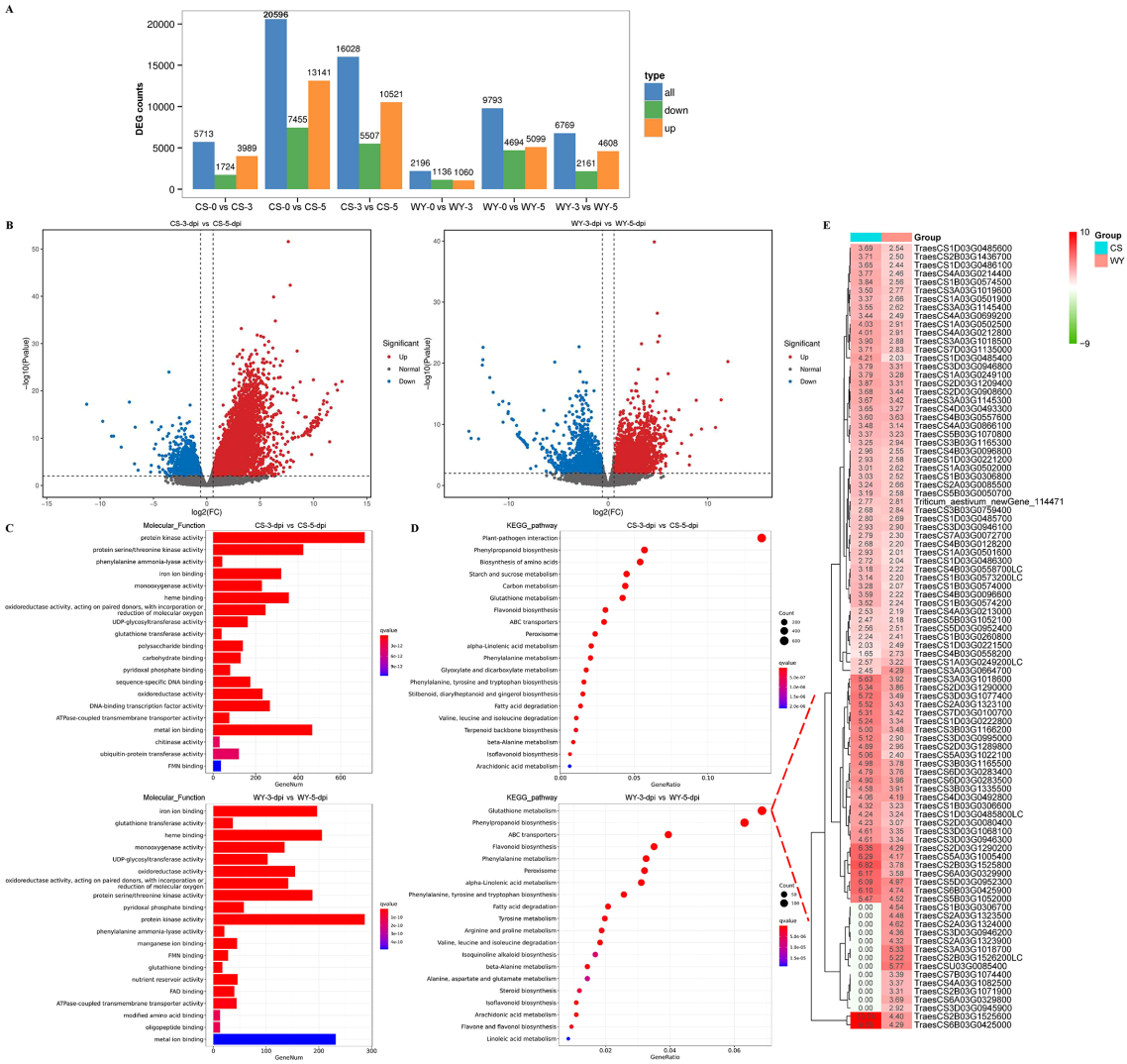


FIGURE 4
Differential expression analysis for up-regulated genes. **(A)** Statistical analysis for all genes. **(B)** Volcano plot of DEGs obtained in the comparison of 3 *vs*. 5-dpi. **(C)** GO term and **(D)** KEGG pathway enrichment analysis for significantly up-regulated DEGs. **(E)** Gene expression heatmap of glutathione metabolism pathway.

GO enrichment analysis of up-regulated DEGs revealed distinct defense responses between the FHB-resistant wheat CS and the susceptible wheat WY. For the 0 *vs.* 3-dpi comparison, two of the top five GO terms differed between CS and WY. In CS, the top terms were "phenylalanine ammonia-lyase activity" and "heme binding", while in WY, they were "hydrolase activity, hydrolyzing O-glycosyl compounds" and "polysaccharide binding" (Supplementary Figure 5B). In the 0 *vs.* 5-dpi comparison, CS showed enrichment for "glutathione transferase activity" and "iron ion binding", while WY had "ATPase-coupled transmembrane transporter activity" and "ATPase activity" (Supplementary Figure 6B). In the 3 *vs.* 5-dpi comparison, three of the top five GO terms were different: CS showed "protein kinase activity", "protein serine/threonine kinase activity", and "phenylalanine ammonia-lyase activity", while WY showed "glutathione transferase activity", "heme binding", and "UDP-glycosyltransferase activity" (Figure 4C).

KEGG pathway enrichment of up-regulated DEGs also highlighted distinct defense responses between CS and WY. In the 0 *vs.* 3-dpi comparison, only one of the top five KEGG pathways was different: "biosynthesis of amino acids" (ko01230) in CS, compared to "plant hormone signal transduction" (ko04075) in WY, although the enrichment of the top five pathways in WY was not significant. In the 0 *vs.* 5-dpi comparison, two of the top five pathways differed: "phenylpropanoid biosynthesis" (ko00940) and "carbon metabolism" (ko02010) in CS, and "glutathione metabolism" (ko00480) and "ABC transporters" (ko02010) in WY. For the 3 *vs.* 5-dpi comparison, four of the top five pathways were different between CS and WY: "plant-pathogen interaction" (ko04626), "biosynthesis of amino acids" (ko01230), "starch and sucrose metabolism" (ko00500), and "carbon metabolism" (ko01200) in CS, compared to "glutathione metabolism" (ko00480), "ABC transporters" (ko02010), "flavonoid

biosynthesis" (ko00941), and "phenylalanine metabolism" (ko00360) in WY (Figure 4D) (Supplementary Figures 5C, 6C).

Additionally, a $\log_2$ (FC) > 2 threshold was applied to filter up-regulated genes enriched in the "glutathione metabolism" pathway in the 3 *vs.* 5-dpi comparison of WY, resulting in the identification of 93 DEGs. The expression heatmap of these up-regulated DEGs showed that 80 DEGs were shared between CS and WY. Among these, two DEGs (TraesCS2B03G1525600 and TraesCS6B03G0425000) in CS had expression levels more than two times higher than in WY. Furthermore, the heatmap also showed that 13 DEGs were differentially expressed only in WY (Figure 4E).

3.6 Integrated analysis of the targeting relationship between miRNA and mRNA

Using small RNA and transcriptome sequencing data, DE_miRNAs and DEGs were identified in the two sample groups. The regulatory relationships between miRNAs and mRNAs were then explored, focusing on the negative regulatory effects of miRNAs on mRNAs. First, DE_miRNAs were used as screening criteria to identify the mRNAs regulated by these miRNAs, followed by an analysis of the pairs of DE_miRNAs and mRNAs with negative regulatory relationships (Dataset S4). Similarly, DEGs were used as screening criteria to identify miRNAs that regulate these DEGs, and DEG-miRNA pairs with negative regulatory relationships were analyzed (Dataset S5). The targeting relationships between miRNAs and mRNAs were visualized using Cytoscape software (v3.10.2). The top 20 DE_miRNAs (or top 18 for 0 *vs.* 3-dpi comparison in WY) most relevant to target gene regulation were selected to construct the miRNA-mRNA targeting relationship network map (Dataset S6).

The results showed that only two miRNAs were shared between CS and WY in the 0 *vs.* 3-dpi comparison, whereas 16 and 15 miRNAs were shared in the 0 *vs.* 5-dpi and 3 *vs.* 5-dpi comparisons, respectively (Figure 5A) (Supplementary Figure 7A). In the 0 *vs.* 3-dpi comparison, ten DE_miRNAs in CS and eight in WY specifically targeted DEGs with negative regulatory relationships, with each DE_miRNA targeting one to four DEGs (Supplementary Figure 7B). In the 0 *vs.* 5-dpi and 3 *vs.* 5-dpi comparisons, all DE_miRNAs specifically expressed in CS and WY targeted DEGs with negative regulatory relationships, with each DE_miRNA targeting 14 to 30 DEGs. Notably, novel_miR_228, which was up-regulated in CS, and tae-miR1122a, which was down-regulated in WY, were present in both comparisons. Novel_miR_228 down-regulated 28 and 23 DEGs in CS, while tae-miR1122a up-regulated 27 and 23 DEGs in WY, respectively (Figure 5B) (Supplementary Figure 7B).

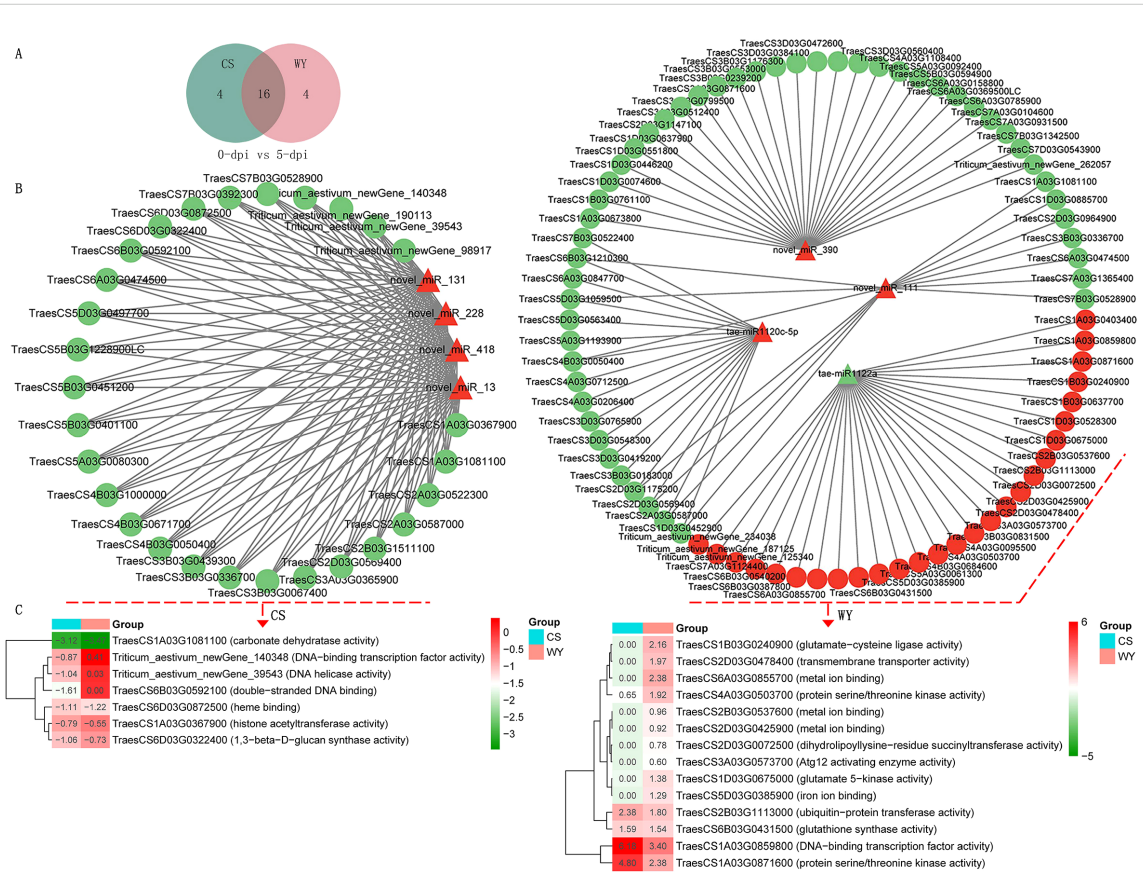


FIGURE 5
Targeting relationships between DE_miRNAs and DEGs. **(A)** Comparison of the top 20 DE_miRNAs most relevant to target gene regulation in CS and WY. **(B)** Targeting relationships between the DE_miRNAs specifically expressed in CS and WY and DEGs. **(C)** Expression heatmap of genes significantly down-regulated by novel_miR_228 and significantly up-regulated by tae-miR1122a with molecular function (MF) GO term. Heatmap represents $\log_2$ (FC) between *F. graminearum* infected and control.

GO analysis of these DEGs revealed that seven DEGs down-regulated by novel_miR_228 in CS had molecular function (MF) terms in the 0 *vs.* 5-dpi comparison, and nine DEGs in the 3 *vs.* 5-dpi comparison. Significant GO term enrichment analysis (using a q-value < 0.05 as the threshold) showed two DEGs, TraesCS1A03G0367900 (histone acetyltransferase activity; q = 1.52e-02) and TraesCS6B03G0592100 (double-stranded DNA binding; q = 1.67e-43), were significantly enriched in the 0 *vs.* 5-dpi comparison of CS, while six DEGs—TraesCS2A03G0587000 (protein kinase activity; q = 9.31e-06), TraesCS2D03G0569400 (protein kinase activity; q = 9.31e-06), TraesCS3D03G0550100 (hydrolase activity; q = 1.61e-02), TraesCS5A03G0080300 (microtubule binding; q = 1.12e-02), TraesCS7A03G0224800 (single-stranded DNA binding; q = 4.39e-04), and TraesCS7A03G0450300 (GTP binding; q = 4.47e-08)—were significantly enriched in the 3 *vs.* 5-dpi comparison of CS. The expression of these DEGs in CS was lower than in WY (Figure 5C) (Supplementary Figure 7C).

Likewise, GO analysis showed that 14 DEGs up-regulated by tae-miR1122a in WY had MF terms in both the 0 *vs.* 5-dpi and 3 *vs.* 5-dpi comparisons. Nine DEGs up-regulated by tae-miR1122a in WY were not differentially expressed in CS in the 0 *vs.* 5-dpi comparison, and three (TraesCS2B03G0537600, TraesCS2D03G0425900, and TraesCS6A03G0855700) were significantly enriched in the MF term for metal ion binding (q = 2.14e-03). In the 3 *vs.* 5-dpi comparison, nine DEGs up-regulated by tae-miR1122a in WY were not differentially expressed in CS, and four (TraesCS2B03G0537600, TraesCS2D03G0425900, TraesCS6A03G0855700, and TraesCS7D03G0982200) were significantly enriched in the MF term for metal ion binding (q = 2.94e-02). The expression of these DEGs in WY was higher than in CS (Figure 5C) (Supplementary Figure 7C).

3.7 KOG functional classification of DEGs regulated by miRNA

After analyzing the targeting relationships between miRNAs and mRNAs, functional classification analysis was performed on the DEGs identified. The KOG functional classification of DE_miRNA-regulated DEGs was conducted. Focusing on the top five functional classes, four classes—T (signal transduction mechanisms), R (general function prediction only), O (posttranslational modification, protein turnover, chaperones), and J (translation, ribosomal structure, and biogenesis)—were shared by CS and WY in the comparison of 0 *vs.* 3-dpi. The fifth class differed, with class S (function unknown) identified in CS and class A (RNA processing and modification) identified in WY. However, in the 0 *vs.* 5-dpi and 3 *vs.* 5-dpi comparisons, the top five functional classes for both CS and WY were identical: T, R, O, G (carbohydrate transport and metabolism), and K (transcription) (Supplementary Figure 8). Similarly, the KOG functional classification of miRNA-regulated DEGs was performed. The results showed that the top five functional classes—T, R, O, G, and Q (secondary metabolites biosynthesis, transport, and catabolism)—were consistent across all three comparisons for both CS and WY (Supplementary Figure 9).

3.8 GO analysis of DEGs regulated by miRNA

GO analysis of DE_miRNA-regulated DEGs revealed 22, 18, and 17 functional classifications in the biological process (BP), cellular component (CC), and molecular function (MF) categories, respectively, for both CS and WY. Among these, the percentage of genes in 16 BP, 15 CC, and 12 MF classifications exceeded 0.1%. The results showed that most DEGs in the BP category were associated with metabolic processes, cellular processes, and single-organism processes. In the CC category, the majority of DEGs were located in the membrane, cell, cell part, membrane part, and organelle. In the MF category, most DEGs were related to catalytic activity and binding functions (Supplementary Figure 10). Similar findings were obtained from the GO analysis of miRNA-regulated DEGs (Supplementary Figure 11).

GO enrichment analysis was conducted on the identified DEGs (Dataset S7, Dataset S8), with the top 20 terms for each comparison selected for display, using a q-value < 0.05 as the significance threshold. GO enrichment based on DE_miRNA-regulated DEGs revealed the most significant BP terms in CS were phenylalanyl-tRNA aminoacylation, RNA modification, and leucyl-tRNA aminoacylation for the 0 *vs.* 3-dpi, 0 *vs.* 5-dpi, and 3 *vs.* 5-dpi comparisons, respectively, whereas those in WY were phenylalanyl-tRNA aminoacylation, defense response, and mitochondrial RNA modification. The BP terms with the highest number of genes in CS included gene silencing by RNA, defense response, and protein folding, while those in WY were gene silencing by RNA, defense response, and nucleic acid phosphodiester bond hydrolysis. For the CC category, the most significant terms in CS were the THO complex (part of the transcription export complex), microtubule, and Ino80 complex, while those in WY were the ROC complex, Golgi membrane, and cell plate. The CC terms with the most genes in CS were cytoplasm (for the first two comparisons) and microtubule (for the third one), while in WY, the terms with the most genes were cytoplasm (for the first comparison) and Golgi membrane (for the subsequent two). In the MF category, the most significant terms in CS were phenylalanine-tRNA ligase activity, double-stranded DNA binding, and carbohydrate binding, while in WY they were calcium transmembrane transporter activity, phosphorylative mechanism, and double-stranded DNA binding. The terms with the most genes in CS were protein binding, catalytic activity, and protein kinase activity, whereas in WY, they were RNA binding, protein kinase activity, and RNA binding (Figure 6).

GO enrichment analysis based on miRNA-regulated DEGs revealed that the most significant BP terms in CS were protein-chromophore linkage, trehalose biosynthetic process, and leucyl-tRNA aminoacylation in the 0 *vs.* 3-dpi, 0 *vs.* 5-dpi, and 3 *vs.* 5-dpi comparisons, respectively. In WY, the most significant BP terms were protein-chromophore linkage in both 0 *vs.* 3-dpi and 0 *vs.* 5-dpi, and trehalose biosynthetic process for 3 *vs.* 5-dpi. The BP terms with the most genes in CS included defense response, carbohydrate metabolic process, and recognition of pollen, while in WY, they were defense response, photosynthesis, and response to oxidative stress. In the CC category, significant terms were observed only in two comparisons of WY (0 *vs.* 3-dpi and 0 *vs.* 5-dpi), with the most

10.3389/fpls.2025.1537605

FIGURE 6
GO term enrichment analysis for DE_miRNA-regulated DEGs. GeneNum denotes the gene number.

significant CC term being photosystem I in both cases. The CC terms with the most genes were the chloroplast thylakoid membrane for both comparisons, though this term was not statistically significant in the second comparison. In the MF category, the most significant MF term in the comparison of 0 *vs.* 3-dpi in CS was calmodulin binding, with the term having the most genes being protein kinase activity. In the other two comparisons (0 *vs.* 5-dpi and 3 *vs.* 5-dpi), the most significant MF term and the term with the most genes in CS were both protein kinase activity. For WY, the most significant MF terms in the 0 *vs.* 3-dpi and 0 *vs.* 5-dpi comparisons were chlorophyll binding and protein serine/threonine kinase activity, respectively. In both cases, the MF term with the most genes was protein kinase activity. In the 3 *vs.* 5-dpi comparison in WY, both the most significant MF term and the MF term with the most genes were protein serine/threonine kinase activity (Supplementary Figure 12).

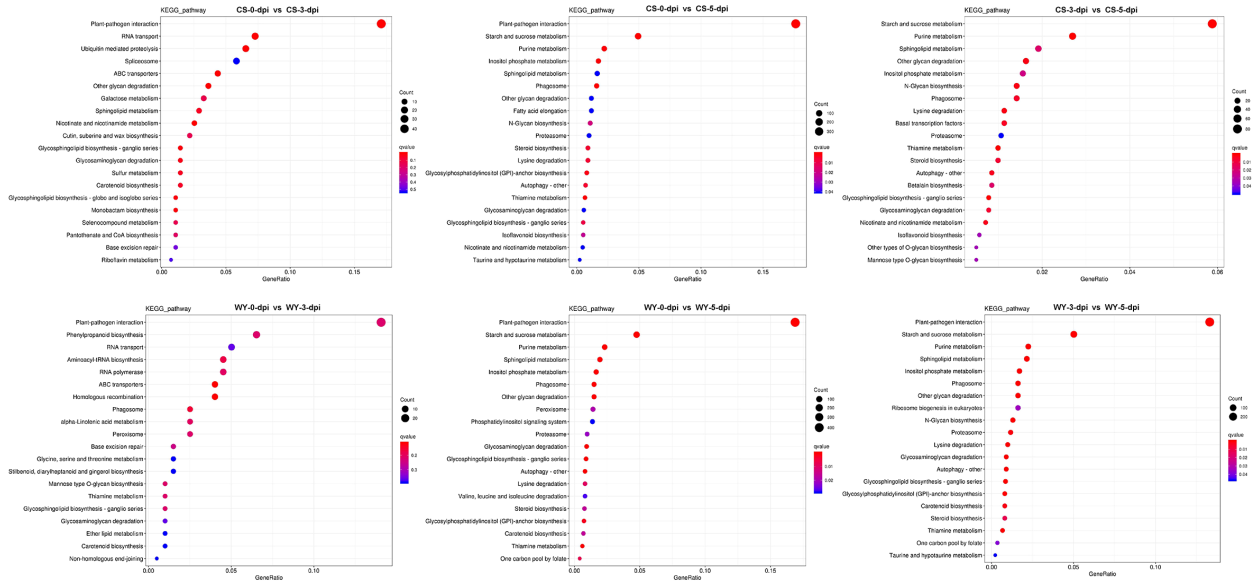


FIGURE 7
KEGG pathway enrichment analysis for DE_miRNA-regulated DEGs. GeneRatio indicates the gene ratio.

3.9 KEGG analysis of DEGs regulated by miRNA

KEGG analysis based on DE_miRNA-regulated DEGs indicated that most metabolic pathways were within the metabolism category, with the pathway containing the most DEGs being the plant-pathogen interaction pathway (Supplementary Figure 13). Similar results were obtained from the KEGG analysis based on miRNA-regulated DEGs (Supplementary Figure 14). KEGG enrichment analysis was conducted on these DEGs (Dataset S9, Dataset S10), and the top 20 pathways for each comparison were selected for display, using q-value < 0.05 as the significance threshold. KEGG enrichment of DE_miRNA-regulated DEGs showed that the most significantly enriched pathway with the highest number of DEGs was the plant-pathogen interaction pathway across all comparisons, except for 3 *vs.* 5-dpi in CS, where the starch and sucrose metabolism pathway was the most significantly enriched (Figure 7). Similarly, KEGG enrichment based on miRNA-regulated DEGs also identified the plant-pathogen interaction pathway as the most significantly enriched pathway (Supplementary Figure 15).

To compare KEGG pathway enrichment between CS and WY at different time points, the top five pathways from each comparison were selected for further analysis. First, the top five KEGG pathways enriched based on DE_miRNA-regulated DEGs were examined. Results showed that only two pathways were identical between CS and WY in the 0 *vs.* 3-dpi comparison: plant-pathogen interaction (ko04626) and RNA transport (ko03013). However, these pathways were not significant in WY. The three specific pathways in CS were ubiquitin-mediated proteolysis (ko04120), spliceosome (ko03040; not significant), and ABC transporters (ko02010). In contrast, the three specific pathways in WY were phenylpropanoid biosynthesis (ko00940), aminoacyl-tRNA biosynthesis (ko00970), and RNA polymerase (ko03020), and none of them were significant. Additionally, both CS and WY had the same top five pathways for the 0 *vs.* 5-dpi comparison. For the 3 *vs.* 5-dpi comparison, four of the top five pathways were shared by CS and WY. The only differing pathway in CS was other glycan degradation (ko00511), while in WY, it was plant-pathogen interaction (ko04626) (Table 2).

The KEGG pathways were enriched based on miRNA-regulated DEGs. In the 0 *vs.* 3-dpi comparison, four of the top five pathways were identical between CS and WY. However, the carbon metabolism pathway (ko02010) in CS and the starch and sucrose metabolism pathway (ko00500) in WY were not significant. The only differing pathway in CS was phenylpropanoid biosynthesis (ko00940), while in WY, it was photosynthesis (ko00195). For the other two comparisons (0 *vs.* 5-dpi, and 3 *vs.* 5-dpi), two of the top five pathways were different between CS and WY in each case, although the other three pathways—plant-pathogen interaction (ko04626), starch and sucrose metabolism (ko00500), and ABC transporters (ko02010)—were identical. In the 0 *vs.* 5-dpi comparison, the two different pathways in CS were flavonoid biosynthesis (ko00941) and glyoxylate and dicarboxylate metabolism (ko00630). In WY, the different pathways were carbon metabolism (ko01200) and photosynthesis (ko00195; not significant). In the 3 *vs.* 5-dpi comparison, the two different pathways in CS were flavonoid biosynthesis (ko00941) and cyanoamino acid metabolism (ko00460). In WY, the different pathways were phenylpropanoid biosynthesis (ko00940; not significant) and phenylalanine metabolism (ko00360) (Table 3).

3.10 RT–qPCR validation of DEGs regulated by miRNA

Based on the integrated analysis of the targeting relationships between miRNA and mRNA, two DE_miRNAs—tae-miR1122a and novel_miR_228—were identified from the top 20 DE_miRNAs and were specifically expressed in WY and CS at 5-dpi. Six DEGs, three with a negative regulatory relationship with tae-miR1122a and the

10.3389/fpls.2025.1537605

TABLE 2 KEGG pathway enrichment analysis of DE_miRNA-regulated DEGs in CS and WY (top 5 pathways).

Comparison	Pathway ID	Description	*q*-value	Gene	DEG	Total DEG	DEG ratio
CS 0 *vs.* 3-dpi	ko04626	Plant-pathogen interaction	0.00375	6392	47	275	17.09%
	ko03013	RNA transport	0.00511	1979	20		7.27%
	ko04120	Ubiquitin mediated proteolysis	0.00511	1659	18		6.55%
	ko03040	Spliceosome	0.54038	2797	16		5.82%
	ko02010	ABC transporters	0.00375	776	12		4.36%
WY 0 *vs.* 3-dpi	ko04626	Plant-pathogen interaction	0.21824	6392	28	199	14.07%
	ko00940	Phenylpropanoid biosynthesis	0.21824	2419	13		6.53%
	ko03013	RNA transport	0.36091	1979	10		5.03%
	ko00970	Aminoacyl-tRNA biosynthesis	0.17576	1175	9		4.52%
	ko03020	RNA polymerase	0.20704	1267	9		4.52%
CS 0 *vs.* 5-dpi	ko04626	Plant-pathogen interaction	1.04e-25	6392	375	2127	17.63%
	ko00500	Starch and sucrose metabolism	2.21e-05	1910	105		4.94%
	ko00230	Purine metabolism	0.00073	754	47		2.21%
	ko00562	Inositol phosphate metabolism	0.00046	526	37		1.74%
	ko00600	Sphingolipid metabolism	0.04071	676	35		1.65%
WY 0 *vs.* 5-dpi	ko04626	Plant-pathogen interaction	1.17e-24	6392	415	2458	16.88%
	ko00500	Starch and sucrose metabolism	1.72e-05	1910	117		4.76%
	ko00230	Purine metabolism	2.64e-05	754	57		2.32%
	ko00600	Sphingolipid metabolism	0.00045	676	48		1.95%
	ko00562	Inositol phosphate metabolism	0.00022	526	41		1.67%
CS 3 *vs.* 5-dpi	ko00500	Starch and sucrose metabolism	7.38e-07	1910	83	1411	5.88%
	ko00230	Purine metabolism	0.00011	754	38		2.69%
	ko00600	Sphingolipid metabolism	0.01862	676	27		1.91%
	ko00511	Other glycan degradation	0.00232	440	20		1.63%
	ko00562	Inositol phosphate metabolism	0.02406	526	22		1.56%
WY 3 *vs.* 5-dpi	ko04626	Plant-pathogen interaction	7.72e-06	6392	297	2232	13.31%
	ko00500	Starch and sucrose metabolism	5.67e-06	1910	112		5.02%
	ko00230	Purine metabolism	0.00018	754	50		2.24%
	ko00600	Sphingolipid metabolism	5.44e-05	676	48		2.15%
	ko00562	Inositol phosphate metabolism	0.00024	526	38		1.70%

other three with novel_miR_228, were selected for RT-qPCR validation of their expression levels. The DEGs targeted by tae-miR1122a were TraesCS1B03G0240900, TraesCS2D03G0425900, and TraesCS3A03G0573700, associated with the KEGG pathway of glutathione metabolism (ko00480), phagosome pathway (ko04145), and autophagy-other (ko04136), respectively. The DEGs targeted by novel_miR_228 were TraesCS2A03G0587000, TraesCS2D03G0569400, and TraesCS4B03G1000000, and both of the first two were associated with the plant-pathogen interaction pathway (ko04626), and the third one had no KEGG pathway annotation. The relative expression data from RT-qPCR for the selected genes were compared with the RNA-seq analysis data, and the results revealed that the expression patterns were largely consistent between the two methods, confirming that the RNA-seq analysis in this study was reliable and suitable for further research (Figure 8).

4 Discussion

4.1 Transcriptomic study for FHB resistance

Fusarium head blight (FHB) significantly threatens wheat production globally, leading to yield loss and grain contamination by

10.3389/fpls.2025.1537605

TABLE 3 KEGG pathway enrichment analysis of miRNA-regulated DEGs in CS and WY (top 5 pathways).

Comparison	Pathway ID	Description	*q*-value	Gene	DEG	Total DEG	DEG ratio
CS 0 *vs.* 3-dpi	ko04626	Plant-pathogen interaction	8.29e-10	6392	108	550	19.64%
	ko00500	Starch and sucrose metabolism	9.02e-08	1910	45		8.18%
	ko00940	Phenylpropanoid biosynthesis	0.01981	2419	35		6.36%
	ko02010	Carbon metabolism	0.06503	1853	26		4.73%
	ko00941	Flavonoid biosynthesis	0.00015	763	18		3.27%
WY 0 *vs.* 3-dpi	ko04626	Plant-pathogen interaction	6.48e-10	6392	57	218	26.15%
	ko00941	Flavonoid biosynthesis	3.80e-07	763	16		7.34%
	ko00195	Photosynthesis	0.01866	1642	14		6.42%
	ko01200	Carbon metabolism	0.04673	1853	14		6.42%
	ko00500	Starch and sucrose metabolism	0.05285	1910	14		6.42%
CS 0 *vs.* 5-dpi	ko04626	Plant-pathogen interaction	1.49e-12	6392	270	1705	15.84%
	ko00500	Starch and sucrose metabolism	1.10e-15	1910	120		7.04%
	ko02010	ABC transporters	0.00101	776	40		2.35%
	ko00941	Flavonoid biosynthesis	0.00746	763	36		2.11%
	ko00630	Glyoxylate and dicarboxylate metabolism	0.00054	546	33		1.94%
WY 0 *vs.* 5-dpi	ko04626	Plant-pathogen interaction	3.65e-08	6392	150	900	16.67%
	ko00500	Starch and sucrose metabolism	3.65e-08	1910	63		7.00%
	ko01200	Carbon metabolism	0.00021	1853	50		5.56%
	ko00195	Photosynthesis	0.05300	1642	35		3.89%
	ko02010	ABC transporters	0.00015	776	28		3.11%
CS 3 *vs.* 5-dpi	ko04626	Plant-pathogen interaction	1.04e-11	6392	226	1378	16.40%
	ko00500	Starch and sucrose metabolism	2.56e-08	1910	85		6.17%
	ko02010	ABC transporters	7.54e-06	776	41		2.98%
	ko00941	Flavonoid biosynthesis	0.00166	763	33		2.39%
	ko00460	Cyanoamino acid metabolism	0.01322	700	28		2.03%
WY 3 *vs.* 5-dpi	ko04626	Plant-pathogen interaction	0.00223	6392	88	593	14.84%
	ko00500	Starch and sucrose metabolism	0.00032	1910	38		6.41%
	ko00940	Phenylpropanoid biosynthesis	0.07981	2419	34		5.73%
	ko00360	Phenylalanine metabolism	3.92e-05	635	21		3.54%
	ko02010	ABC transporters	0.00034	776	21		3.54%

mycotoxins. Breeding and utilizing FHB-resistant wheat cultivars provide effective and eco-friendly solutions to mitigate these effects. A comprehensive understanding of the molecular mechanisms underlying wheat-pathogen interactions has laid the genetic foundation for wheat breeding programs aimed at combating FHB. Recent transcriptomic studies advanced our knowledge of key molecular pathways associated with FHB resistance and susceptibility (Kazan and Gardiner, 2018), and will aid in identifying key genes and mechanisms involved in the plant's defense response. For example, previous studies have shown that hormone biosynthesis and signal transduction pathways actively respond to FHB infection, and contribute to plant defense reaction (Biselli et al., 2018). Specifically, salicylic acid (SA) and jasmonic acid (JA) positively affect FHB resistance, while auxin and abscisic acid (ABA) are linked to susceptibility of FHB. Ethylene has dual roles in the interaction with *F. graminearum* (Wang et al., 2018). Additionally, resistant genotypes exhibit early and intense expression of defense-related genes, including those involved in redox homeostasis and secondary metabolite biosynthesis (Xiao et al., 2013). Differentially expressed miRNAs and lncRNAs play roles in regulating gene expression related to biotic and abiotic stress responses, respectively (Biselli et al., 2018; Soresi et al., 2023). Here, a comprehensive transcriptomic analysis of two Sichuan

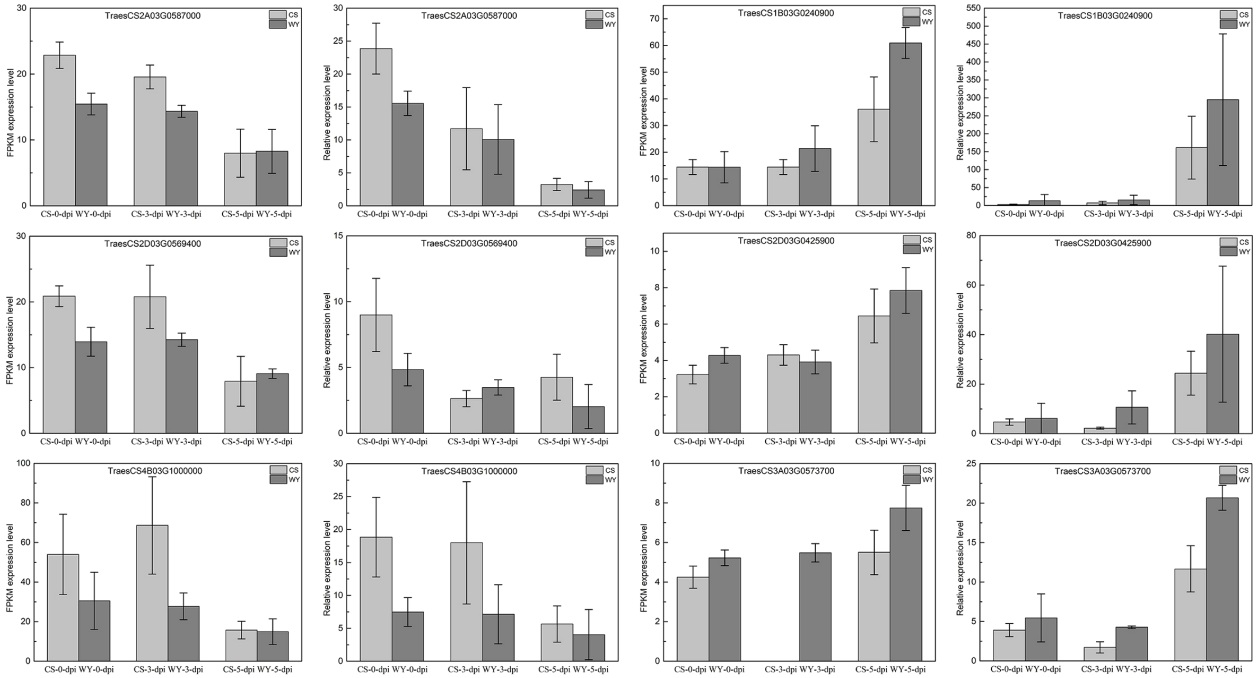


FIGURE 8
Relative expression comparison for selected genes using RNA-seq and RT-qPCR data.

wheat landraces with contrasting FHB resistance (WY was resistant, while CS was susceptible) was conducted based on the prior research (Wan et al., 1997). Analysis of the expression and regulatory networks of lncRNAs, circRNAs, mRNAs, and miRNAs provides a detailed view of the regulatory differences between resistant and susceptible genotypes.

4.2 Differential expression of lncRNAs, circRNAs, and mRNAs

Our findings reveal significant differences in RNA expression profiles between WY and CS at different stages post-inoculation (Figures 1, 2). The differential expression of lncRNAs is particularly noteworthy. Previous studies have indicated that lncRNAs modulate the expression of defense-related genes, such as those involved in the JA pathway, which is crucial for plant disease resistance (Chen et al., 2023). Specific lncRNAs, like ALEX1 in rice, enhance resistance to bacterial blight by activating JA signaling and increasing JA content (Yu et al., 2019). Furthermore, lncRNAs can function as ceRNAs, decoying miRNAs to regulate the expression of target genes involved in immune responses. For instance, lncRNAs in tomato modulate MYB transcription factors by decoying miR159, enhancing resistance to *Phytophthora infestans* (Cui et al., 2020). LncRNAs have also been shown to regulate pathogen resistance and often peak early during pathogen invasion (Duan et al., 2020). Our results confirm this, with WY and CS displaying substantial lncRNA activity at 3 and 5-dpi, suggesting that early lncRNA activity is crucial for an effective defense against *F. graminearum*. Moreover, the greater number of differentially expressed lncRNAs observed in the field conditions compared to greenhouse settings suggests that environmental factors significantly influence lncRNA-mediated responses (Soresi et al., 2023).

In addition to lncRNAs, which are known for their regulatory roles, the circRNAs have emerged as another crucial class of regulatory RNAs. The role of circRNAs in responding to FHB is still not well-defined. Our data showed limited differential expression of circRNAs in both genotypes (Figures 1, 2), aligning with findings that indicate a reduction in circRNA expression following *F. graminearum* infection, possibly due to their involvement in fine-tuning gene expression during initial infection stages (Yin et al., 2022). Together, these results suggest that lncRNAs actively mediate gene regulation during early infection, while circRNAs may play a more subdued role, potentially involved in more sustained immune responses.

The differentially expressed mRNAs was more pronounced in CS than in WY (Table 1), consistent with the previous studies that susceptible genotypes often exhibit a broader transcriptional response. This reflects a less-targeted response, activating a wide array of genes rather than focusing on specific defense-related pathways (Erayman et al., 2015). In contrast, resistant genotypes like WY likely mount a more efficient and focused response, resulting in fewer DEGs but potentially more effective pathogen resistance mechanisms (Walker et al., 2024).

4.3 Integrated analysis of targeting relationships among RNAs

To further understand the regulatory roles of these RNAs, we analyzed their interactions and targeting relationships. Focusing on miRNAs as central regulators, we observed primary targeting

relationships with mRNAs (Figure 3), particularly at 3 and 5 dpi. This miRNA-mRNA interaction network suggests that miRNAs play a vital role in silencing pathogen-related genes, curbing the spread of infection and enhancing plant resistance (Jiao and Peng, 2018; MaChado et al., 2018; Fu et al., 2023). In contrast, fewer interactions were observed when lncRNAs or circRNAs were analyzed as the central regulators (Supplementary Figures 2, Supplementary Figures 3). This lack of significant ceRNA relationships at later infection stages indicates that while lncRNAs and circRNAs may be important during initial pathogen recognition, their influence diminishes as the infection progresses. This temporal division of labor highlights miRNAs as the primary regulatory players in sustained immune responses, while lncRNAs and circRNAs contribute more to the early stages of infection.

4.4 Differential expression of up-regulated genes and their role in wheat FHB resistance

The differential expression analysis showed that CS had more up-regulated DEGs than WY across all comparisons (Figures 4A, B). In CS, the number of up-regulated genes exceeded the down-regulated ones at all time points, indicating a stronger overall transcriptional response to infection. However, this response was likely non-specific and may contribute to the higher susceptibility of CS to FHB. In contrast, WY displayed more balanced gene expression, with up-regulated DEGs outnumbering down-regulated ones only at 3 *vs.* 5 dpi. This suggests that while both genotypes mount a defense response to *F. graminearum*, a broader transcriptional activation of CS may lead to an inflammatory response that fails to efficiently control pathogen spread (Pan et al., 2018; Erayman et al., 2015).

The GO term enrichment analysis revealed differences in the molecular functions (MF) of the up-regulated DEGs between CS and WY (Figure 4C). At 0 *vs.* 3 dpi, CS showed enrichment for phenylalanine ammonia-lyase activity and heme binding, both of which are important for phenylpropanoid biosynthesis and managing oxidative stress. These pathways help protect the plant from damage and activate secondary metabolites involved in defense against fungal pathogens (Ramaroson et al., 2022). In contrast, WY was enriched for hydrolase activity and polysaccharide binding, suggesting a more localized, mechanical defense, likely related to altering the cell wall to prevent fungal invasion. At 0 *vs.* 5 dpi, CS enriched for glutathione transferase activity and iron ion binding, both involved in detoxifying reactive oxygen species (ROS) and maintaining redox balance. These processes are crucial during pathogen infection (Van Breusegem et al., 2008). WY, on the other hand, showed enrichment for ATPase-coupled transmembrane transporter activity and ATPase activity, which are involved in ion transport and energy metabolism, possibly indicating a different or less efficient stress response. At 3 *vs.* 5 dpi, CS showed enrichment for protein kinase activity and protein serine/threonine kinase activity, both crucial for signalling immune responses (Pan et al., 2018; Jiang et al., 2022). The presence of phenylalanine ammonia-lyase activity in CS also suggests ongoing activation of pathways important for producing lignin and antimicrobial compounds. WY, by contrast, was enriched for glutathione transferase activity, heme binding, and UDP-glycosyltransferase activity, which are involved in detoxification, ROS management, and modifying secondary metabolites (Wang et al., 2020a; Gharabli et al., 2023; Walker et al., 2024). These differences highlight the more specialized defense mechanisms in WY, which may explain its greater resistance to FHB.

The KEGG pathway enrichment analysis highlighted further differences in the defense mechanisms of CS and WY (Figure 4D). At 0 *vs.* 3 dpi, CS showed enrichment in the biosynthesis of amino acids pathway, which is important for producing amino acids needed for protein synthesis and secondary metabolite formation. In contrast, WY enriched for plant hormone signal transduction, suggesting that hormonal regulation (such as JA and SA) plays a key role in its early defense response (Pieterse et al., 2012; Pan et al., 2018). This supports the idea that the resistance of WY to FHB may involve a more coordinated immune response, with specific hormones activated by fungal infection. At 0 *vs.* 5 dpi, CS enriched for pathways related to phenylpropanoid biosynthesis and carbon metabolism, both crucial for stress responses, secondary metabolite production, and energy balance during infection. In contrast, WY enriched for glutathione metabolism and ABC transporters, which help detoxify oxidative stress and remove harmful substances (Wang et al., 2020b; Gullner et al., 2018). This suggests that the resistance of WY may depend more on detoxification and transporting stress-related compounds. At 3 *vs.* 5 dpi, the differences were even more pronounced. CS enriched for plant-pathogen interaction and amino acid biosynthesis, which are important for immune recognition and defense protein synthesis. WY, however, enriched for glutathione metabolism and ABC transporters, helping maintain cellular integrity during stress (Wang et al., 2020b; Gullner et al., 2018). Additionally, WY showed enrichment in flavonoid biosynthesis and phenylalanine metabolism, which suggests a shift towards producing secondary metabolites as a defense strategy to limit pathogen growth.

A detailed analysis of the up-regulated genes in the glutathione metabolism pathway during the 3 *vs.* 5 dpi comparison in WY identified 93 DEGs, of which 80 were shared between CS and WY (Figure 4E). These genes are involved in detoxifying ROS and maintaining cellular redox balance, which are essential for plant defense against oxidative stress caused by fungal infection. Interestingly, 13 DEGs were only differentially expressed in WY, suggesting a unique resistance mechanism in this genotype. The heatmap analysis showed that two genes, TraesCS2B03G1525600 and TraesCS6B03G0425000, had significantly higher expression levels in CS compared to WY, which may indicate a larger-scale response in CS, but this doesn't necessarily mean a more effective defense. The differences in gene expression highlight the distinct defense strategies between the two genotypes, with WY showing a more localized, focused response to oxidative stress.

4.5 Integrated analysis of miRNA-mRNA targeting relationships

The results show a clear difference in miRNA profiles between the two wheat genotypes, especially at different time points after

infection. At 0 *vs.* 3 dpi, only two miRNAs were common between CS and WY, indicating that each genotype activates different miRNA pathways early in the infection. However, by 0 *vs.* 5 dpi and 3 *vs.* 5 dpi, 16 and 15 miRNAs were shared between the genotypes, suggesting a more similar response as the infection progresses. Notably, novel_miR_228 and tae-miR1122a were key miRNAs with different roles in the two genotypes. Novel_miR_228, which was up-regulated in CS, down-regulated 28 and 23 genes related to defense in CS at 0 *vs.* 5 dpi and 3 *vs.* 5 dpi, respectively (Figure 5B) (Supplementary Figure 7B). This suggests that it may weaken the defense response of host, making CS more susceptible to FHB. On the other hand, tae-miR1122a, which was down-regulated in WY, up-regulated 27 and 23 defense-related genes in WY at the same time points, suggesting that it helps enhance resistance in WY. These findings support the idea that miRNAs play a role in regulating the defense mechanisms of host in response to fungal stress, contributing to FHB resistance (MaChado et al., 2018; Jin et al., 2020).

The GO analysis of the DEGs regulated by the identified miRNAs revealed important molecular functions linked to FHB resistance. In CS, novel_miR_228 down-regulated genes related to key functions like histone acetyltransferase activity and DNA binding, which are essential for transcription and DNA repair. This down-regulation could weaken the ability of host to manage gene expression and repair DNA, making it more vulnerable to fungal damage. Genes like TraesCS1A03G0367900 and TraesCS6B03G0592100 showed strong associations with these functions, highlighting how *F. graminearum* might exploit these weaknesses to infect the host. In contrast, in WY, tae-miR1122a up-regulated genes involved in metal ion binding, such as TraesCS2B03G0537600, TraesCS2D03G0425900, and TraesCS6A03G0855700 (Figure 5C) (Supplementary Figure 7C). These genes help maintain cellular stability under stress, suggesting that they play a role in the defense of plants (Wang et al., 2020b). The up-regulation of these genes in WY may support antioxidant and detoxification pathways, helping the plant cope with the oxidative damage caused by fungal infection (Gullner et al., 2018). The significant enrichment of these genes in WY points to their potential role in strengthening the ability of host to resist fungal stress by binding toxic ions released by the pathogen, thus helping maintain cellular integrity and defense against FHB.

The comparison between WY and CS shows a clear difference in how they respond to *F. graminearum* infection. In WY, the expression of defense-related genes is stronger, with down-regulation of miRNAs like tae-miR1122a and up-regulation of their target genes involved in metal ion binding and stress responses. This suggests that the resistance of WY to FHB may be due to better regulation of metal ion balance, oxidative stress, and immune signaling, all of which are essential for protecting cells during infection. In CS, the up-regulation of novel_miR_228 and the down-regulation of key defense genes indicate a weaker response to the fungal infection. This may explain why CS, a more susceptible genotype, struggles to mount a strong defense against *F. graminearum*, making it more prone to disease.

4.6 GO term and KEGG pathway enrichment of DEGs regulated by miRNA

GO and KEGG pathway analyses further illustrate the functional divergence between WY and CS in response to FHB. In GO classification, the DEGs of WY were enriched in defense-related categories, including gene silencing by RNA and defense response (Figure 6). This enrichment suggests that WY is genetically predisposed to actively respond to pathogen invasion, prioritizing immune functions over general metabolic processes. In the cellular component (CC) category, the DEGs of WY were primarily localized to membrane and cytoplasmic components, critical sites for pathogen recognition and signaling. In contrast, CS displayed a more generalized distribution across organelles, indicative of a less-targeted defense approach.

KEGG pathway enrichment revealed that immune pathways were central to the response of WY, particularly the plant-pathogen interaction pathway (Figure 7). This pathway was consistently enriched across all time points, reflecting a robust immune response in WY, channeling resources toward defense mechanisms upon pathogen detection (Zaynab et al., 2018). Conversely, the response of CS emphasized starch and sucrose metabolism pathways (Table 2), suggesting a baseline metabolic stability rather than a targeted immune response. This finding is consistent with previous studies that susceptible genotypes often allocate resources to general metabolism at the expense of pathogen-specific defenses (Brauer et al., 2019).

Additionally, pathways like phenylpropanoid biosynthesis and phenylalanine metabolism were enriched in WY but not in CS (Table 3). These pathways are known to produce secondary metabolites that strengthen structural defenses and stress responses, underscoring the preparedness of WY for pathogen attack (Zaynab et al., 2018). The divergence in pathway enrichment between WY and CS highlights how WY allocates resources more strategically, channeling efforts toward immune functions while CS appears to prioritize metabolic maintenance. This strategic resource allocation likely contributes to enhancing the resistance of WY to FHB.

4.7 Role of miRNAs in regulating defense mechanisms

This study highlighted the critical role of miRNAs in response to FHB resistance. miRNAs are known to target specific mRNAs, modulating gene expression in response to stressors, including pathogens. For example, miRNAs are integral to the defense mechanisms of plants against pathogens. They regulate the expression of genes involved in phytohormone signaling, ROS production, and nucleotide-binding site-leucine-rich repeat (NBS-LRR) gene expression, which are critical for mounting an effective defense response (Yang et al., 2021). The interaction between miRNAs and lncRNAs complicated this regulatory network, further fine-tuning the immune response of plants (Song et al.,

2021). We found that WY exhibited a notably stronger miRNA response than CS, particularly at later infection stages (Table 1). miRNAs in WY were enriched in defense-related pathways such as glutathione metabolism and plant-pathogen interactions, crucial for oxidative stress regulation—an essential factor in plant defense that limits pathogen spread (Dorion et al., 2021). The enrichment of miRNAs in plant-pathogen interaction pathways supports their role in enhancing immune responses, allowing WY to recognize and respond more effectively to *F. graminearum* invasion. In contrast, miRNA pathways in CS were predominantly linked to general metabolic processes rather than defense-specific responses, suggesting that the miRNAs of CS may not be primed for targeted pathogen defense. This aligns with previous studies that susceptible genotypes often prioritize maintaining metabolic homeostasis over rapid immune activation, leaving them more vulnerable to pathogens (Brauer et al., 2019). These differences indicate that miRNAs in resistant genotypes like WY are more adept at regulating specific pathways linked to pathogen defense, facilitating a faster and more efficient immune response.

In summary, this study provides valuable insight into the molecular mechanisms underlying FHB resistance in two Sichuan wheat landraces WY (HR) and CS (HS). Through a comprehensive transcriptomic analysis, we identified distinct differences in how these two genotypes respond to *F. graminearum*. The resistant genotype, WY, showed a targeted miRNA response, particularly at later infection stages, which effectively regulated pathways important for defense, such as glutathione metabolism and phenylpropanoid biosynthesis. In contrast, the susceptible genotype, CS, exhibited a broader transcriptional response focused more on general metabolism than specific pathogen defense. Our findings emphasize the critical role of miRNA-mRNA interactions in enhancing the resistance of WY. Although lncRNAs and circRNAs were also involved in resistance, their contributions were limited in comparison with miRNAs. This research not only emphasized the importance of miRNA-mediated defense mechanisms but also suggested potential molecular targets to improve FHB resistance in breeding programs. Future studies should focus on validating these pathways and incorporating deep learning method and genomic selection (GS) (Chen et al., 2024; Wang et al., 2024b; Ma et al., 2024) techniques to develop more resilient wheat varieties, ensuring sustainable wheat production amid ongoing challenges posed by FHB disease.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRJNA1183537.

Author contributions

LW: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. JW: Investigation, Writing – review & editing. SS: Investigation, Writing – review & editing. ZY: Writing – review & editing. XH: Conceptualization, Resources, Writing – review & editing, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Funding was provided by "the Key Research and Development Program of Science and Technology Department of Sichuan Province, grant number 2021YFS0342" and "Fundamental Research Funds of China West Normal University, grant number 18Q053 and 20A018".

Acknowledgments

We would like to thank Dr. Youping Xu and Ms. Tingting Wang for preparing the inoculum, and thank Professor Zehong Yan for providing wheat materials and revising the article.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1537605/full#supplementary-material

References

Alaux, M., Dyer, S., and Sen, T. Z. (2023). "Wheat data integration and FAIRification: IWGSC, grainGenes, ensembl and other data repositories," in *The wheat genome*. Eds. R. Appels, K. Eversole, C. Feuillet and D. Gallagher (Compendium of Plant Genomes. Springer, Cham). doi: 10.1007/978-3-031-38294-9_2

Bai, G., and Shaner, G. (2004). Management and resistance in wheat and barley to fusarium head blight. *Annu. Rev. Phytopathol.* 42, 135–161. doi: 10.1146/annurev.phyto.42.040803.140340

Bernardo, A., Bai, G., Guo, P., Xiao, K., Guenzi, A. C., and Ayoubi, P. (2007). Fusarium graminearum-induced changes in gene expression between Fusarium head blight-resistant and susceptible wheat cultivars. *Funct. Integr. Genomics* 7, 69–77. doi: 10.1007/s10142-006-0028-1

Biselli, C., Bagnaresi, P., Faccioli, P., Hu, X., Balcerzak, M., Mattera, M. G., et al. (2018). Comparative transcriptome profiles of near-isogenic hexaploid wheat lines differing for effective alleles at the 2DL FHB resistance QTL. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00037

Bo, X., and Wang, S. (2005). TargetFinder: a software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. *Bioinformatics* 21, 1401–1402. doi: 10.1093/bioinformatics/bti211

Brauer, E. K., Rocheleau, H., Balcerzak, M., Pan, Y. L., Fauteux, F., Liu, Z. Y., et al. (2019). Transcriptional and hormonal profiling of Fusarium graminearum-infected wheat reveals an association between auxin and susceptibility. *Physiol. Mol. Plant Pathol.* 107, 33–39. doi: 10.1016/j.pmpp.2019.04.006

Buerstmayr, H., Ban, T., and Anderson, J. A. (2009). QTL mapping and marker assisted selection for Fusarium head blight resistance in wheat: a review. *Plant Breed.* 128, 1–26. doi: 10.1111/j.1439-0523.2008.01550.x

Buerstmayr, M., Steiner, B., and Buerstmayr, H. (2020). Breeding for Fusarium head blight resistance in wheat-progress and challenges. *Plant Breed.* 139, 429–454. doi: 10.1111/pbr.12797

Chen, D., Zhang, Z., Chen, Y., Li, B., Chen, T., and Tian, S. (2023). Transcriptional landscape of pathogen-responsive lncRNAs in tomato unveils the role of hydrolase encoding genes in response to Botrytis cinerea invasion. *Plant Cell Environ.* 47, 651–663. doi: 10.1111/pce.14757

Chen, Y., Kistler, H. C., and Ma, Z. (2019). Fusarium graminearum trichothecene mycotoxins: Biosynthesis, regulation, and management. *Annu. Rev. Phytopathol.* 57, 15–39. doi: 10.1146/annurev-phyto-082718-100318

Chen, Y., Wang, W., Yang, Z., Peng, H., Ni, Z., Sun, Q., et al. (2024). Innovative computational tools provide new insights into the polyploid wheat genome. *Abiotech* 5, 52–70. doi: 10.1007/s42994-023-00131-7

Cui, J., Jiang, N., Hou, X., Wu, S., Zhang, Q., Meng, J., et al. (2020). Genome-wide identification of lncRNAs and analysis of ceRNA networks during tomato resistance to Phytophthora infestans. *Phytopathology* 110, 456–464. doi: 10.1094/PHYTO-04-19-0137-R

Dorion, S., Ouellet, J. C., and Rivoal, J. (2021). Glutathione metabolism in plants under stress: beyond reactive oxygen species detoxification. *Metabolites* 11, 641. doi: 10.3390/metabo11090641

Doron, B., Manda, W., Aaron, G., Marks, D. S., and Chris, S. (2008). The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 36, 149–153. doi: 10.1093/nar/gkm995

Duan, X., Song, X., Wang, J., and Zhou, M. (2020). Genome-wide identification and characterization of Fusarium graminearum-responsive lncRNAs in Triticum aestivum. . *Genes* 11, 1135. doi: 10.3390/genes11101135

Erayman, M., Turktas, M., Akdogan, G., Gurkok, T., Inal, B., Ishakoglu, E., et al. (2015). Transcriptome analysis of wheat inoculated with Fusarium graminearum. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00867

Fahlgren, N., Howell, M. D., Kasschau, K. D., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., et al. (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PloS One* 2, e219. doi: 10.1371/journal.pone.0000219

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52. doi: 10.1093/nar/gkr688

Fu, K., Wu, Q., Jiang, N., Hu, S., Ye, H., Hu, Y., et al. (2023). Identification and expressional analysis of siRNAs responsive to Fusarium graminearum infection in wheat. *Int. J. Mol. Sci.* 24, 16005. doi: 10.3390/ijms242116005

Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for *de novo* circular RNA identification. *Genome Biol.* 16, 4. doi: 10.1186/s13059-014-0571-3

Gharabli, H., Della Gala, V., and Welner, D. H. (2023). The function of UDP-glycosyltransferases in plants and their possible use in crop protection. *Biotechnol. Adv.* 67, 108182. doi: 10.1016/j.bioteChadv.2023.108182

Gullner, G., Komives, T., Király, L., and Schröder, P. (2018). Glutathione S-transferase enzymes in plant-pathogen interactions. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01836

Hu, W. J., Fu, L. P., Gao, D. R., Li, D. S., Liao, S., and Lu, C. B. (2023). Marker-assisted selection to pyramid Fusarium head blight resistance loci Fhb1 and Fhb2 in the high-quality soft wheat cultivar Yangmai 15. *J. Integr. Agr.* 22, 360–370. doi: 10.1016/j.jia.2022.08.057

Hu, X., Rocheleau, H., McCartney, C., Biselli, C., Bagnaresi, P., Balcerzak, M., et al. (2019). Identification and mapping of expressed genes associated with the 2DL QTL for fusarium head blight resistance in the wheat line Wuhan 1. *BMC Genet.* 20, 47. doi: 10.1186/s12863-019-0748-6

Jiang, L., Zhang, S., Su, J., Peck, S. C., and Luo, L. (2022). Protein kinase signaling pathways in plant-Colletotrichum interaction. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.829645

Jiao, J., and Peng, D. (2018). Wheat microRNA1023 suppresses invasion of Fusarium graminearum via targeting and silencing FGSG_03101. *J. Plant Interact.* 13, 514–521. doi: 10.1080/17429145.2018.1528512

Jin, X., Jia, L., Wang, Y., Li, B., Sun, D., and Chen, X. (2020). Identification of Fusarium graminearum-responsive miRNAs and their targets in wheat by sRNA sequencing and degradome analysis. . *Funct. Integr. Genomics* 20, 51–61. doi: 10.1007/s10142-019-00699-8

Kazan, K., and Gardiner, D. M. (2018). Transcriptomics of cereal-Fusarium graminearum interactions: what we have learned so far. *Mol. Plant Pathol.* 19, 764–778. doi: 10.1111/mpp.12561

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

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 278. doi: 10.1186/s13059-019-1910-1

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162. doi: 10.1093/nar/gky1141

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9, 357–359. doi: 10.1038/nmeth.1923

Li, G., Zhou, J., Jia, H., Gao, Z., Fan, M., Luo, Y., et al. (2019a). Mutation of a histidine-rich calcium-binding-protein gene in wheat confers resistance to Fusarium head blight. *Nat. Genet.* 51, 1106–1112. doi: 10.1038/s41588-019-0426-7

Li, T., Zhang, H., Huang, Y., Su, Z., Deng, Y., Liu, H., et al. (2019b). Effects of the Fhb1 gene on Fusarium head blight resistance and agronomic traits of winter wheat. *Crop J.* 7, 799–808. doi: 10.1016/j.cj.2019.03.005

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

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, X., Wang, H., Wu, S., Han, B., Cui, D., Liu, J., et al. (2024). DeepCCR: large-scale genomics-based deep learning method for improving rice breeding. *Plant Biotechnol. J.* 22 (10), 2691–2693. doi: 10.1111/pbi.14384

Ma, Z., Xie, Q., Li, G., Jia, H., Zhou, J., Kong, Z., et al. (2020). Germplasms, genetics and genomics for better control of disastrous wheat Fusarium head blight. *Theor. Appl. Genet.* 133, 1541–1568. doi: 10.1007/s00122-019-03525-8

MaChado, A. K., Brown, N. A., Urban, M., Kanyuka, K., and Hammond-Kosack, K. E. (2018). RNAi as an emerging approach to control Fusarium head blight disease and mycotoxin contamination in cereals. *Pest Manage. Sci.* 74, 790–799. doi: 10.1002/ps.4748

Oliveros, J. C. (2015). "Venny," in *An interactive tool for comparing lists with Venn's diagrams*. Available at: https://bioinfogp.cnb.csic.es/tools/venny/index.html (Accessed May 25, 2024).

Pan, Y., Liu, Z., Rocheleau, H., Fauteux, F., Wang, Y., McCartney, C., et al. (2018). Transcriptome dynamics associated with resistance and susceptibility against fusarium head blight in four wheat genotypes. *BMC Genomics* 19, 642. doi: 10.1186/s12864-018-5012-3

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

Pertea, G., and Pertea, M. (2020). GFF utilities: gffRead and gffcompare. *F1000Research* 9, 304. doi: 10.12688/f1000research.23297.1

Pieterse, C. M., van der Does, D., Zamioudis, C., Leon-Reyes, A., and Van Wees, S. C. (2012). Hormonal modulation of plant immunity. *Annu. Rev. Cell Dev. Bi.* 28, 489–521. doi: 10.1146/annurev-cellbio-092910-154055

Ramaroson, M. L., Koutouan, C., Helesbeux, J. J., Le Clerc, V., Hamama, L., Geoffriau, E., et al. (2022). Role of phenylpropanoids and flavonoids in plant resistance to pests and diseases. *Molecules* 27, 8371. doi: 10.3390/molecules27238371

Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA* 10, 1507–1517. doi: 10.1261/rna.5248604

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358. doi: 10.1016/j.cell.2011.07.014

Song, L., Fang, Y., Chen, L., Wang, J., and Chen, X. (2021). Role of non-coding RNAs in plant immunity. *Plant Commun.* 2, 100180. doi: 10.1016/j.xplc.2021.100180

Soresi, D., Díaz, M., Bagnaresi, P., Gallo, C., and Carrera, A. (2023). Gene and lncRNA expression patterns associated with Qfhs.ndsu-3AS Fusarium head blight resistance QTL in durum wheat. *Can. J. Plant Pathol.* 45, 41–56. doi: 10.1080/07060661.2022.2091663

Su, Z., Bernardo, A., Tian, B., Chen, H., Wang, S., Ma, H., et al. (2019). A deletion mutation in TaHRC confers Fhb1 resistance to Fusarium head blight in wheat. *Nat. Genet.* 51, 1099–1105. doi: 10.1038/s41588-019-0425-8

Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., et al. (2023). SRplot: A free online platform for data visualization and graphing. *PloS One* 18, e0294236. doi: 10.1371/journal.pone.0294236

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Van Breusegem, F., Bailey-Serres, J., and Mittler, R. (2008). Unraveling the tapestry of networks involving reactive oxygen species in plants. *Plant Physiol.* 147, 978–984. doi: 10.1104/pp.108.122325

Walker, P. L., Belmonte, M. F., McCallum, B. D., McCartney, C. A., Randhawa, H. S., and Henriquez, M. A. (2024). Dual RNA-sequencing of Fusarium head blight resistance in winter wheat. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1299461

Walter, S., Kahla, A., Arunachalam, C., Perochon, A., Khan, M. R., Scofield, S. R., et al. (2015). A wheat ABC transporter contributes to both grain formation and mycotoxin tolerance. *J. Exp. Bot.* 66, 2583–2593. doi: 10.1093/jxb/erv048

Wan, Y. F., Yen, C., and Yang, J. L. (1997). Sources of resistance to head scab in Triticum. *Euphytica* 94, 31–36. doi: 10.1023/A:1002982005541

Wang, R., Huang, J., Liang, A., Wang, Y., Mur, L. A. J., Wang, M., et al. (2020b). Zinc and copper enhance cucumber tolerance to fusaric acid by mediating its distribution and toxicity and modifying the antioxidant system. *Int. J. Mol. Sci.* 21, 3370. doi: 10.3390/ijms21093370

Wang, X., Li, G., Jia, H., Cheng, R., Zhong, J., Shi, J., et al. (2024). Breeding evaluation and precise mapping of Fhb8 for Fusarium head blight resistance in wheat (Triticum aestivum). *Plant Breed.* 143, 26–33. doi: 10.1111/pbr.13113

Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang, K., et al. (2020a). Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. *Science* 368, 822–823. doi: 10.1126/science.aba5435

Wang, L., Wang, L., Li, Q., Liu, Z., Surendra, A., Pan, Y., et al. (2018). Integrated transcriptome and hormone profiling highlight the role of multiple phytohormone pathways in wheat resistance against fusarium head blight. *PloS One* 13, e0207036. doi: 10.1371/journal.pone.0207036

Wang, H., Yan, S., Wang, W., Cheng, Y., Hong, J., He, Q., et al. (2024). Cropformer: An interpretabl deep learning framework for crop genome prediction. *Plant Commun.* doi: 10.1016/j.xplc.2024.101223

Xiao, J., Jin, X. H., Jia, X. P., Wang, H. Y., Cao, A. Z., Zhao, W. P., et al. (2013). Transcriptome-based discovery of pathways and genes related to resistance against Fusarium head blight in wheat landrace Wangshuibai. *BMC Genom.* 14, 197. doi: 10.1186/1471-2164-14-197

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483

Yang, X., Zhang, L., Yang, Y., Schmid, M., and Wang, Y. (2021). miRNA mediated regulation and interaction between plants and pathogens. *Int. J. Mol. Sci.* 22, 2913. doi: 10.3390/ijms22062913

Yin, J., Han, X., Zhu, Y., Fang, Z., Gao, D., and Ma, D. (2022). Transcriptome profiles of circular RNAs in common wheat during Fusarium head blight disease. . *Data* 7, 121. doi: 10.3390/data7090121

Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11, R14. doi: 10.1186/gb-2010-11-2-r14

Yu, Y., Zhou, Y., Feng, Y., He, H., Lian, J., Yang, Y., et al. (2019). Transcriptional landscape of pathogen-responsive lncRNAs in rice unveils the role of ALEX1 in jasmonate pathway and disease resistance. *Plant Biotechnol. J.* 18, 679–690. doi: 10.1111/pbi.13234

Zaynab, M., Fatima, M., Abbas, S., Sharif, Y., Umair, M., Zafar, M. H., et al. (2018). Role of secondary metabolites in plant defense against pathogens. *Microb. Pathogenesis* 124, 198–202. doi: 10.1016/j.micpath.2018.08.034

Zhang, S., Shen, S., Yang, Z., Kong, X., Liu, F., and Zhen, Z. (2020). Coding and Non-coding RNAs: Molecular basis of forest-insect outbreaks. *Front. Cell Dev. Biol.* 8. doi: 10.3389/fcell.2020.00369

Zhang, Y. D., Yang, Z. B., Ma, H. C., Huang, L. Y., Ding, F., Du, Y. Y., et al. (2021). Pyramiding of Fusarium head blight resistance quantitative trait loci, Fhb1, Fhb4, and Fhb5, in modern Chinese wheat cultivars. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.694023

Zhang, F., Zhang, H., Liu, J., Ren, X., Ding, Y., Sun, F., et al. (2024). Fhb9, a major QTL for Fusarium head blight resistance improvement in wheat. *J. Integr. Agr.* doi: 10.1016/j.jia.2024.03.045

Zhu, Z., Hao, Y., Mergoum, M., Bai, G., Humphreys, G., Cloutier, S., et al. (2019). Breeding wheat for resistance to Fusarium head blight in the Global North: China, USA, and Canada. *Crop J.* 7, 730–738. doi: 10.1016/j.cj.2019.06.003

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 May 2025
DOI 10.3389/fpls.2025.1546580

OPEN ACCESS

EDITED BY
Yuqing He,
Huazhong Agricultural University, Wuhan, China

REVIEWED BY
Gurjeet Singh,
Texas A and M University, United States
Ujjawal Kumar Singh Kushwaha,
Nepal Agricultural Research Council, Nepal
Qiaojun Lou,
Shanghai Agrobiological Gene Center, China

*CORRESPONDENCE
Longzhi Han
hanlongzhi@caas.cn
Jianxin Deng
djxin555@hotmail.com
Xiaoding Ma
maxiaoding@caas.cn

†These authors have contributed equally to this work

RECEIVED 17 December 2024
ACCEPTED 21 April 2025
PUBLISHED 13 May 2025

CITATION
Xue L, Wang S, Zhang Q, Han B, Cui D, Han L, Deng J and Ma X (2025) Identification of key genes associated with mesocotyl length through a genome-wide association study in rice.
Front. Plant Sci. 16:1546580.
doi: 10.3389/fpls.2025.1546580

Identification of key genes associated with mesocotyl length through a genome-wide association study in rice

Li Xue[1,2†], Sen Wang[2,3†], Qiuyu Zhang[2,4], Bing Han[2], Di Cui[2], Longzhi Han[2*], Jianxin Deng[1*] and Xiaoding Ma[2*]

[1]College of Agriculture, Yangtze University, Hubei, China, [2]State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [3]National Key Laboratory, Sichuan Agricultural University, Sichuan, China, [4]Chongqing Engineering Research Center of Specialty Crop Resources, Chongqing Normal University, Chongqing, China

Mesocotyl length is a key trait affecting seedling emergence and establishment in dry direct-seeded rice, with longer mesocotyls promoting rapid and uniform emergence, thereby forming larger effective populations. Therefore, mining genes associated with mesocotyl length will facilitate the development of rice varieties suitable for dry direct seeding. In this study, 300 rice germplasm resources with a wide range of sources were selected as experimental materials. Phenotypic traits such as mesocotyl length and seedling emergence rate were systematically determined in each variety by setting different mulch depth treatments. Genome-wide association analysis (GWAS) was used to locate QTL controlling mesocotyl length and predict candidate genes. The results showed that mesocotyl length increased significantly with greater soil cover depth, while excessively deep sowing treatments inhibited seedling emergence. The GWAS analysis identified four QTLs associated with mesocotyl length and two QTLs associated with seedling emergence, with phenotypic contributions of 6.96-8.48%. Among them, the mesocotyl length-related QTL *qML3* located at 28.03-28.43 Mb on chromosome 3 was detected at both sowing depths. Gene annotation analysis identified nine candidate genes related to plant hormones and transcription factors for *qML3*. Further investigation revealed three genes (*LOC_Os03g49250*, *LOC_Os03g49400*, and *LOC_Os03g49510*) exhibiting distinct haplotypes with significant differences in mesocotyl length, suggesting they may be causal genes for *qML3*. The results provide new clues to elucidate the molecular mechanism of rice mesocotyl development and lay an important foundation for subsequent gene function verification and molecular breeding. In the future, the functions of these candidate genes will be verified by transgenic and other methods, and molecular markers will be developed for genetic improvement of drought-tolerant rice varieties.

KEYWORDS

rice, mesocotyl length, genome-wide association analysis, haplotype, candidate gene

1 Introduction

Rice (*Oryza sativa* L.) is a major global food crop that is traditionally grown by transplanting seedlings into paddy fields (Zhao et al., 2018). However, conventional transplanted rice production systems face significant challenges because of climate change, urbanization, water scarcity, and labor shortages in many regions (Sagare et al., 2020; Zhang et al., 2023). Dry direct seeding, as an alternative to transplanting, saves time and labor while reducing water and energy consumption (Liu et al., 2019; Ohno et al., 2018; Zhan et al., 2019). Additionally, direct dry seeding eliminates the transplanting step, not only accelerating seedling cultivation but also increasing yield (Ishfaq et al., 2018; Muhammad et al., 2020). Although direct seeding of rice has multiple advantages over traditional planting methods, its popularization still faces several challenges, such as a low seedling emergence rate (ER), poor seedling establishment, weed infestation, and susceptibility to lodging (Mahender et al., 2015; Lee et al., 2017). Therefore, a key challenge at this stage is to identify effective methods to promote rapid and high-quality germination and growth of direct-seeded rice.

The mesocotyl, which forms when rice seeds germinate in the dark, is the organ that connects the germinal sheath to the seed root (Schillinger et al., 1998; Zhang et al., 2012). Light is a key factor influencing mesocotyl length (ML). Typically, light inhibits ML, whereas darkness significantly promotes its elongation. The study revealed that under light-free conditions, the cells of the mesocotyl not only actively divided but also exhibited particularly pronounced elongation and growth. This process is accompanied by increased levels of phytohormones such as gibberellins (GA), auxins (IAA), and cytokinins (CK), which act synergistically to strongly promote mesocotyl elongation (Li et al., 2012, 2013). In addition, temperature also affects the elongation of the ML. The optimal temperature range is 25-30°C, within which the ML increases with the rise in temperature. Beyond the suitable temperature range, as the temperature continues to rise, the growth rate of the rice mesocotyl slows down (Cao et al., 2002). Temperature is another factor that regulates the elongation of the mesodermal axis. The optimal temperature promotes growth, while both high and low temperatures are detrimental to the elongation of the mesodermal axis. Thus, the mesocotyl is not only a central feature of seed germination adapted to dark conditions but also a crucial factor in regulating the initial morphogenesis and growth strategy of rice seedlings. Rice seeds with longer mesocotyls demonstrate superior seedling emergence and uniformity (Turner et al., 1982; Mgonja et al., 1993; Chung, 2010). ML is recognized as a key trait for breeding direct-seeded rice varieties (Zhan et al., 2019). Therefore, an economical and effective method to promote the adoption of direct seeding technology is to investigate the locus ML, analyze the underlying genetic mechanisms, and screen and innovate germplasm with long mesocotyls.

ML can be influenced by external environmental factors but is also determined by intrinsic genetic mechanisms (Gray et al., 1998; Kato and Katsura, 2014; Liu et al., 2023). ML is typically a quantitative inherited trait controlled by multiple genes (Wu et al., 2005; Wang et al., 2021; Zhang et al., 2023). Therefore, mining of ML-related genes and molecular breeding could facilitate the development of direct-seeded rice varieties. With the rapid development of molecular marker technology and genome sequencing technology, many ML-related QTLs have been identified in various natural populations through biparental QTL analysis and genome-wide association studies (Lee et al., 2017; Liu et al., 2019; Zhan et al., 2019; Liu et al., 2023; Wang et al., 2023);. Eleven QTLs were identified using a population of recombinant inbred lines (RILs) from an interspecific cross between *O. sativa* and *O. rufipogon* (Cai and Morishima, 2002). In a population of backcrossed inbred lines (BILs) from a cross between Kasalath and Nipponbare, five QTLs for ML were detected (Lee et al., 2011). Three major QTLs, designated *qMel-1*, *qMel-3*, and *qMel-6*, located on chromosomes 1, 3, and 6, respectively, were identified through the analysis of 98 self-crossed lines from the Kasalath and Nipponbare crosses (Lee et al., 2017). In addition, 9, 17, 16, and 11 QTLs for ML were identified using GWAS (Wu et al., 2015; Lu et al., 2016; Liu et al., 2019; Kwon, 2021). To date, several ML-related genes have been cloned. Xiong et al. (2017) reported that ethylene (ETH) regulates ML by repressing *GY1* gene expression and controlling the biosynthesis of jasmonic acid (JA). Sun et al. (2018) reported that the *OsGSK2* gene regulates ML by coordinating the signaling of strigolactone (SL) and brassinolide (BR). Zheng et al. (2020) reported that the karrikin signaling pathway and the BR signaling pathway may jointly regulate mesoembryo axis elongation.

ML as a crucial indicator of rice growth and development, demonstrates significant pleiotropic effects through its regulatory genes in crop breeding. Choi (2003) reported that *OsEXP4* overexpression produced a biphasic phenotype with 12% of plants showing increased height and 88% displaying reduced stature compared to controls; the overexpressing lines exhibited 31% and 97% increases in coleoptile and ML, respectively, while antisense plants showed 28% and 43% reductions. Ma et al. (2013) found that ethylene treatment significantly promoted coleoptile and mesocotyl elongation in etiolated *MHZ7* overexpressing plants while increasing grain length in transgenic plants, whereas *mhz7* mutants exhibited reduced thousand-grain weight. Lv et al. (2021) demonstrated that *OsPAO5* knockout across different genetic backgrounds significantly enhanced mesocotyl elongation, improved emergence rate and speed under deep-sowing conditions, and concurrently increased grain length, grain weight, and grains per panicle, ultimately boosting single-plant yield. Through analysis of 165 rice accessions, Meng et al. (2023) further identified the superior gene *OsML1*, which not only markedly improved emergence rate in deep sowing but also conferred comprehensive enhancements in agronomic traits including plant height, panicle length, and grain width in overexpression lines.

Discovering rice resources with long mesocotyls, identifying and cloning the genes that control ML, and analyzing their molecular regulatory mechanisms are key steps in breeding rice varieties with long mesocotyls. In this study, a GWAS was conducted on a rice population consisting of 300 genotypes to

assess mesocotyl phenotypes. Four new QTLs were ultimately identified, and three new candidate genes for the ML were initially predicted. This study lays the foundation for the cloning of new ML genes and analyzing the genetic mechanisms of rice ML.

2 Materials and methods

2.1 Plant materials

In this study, we selected 300 rice varieties, all of which were derived from the National Gene Bank of China. Based on previous research results, we selected samples including local varieties, superior varieties, foundation varieties and their derivatives to ensure a wide distribution and high genetic diversity. (Supplementary Table S1) (Cui et al., 2022). These accessions were grown in Hainan Province, China, in 2022. The seeds were harvested in April 2023, air-dried, and stored in the laboratory of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China.

2.2 Mesocotyl length measurement and emergence rate evaluation

Seeds were placed in an oven at 45°C for one week to break dormancy. For each treatment, 50 full and uniform seeds were independently selected and sterilized with 1.4% sodium hypochlorite solution and then placed in an oven at 35°C for 2 days to promote germination. To assess the effect of different sowing depths on mesocotyl length (ME), three mulching depth treatments of 2, 5 and 7 cm were set up in 84-well plastic frames, with three biological replicates at each depth, and 20 high-quality germinated seeds were sown in each replicate, and then incubated for 7 days in the dark at 28°C. The mesocotyl length of each seedling was measured using a ruler (distance from root base to the node of the embryonic sheath) (Huang et al., 2022). The plates were incubated in a 28°C dark growth chamber for 7 days, after which the mesocotyl length (ML) was measured. For the emergence rate (ER), germinated seeds were sown at depths of 5 cm and 7 cm respectively, and cultivated under conditions of 30°C and 75% relative humidity with 20,000 lux white fluorescent lighting (14 h light/10 h dark photoperiod) for 7 days. All experiments included three biological replicates, with at least 10 seedlings measured per replicate. Phenotypic variation analysis, correlation analysis, and Student's t-tests were performed using IBM SPSS v26.

2.3 Genome-wide association study and candidate genes identification

SNPs with a minor allele frequency (MAF) greater than 5% and deletion under 40% were selected for association analysis, and 711,268 variants were used for GWAS. The Yeo–Johnson transformation was applied to normalize the data for GWAS analysis (Yeo and Johnson, 2000). The GWAS was performed using a mixed linear model (MLM) and the script run_pipeline.pl in the TASSEL software (Bradbury et al., 2007). The significance threshold was set at 1.0×10^{-7}. Principal component analysis (PCA) was used to characterize and differentiate the samples using Plink software (Purcell et al., 2007). SNP density heatmaps and Manhattan and Q–Q plots were generated using the CMplot package in R software (Yin et al., 2021).

QTLs were screened based on the results of the Manhattan plot, where column peaks exceeding the threshold line were considered as candidate QTLs. According to existing studies, genome-wide linkage disequilibrium (LD) attenuation in cultivated rice usually occurs in the range of 100–200 kb. Therefore, we designated the 200 kb region upstream and downstream of the QTL peak SNP locus as the QTL region. We extracted the genes encoding the target intervals from the Rice Expression Database (http://expression.ic4r.org/) and screened them using gene function annotation information from the China Rice data Center (https://ngdc.cncb.ac.cn/red/index). First, transposon- and reverse transcript-transposon-related genes were excluded, and then candidate genes related to the target QTL were preliminarily screened based on the gene functional annotations in combination with existing research reports on rice mesocotyl and internode elongation genes (Kawahara et al., 2013).

2.4 Haplotype analysis

Haplotype analyses were performed on the QTL candidate genes that were repeatedly detected under two different environmental treatments. High-confidence SNPs in the coding regions of the candidate genes within the QTL regions were extracted from the Rice Variation Atlas v2.0 database (http://ricevarmap.ncpgr.cn/) (Zhao et al., 2021), and the haplotypes of the candidate genes within the QTL regions were analyzed. The significance of the differences between the phenotypic values and haplotypes was determined by independent samples t-test ($P<0.05$), and genes showing significant differences in phenotypic value between the different haplotypes were screened by comparing the phenotypic values of the different haplotypes (≥10 materials).

3 Results

3.1 Phenotypic variation

Firstly, the ML of 300 rice materials was evaluated under different overburden conditions (2, 5, 7cm). The results showed that the ML exhibited a continuous distribution with considerable genetic variation, indicating that this diverse population is an ideal material for GWAS studies (Figure 1). At a seeding depth of 2 cm, the ML ranged from 0 to 16.21 mm, with an average of 4.77 mm. At a seeding depth of 5 cm, the ML ranged from 0 to 33.90 mm, with an average of 5.61 mm. At a seeding depth of 7 cm, the ML ranged from 0 to 46.00 mm, with an average of 8.36 mm. In addition, we

10.3389/fpls.2025.1546580

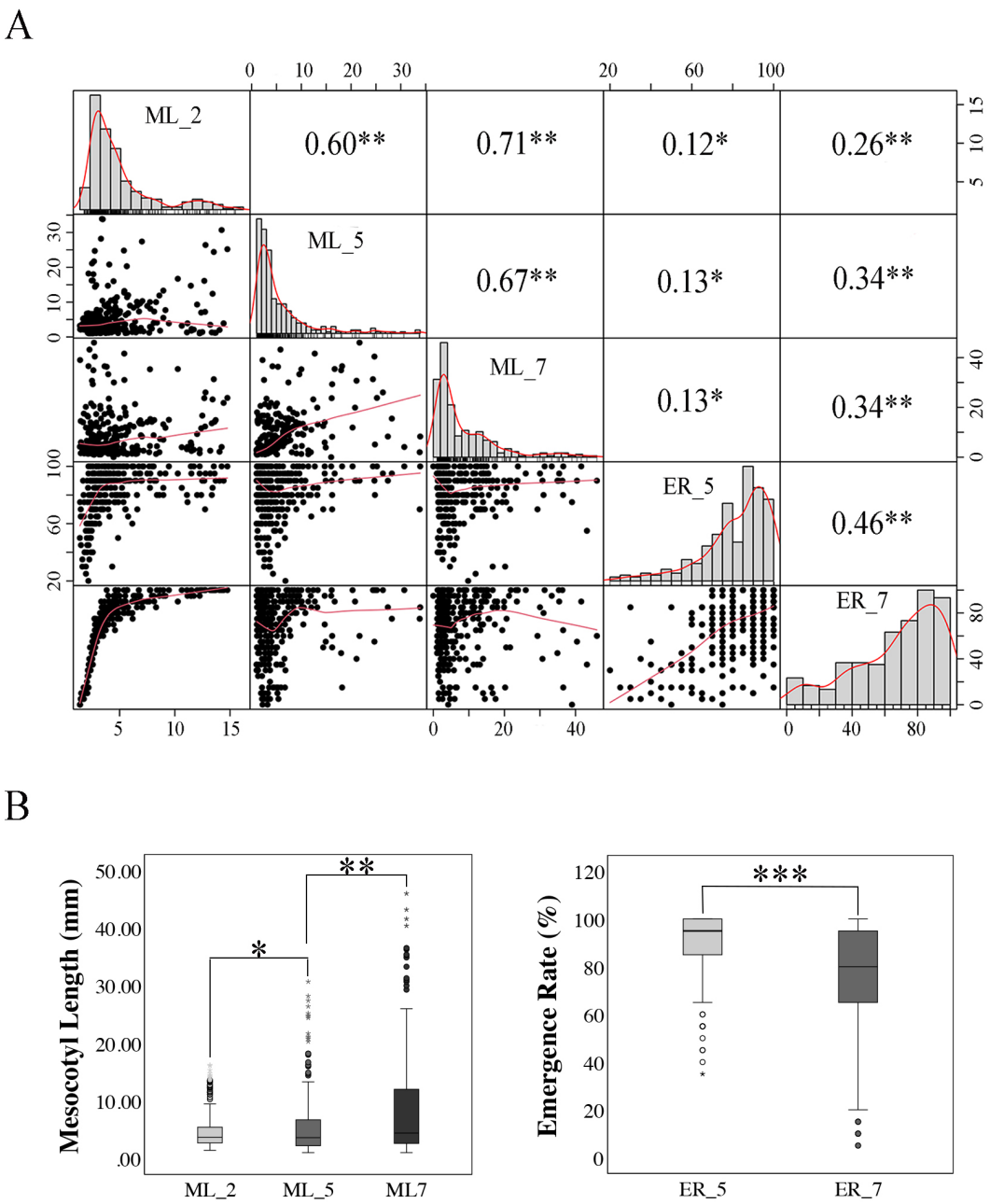


FIGURE 1
Correlation and differential analysis of mesocotyl length and emergence rate under three deep sowing depths. **(A)** Phenotypic distribution and correlation analysis of mesocotyl length and emergence rate under 2cm, 5 cm and 7 cm deep sowing conditions. Numbers in upper right corner are correlation coefficients (r), * indicates the significance at $p \leq 0.05$. **(B)** Phenotypic variations of phenotypic traits: mesocotyl length and emergence rate under 5 cm and 7 cm deep sowing conditions. Significance levels are indicated by asterisks: *P < 0.05, **P < 0.01, ***P < 0.001 (two-tailed *t-test*). ML_2, ML_5 and ML_7: mesocotyl length under deep sowing depths of 2cm, 5 cm and 7 cm; ER_5 and ER_7: emergence rate under deep sowing depths of 5 cm and 7 cm.

observed the ER. The results showed that under at 5 cm and 7 cm soil coverings, there were differences in the ERs among the different varieties. At a seeding depth of 5 cm, the ER ranged from 20% to 100%, with an average of 82.36%. At a seeding depth of 7 cm, the ER ranged from 5% to 100%, with an average of 68.88%.

Correlation analyses showed that ML exhibited a significant positive correlation among the three sowing depths: the correlation coefficients were 0.60 between 2 cm and 5 cm soil depth, 0.67 between 5 cm and 7 cm soil depth, and 0.71 between 2 cm and 7 cm soil depth. Additionally, the ERs at 5 cm and 7 cm sowing depths also exhibited a positive correlation, with a correlation coefficient of 0.46 (Figure 1A). In addition, independent samples t-test results indicated that there were significant differences in ML and ER under different sowing depths. A comparison of ML and ER under

different sowing depths revealed that the ML followed the order of 7 cm > 5 cm > 2 cm, while the ER was higher at 5 cm than at 7 cm ($P < 0.05$) (Figure 1B). This indicates that as the sowing depth increases, the ML also increases, but the seedling ER decreases.

3.2 Genome-wide association analysis and candidate gene identification

The sequencing data of the materials were input and filtered, yielding 711,268 high-quality SNP markers distributed across 12 chromosomes (Supplementary Figure S1). The PCA results indicated that the first 20 principal components accounted for a substantial proportion of the genetic variance, exhibiting a decreasing trend, whereas the 10th and subsequent components explained less than 2% of the genetic variance (Supplementary Figure S2). Therefore, in this study, the group structure was corrected by principal component analysis, using the first 10 principal components as covariates. Calculations were performed using the MLM.

Genome-wide linkage disequilibrium (LD) attenuation in cultivated rice has been reported to occur within a range of 100–200 kb (Huang et al., 2010). We defined the region 200 kb upstream and downstream of the significant SNP as the QTL region associated. A total of six unique loci were detected, which explained the phenotypic variations ranged from 6.96% to 8.48% (Table 1). For ML, two significant loci were detected on chromosomes 3 and 10 at a sowing depth of 2 cm, one significant locus was detected on chromosome 3 at a sowing depth of 5 cm, and 1 significant locus were detected on chromosome 6 at a sowing depth of 7 cm, collectively explaining 6.96% to 7.54% of the phenotypic variation (Figure 2, Table 1). For ER, one significant locus was detected on chromosomes 5 at a sowing depth of 5 cm, one significant locus was detected on chromosome 4 at a sowing depth of 7 cm, with explanation rates of 8.4% and 8.48%, respectively (Supplementary Figure S3; Table 1). Among them, *qML3* detected under the 2 cm sowing depth and *qML3* detected under the 5 cm sowing depth are located at the same position in the genome, representing co-localized loci. For ease of description, we will refer to them collectively as *qML3* in the following text.

3.3 Haplotype analysis of candidate genes

By analysing the functional variants within 200 kb upstream and downstream of the peak SNPs and their LD block characteristics (Figure 3A), we performed a preliminary screening of candidate genes for the qML3 locus. The genetic information of 300 rice varieties has been annotated with reference to the 7th version of the MSU Rice Genome Annotation Project (Rice 7) (http://rice.plantbiology.msu.edu/) (Kawahara et al., 2013) on the rice IRGSP-1.0 genome. There were 68 genes in the *qML3* candidate QTL interval, with 44 genes remaining after the removal of putative proteins, reduced transposons, and transposon-encoding genes (Supplementary Table S3). Many studies have reported the association between the activities of phytohormones such as abscisic acid (ABA), BR, SL, CK, ETH, JA, GA and IAA with ML. Therefore, we focused on genes encoding phytohormone-related genes within the specified intervals, resulting in the identification of 9 candidate genes (Table 2).

To identify potential candidate genes, we performed haplotype analysis of nine genes based on non-synonymous mutations in the coding region of each gene. Among these genes, 2 genes showed no significant phenotypic differences between their different haplotypes (Supplementary Figure S4), and 4 genes had only one haplotype. Of particular interest, 3 genes (*LOC_Os03g49250*, *LOC_Os03g49400* and *LOC_Os03g49510*) showed significant phenotypic differences ($P < 0.05$) between their different haplotypes (Figure 3). For *LOC_Os03g49250*, one major single-base variations in the CDS region that cause amino acid substitution. The lines carrying Hap_2^{LOC_Os03g49250} exhibited significantly longer ML than those carrying Hap_1^{LOC_Os03g49250} under the two sowing depth conditions ($P < 0.05$) (Figures 3B, E). For *LOC_Os03g49400*, there were eight major single-base variations in the CDS region that cause amino acid substitution. The lines carrying Hap_3^{LOC_Os03g49400} exhibited significantly longer ML than those carrying Hap_1^{LOC_Os03g49400} and Hap_2 $^{LOC\_Os03g49400}$ under the two sowing depth conditions ($P < 0.05$) (Figures 3C, E). For *LOC_Os03g49510*, there were three major single-base variation in its CDS region that caused amino acid substitution, and ML exhibited significantly different between the two haplotypes identified based on this variation. Under the two sowing depth

TABLE 1 Candidate QTLs identified from genome-wide association analysis.

Treatment	QTL	Chr	Position (Mb)	*P* value	R^2 (%)	Co-location QTL	Reference
Stress (2cm)	*qML3*	3	28.03-28.43	1.5059e-10	7.54	*qml3, qML3, qML3*	Cao et al., 2002; Niu et al., 2019; Huang et al., 2022
Stress (2cm)	*qML10*	10	3.29-3.69	9.9445e-10	6.96		
Stress (5cm)	*qML3*	3	28.03-28.43	5.6478e-08	7.19	*qml3, qML3, qML3*	Cao et al., 2002; Niu et al., 2019; Huang et al., 2022
Stress (7cm)	*qML6*	6	3.14-3.54	7.9207e-09	7.09		
Stress (5cm)	*qER5*	5	2.61-3.01	3.3156e-08	8.4		
Stress (7cm)	*qER4*	4	0-0.21	1.5602e-08	8.48		

ML, Mesocotyl length; ER, Emergence Rate; QTLs, quantitative trait loci; Chr, chromosome; R^2, phenotypic variation rate; Position, 200 kb upstream and downstream of the significant SNP locus used to define the QTL interval; Co-location QTLs, QTLs for shoot length identified by the previous studies.

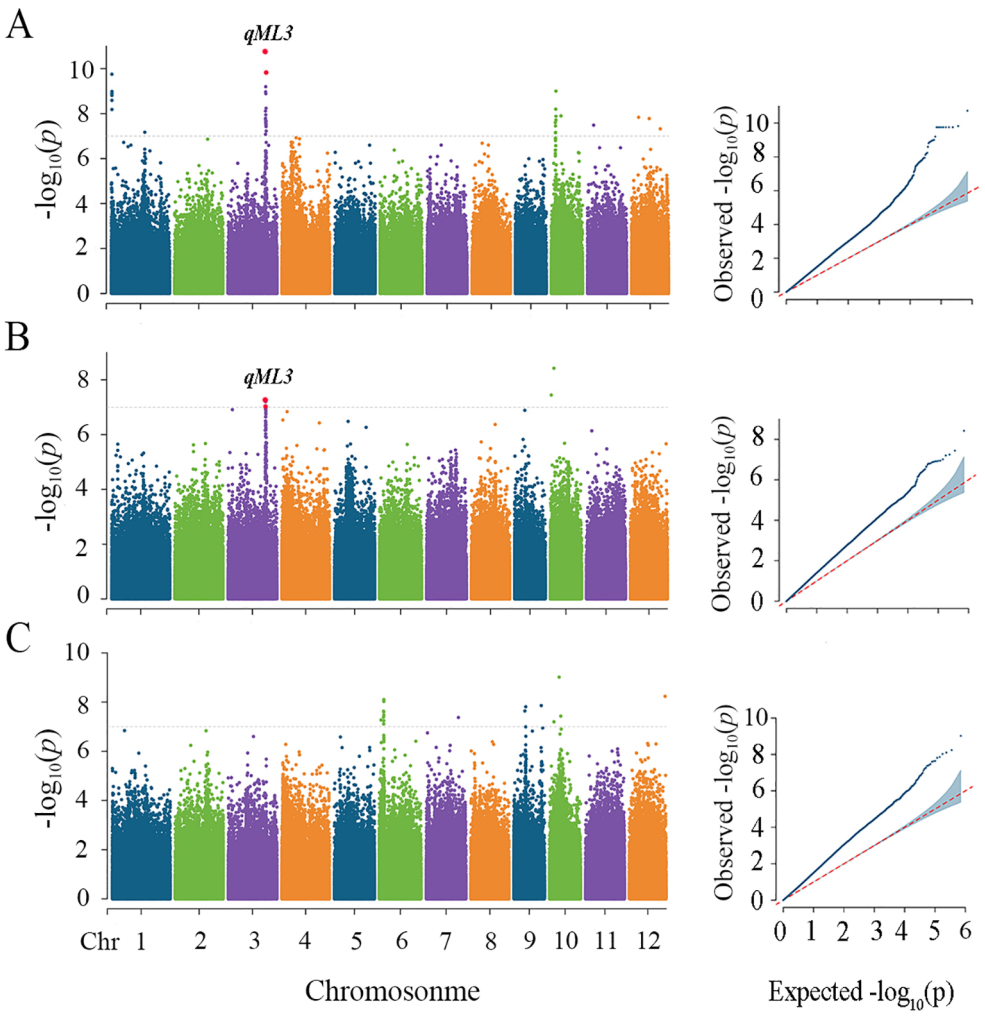


FIGURE 2
Genome-wide association studies for mesocotyl length among three deep sowing depths. Manhattan plots and Quantile-quantile plots of mesocotyl lengths in soil culture with a sowing depth of 2 cm **(A)**, 5cm **(B)** and 7cm **(C)**. Negative log10-transformed p values from a genome-wide scan are plotted against position on each of 12 chromosomes. The gray horizontal dashed lines indicate the genome-wide significance threshold. The red dots represent SNPs that were detected at both depths.

TABLE 2 Candidate genes for gene annotation.

Gene	Functional annotation
LOC_Os03g49050	possible lysine decarboxylase domain containing protein
LOC_Os03g49132	ZOS3-16 - C2H2 zinc finger protein
LOC_Os03g49170	zinc finger family protein
LOC_Os03g49250	OsFBO16 - F-box and other domain containing protein
LOC_Os03g49400	ethylene-insensitive protein
LOC_Os03g49480	elongation of fatty acids protein 2
LOC_Os03g49500	ethylene receptor
LOC_Os03g49510	phosphatidylinositol-4-phosphate 5-kinase
LOC_Os03g49620	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1

High-confidence candidate genes for mesocotyl lengths in *qML3*.

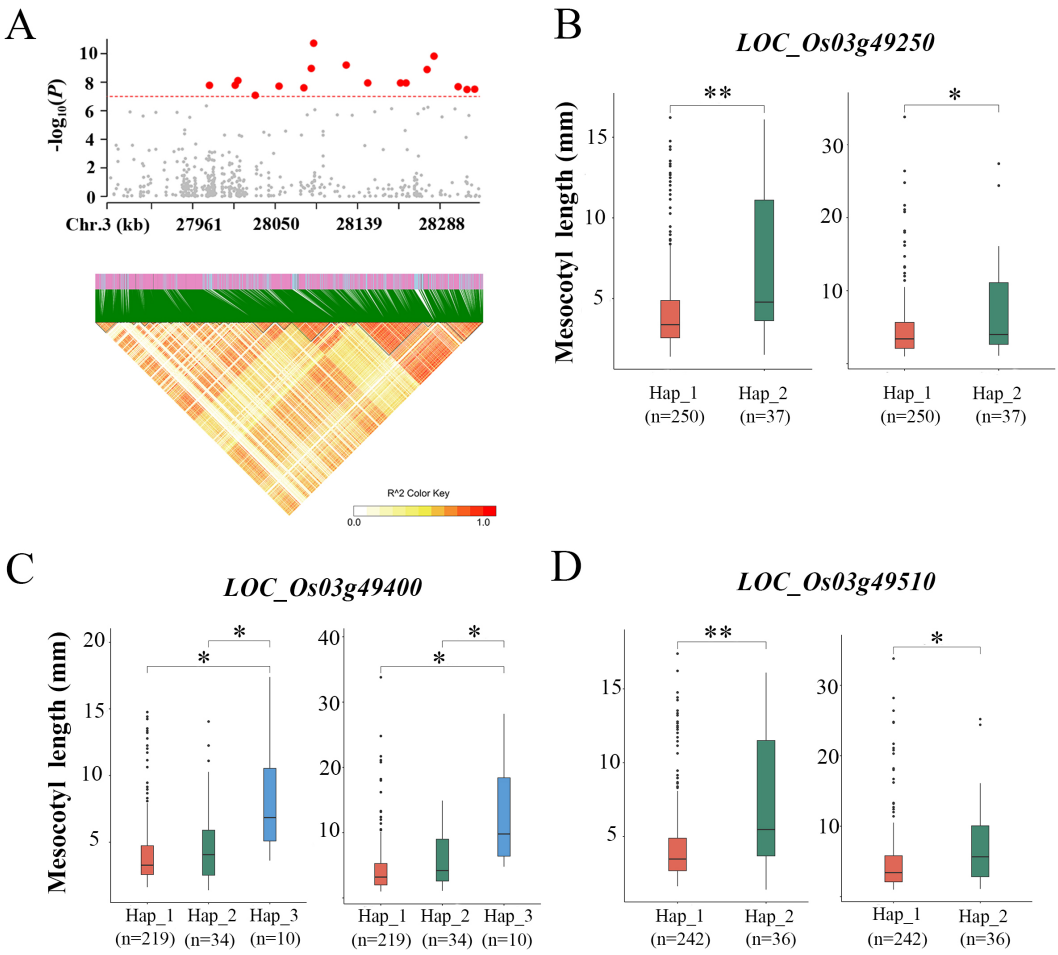
conditions, the lines carrying Hap_2 $^{LOC\_Os03g49510}$ exhibited significantly longer ML than those carrying Hap_1 $^{LOC\_Os03g49510}$ ($P < 0.05$) (Figures 3D, E).

4 Discussion

4.1 Sowing depth affects mesocotyl length and seedling emergence rate

Soil covering depth significantly influences the elongation of rice mesocotyl and its ER. Studies have shown that as sowing depth increases, ML exhibits a significant growth trend. This adaptive elongation mechanism contributes to improving rice ERs (Zhao et al., 2018; Lu et al., 2016). Research by Lee et al. (1999) demonstrated that at sowing depths of 5, 7, and 10 cm, ML significantly increased with depth, a finding consistent with the trend observed in our experiment at sowing depths of 2, 5, and 7

10.3389/fpls.2025.1546580



E

	LOC_Os03g49250			*LOC_Os03g49400*									*LOC_Os03g49510*				
Haplotype	28, 047, 397	ML_2 (mm)	ML_5 (mm)	28, 117, 565	28, 117, 597	28, 120, 272	28, 120, 306	28, 120, 405	28, 121, 513	28, 122, 331	28, 123, 654	ML_2 (mm)	ML_5 (mm)	28, 184, 475	28, 184, 506	28184516	ML_2 (mm)	ML_5 (mm)
Hap_1	T	4.38	4.78	A	T	C	G	A	T	T	A	4.24	4.54	C	G	T	4.52	5.05
Hap_2	C	6.68	7.5	G	C	C	T	G	G	A	G	4.91	6.08	A	C	G	6.91	7.54
Hap_3				A	T	A	G	A	T	T	A	8.91	12.96					

FIGURE 3
Gene structure and haplotype analysis of the candidate genes underlying *qML3*. **(A)** Local Manhattan plot (top) and linkage disequilibrium heatmap (bottom). **(B–D)** The boxplot illustrates the distribution of mesocotyl length across haplotype groups of *LOC_Os03g49250, LOC_Os03g49400* and *LOC_Os03g49510* under different sowing depths (2 cm, right; 5 cm, left). The middle line in each boxplot represents the median. $P < 0.05$. Significance levels are indicated by asterisks: *$P < 0.05$, **$P < 0.01$ (two-tailed t-test). **(E)** Haplotype analysis of the three genes in the region. ML_2 and ML_5: mesocotyl length under deep sowing depths of 2cm and 5cm.

cm. In addition, this pattern was also confirmed by Liu et al. (2019) who observed a significant increase in ML with increasing sowing depth at three mulch depths of 2 cm, 4 cm and 6 cm. In direct-seeded rice, deeper soil covering promotes mesocotyl elongation to facilitate seedling emergence. However, significant differences in mesocotyl elongation ability exist among different varieties: varieties with strong elongation ability can maintain a high ER under deep soil covering, while those with weak elongation ability may struggle to emerge, ultimately affecting yield. In this study, statistical analysis of seedling ERs under different sowing depths revealed that the ER

at a 7 cm sowing depth was significantly lower than that at a 5 cm depth. These results indicate that sowing depth is a critical factor influencing ML and ER. Therefore, selecting an appropriate sowing depth is crucial for the practical production of direct-seeded rice.

4.2 *qML3* exhibits significant potential value in rice direct seeding breeding

Previous studies have identified QTLs and candidate genes associated with mesocotyl elongation in rice using linkage mapping or GWAS techniques. In this study, through GWAS analysis, we identified four QTLs related to ML and two QTLs associated with ER. Among these, a significant co-localized locus for ML, named *qML3*, was detected on chromosome 3 under both 2 cm and 5 cm sowing depths. This locus overlaps with previously identified ML QTLs (Cao et al., 2002; Niu et al., 2019; Huang et al., 2022), underscoring the authenticity and reliability of *qML3*. Furthermore, *qML3* may be closely related to the adaptability of mesocotyl elongation in shallow soil conditions, providing important insights into the genetic regulatory mechanisms of ML in such environments. Regarding ER, we did not detect any significant loci co-localized with ML, suggesting that the genetic regulation of ER may be more complex or significantly influenced by other environmental factors.

4.3 Candidate mesocotyl length genes

Using GWAS, we initially identified 68 candidate genes within the *qML3* interval that might be related to mesocotyl elongation. From these, nine candidate genes with homology to known functional genes were screened by functional annotation analysis and combined with previous studies. Further haplotype analysis showed that the haplotype variants of three genes, *LOC_Os03g49250*, *LOC_Os03g49400* and *LOC_Os03g49510*, were significantly associated with mesocotyl elongation, suggesting that they may be potential candidate genes for *qML3*.

All three candidate genes screened in this study regulate mesocotyl and internode elongation through the phytohormone signalling pathway. *LOC_Os03g49250* encodes an F-box protein, a family of proteins that play important roles in plant signalling (Yan et al., 2017). Its homologous gene, D3, was shown to specifically recognise the OsGSK2-phosphorylated CYCLIN U2 protein and promote its degradation through the ubiquitin-proteasome pathway, thereby negatively regulating rice mesocotyl elongation (Sun et al., 2018). The ethylene-insensitive protein (EIN) encoded by *LOC_Os03g49400* is a core component of the ethylene signalling pathway, and it has been shown that ethylene affects mesocotyl development by regulating the expression of JA and BR related genes via EIN2/EIL2 (Xiong et al., 2017; Jun et al., 2004). *LOC_Os03g49510* encodes phosphatidylinositol 4-phosphate 5-kinase, a family of enzymes involved in growth hormone signalling and plant developmental processes through the regulation of phospholipid metabolism (Liang et al., 2018). It was found that OsPIP5K1 and DWT1/DWL2 synergistically regulate rice growth and development through phosphatidylinositol signalling, and its functional deficiency leads to disturbances in phospholipid metabolism, triggering typical phenotypes such as plant dwarfing and internode shortening (Fang et al., 2020).

During plant growth, the gradual increase in mesocotyl length is typically regulated by phytohormones through complex signaling pathways. Generally, phytohormones control mesocotyl elongation by modulating either cell division or cell expansion (Zhan et al., 2019). In this study, the three identified candidate genes were found to be associated with the biosynthesis and signalling pathways of plant hormones, suggesting their potential involvement in mesocotyl elongation by regulating the expression of hormone-related proteins. The molecular pathways by which the proteins encoded by these genes regulate mesocotyl elongation growth and the correlation (positive or negative) between changes in their expression levels (up- or down-regulation) and the rate of mesocotyl elongation need to be further investigated. In the following work, we will use molecular biology techniques such as gene editing, genetic transformation and DNA insertion to systematically verify the biological functions of these candidate genes in mesocotyl elongation.

5 Conclusions

This study utilized GWAS to analyze the ML and seedling ER of 300 rice germplasm resources. By analyzing two traits, ML and ER, a total of six QTLs were identified. Among them, the ML-related locus *qML3* was detected under two different sowing depths and contributed the most to phenotypic variation, indicating its stable genetic effects across multiple environmental conditions. Further prediction of candidate genes for *qML3* identified three potential candidate genes associated with ML: *LOC_Os03g49250*, *LOC_Os03g49400*, and *LOC_Os03g49510*. The functions of these genes require further validation, but their potential mechanisms provide new research directions for the regulation of mesocotyl elongation. In the future, these genes are expected to be applied in molecular breeding for ideal rice plant types, offering theoretical foundations and genetic resources for the improvement of direct-seeded rice.

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

LX: Investigation, Software, Writing – original draft. SW: Writing – original draft. QZ: Writing – review & editing. BH:

Writing – review & editing. DC: Writing – review & editing. LH: Writing – review & editing. JD: Writing – review & editing. XM: Conceptualization, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Biological Breeding-National Science and Technology Major Project (2022ZD04017), the National Crop Germplasm Resources Center (NCGRC-2023-02) and the Agricultural Science and Technology Innovation Program.

Conflict of interest

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

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1546580/full#supplementary-material

References

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Cai, H. W., and Morishima, H. (2002). QTL clusters reflect character associations in wild and cultivated rice. *Theor. Appl. Genet.* 104, 1217–1228. doi: 10.1007/s00122-001-0819-7

Cao, L. Y., Zhu, J., Yan, Q. C., He, L. B., Wei, X. H., and Cheng, S. H. (2002). Mapping QTLs with epistasis for mesocotyl length in a DH population from indica-japonica cross of rice (Oryza sativa). *Chin. J. Rice Sci.* 03), 24–27. doi: 10.16819/j.1001-7216.2002.03.005

Choi, D. (2003). Regulation of expansin gene expression affects growth and development in transgenic rice plants. *Plant Cell* 15, 1386–1398. doi: 10.1105/tpc.011965

Chung, N. J. (2010). Elongation habit of mesocotyls and coleoptiles in weedy rice with high emergence ability in direct-seeding on dry paddy fields. *Crop Pasture Sci.* 61, 911–917. doi: 10.1071/CP10099

Cui, D., Zhou, H., Ma, X., Lin, Z., Sun, L., Han, B., et al. (2022). Genomic insights on the contribution of introgressions from Xian/Indica to the genetic improvement of Geng/Japonica rice cultivars. *Plant communications* 3, 100325. doi: 10.1016/j.xplc.2022.100325

Fang, F., Ye, S., Tang, J., Bennett, M. J., and Liang, W. (2020). DWT1/DWL2 act together with OsPIP5K1 to regulate plant uniform growth in rice. *New Phytol.* 225, 1234–1246. doi: 10.1111/nph.16216

Gray, W. M., Ostin, A., Sandberg, G., Romano, C. P., and Estelle, M. (1998). High temperature promotes auxin-mediated hypocotyl elongation in Arabidopsis. *Proc. Natl. Acad. Sci. United States America* 95, 7197–7202. doi: 10.1073/pnas.95.12.7197

Huang, Q., Ju, C. Y., Cheng, Y. B., Cui, D., Han, B., Zhao, Z. W., et al. (2022). QTL mapping of mesocotyl elongation and confirmation of a QTL in dongxiang common wild rice in China. *Agronomy* 12, 1800. doi: 10.3390/agronomy12081800

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695

Ishfaq, M., Zulfiqar, U., Anjum, S. A., and Ahmad, M. (2018). Optimizing row spacing for direct seeded aerobic rice under dry and moist fields. *Pakistan J. Agric. Res.* 31, 291–299. doi: 10.17582/journal.pjar/2018/31.4.291.299

Jun, S. H., Han, M. J., and Lee, S. (2004). OsEIN2 is a positive component in ethylene signaling in rice. *Plant Cell Physiol.* 45, 281–289. doi: 10.1093/pcp/pch033

Kato, Y., and Katsura, K. (2014). Rice adaptation to aerobic soils: Physiological considerations and implications for agronomy. *Plant Production Sci.* 17, 1–12. doi: 10.1626/pps.17.1

Kawahara, Y., Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (New York NY)* 6, 4. doi: 10.1186/1939-8433-6-4

Kwon, S. W. (2021). Genome-wide association study (GWAS) of mesocotyl length for direct seeding in Rice. *Agronomy* 11, 1–17. doi: 10.3390/agronomy11122527

Lee, S. S., Kim, J. H., and Hong, S. B. (1999). Effects of priming and growth regulator treatment of seed on emergence and seedling growth of rice. *Korean J. Crop Sci.* 44, 134–137. Available at: https://koreascience.kr/article/JAKO199911922913759.page (Accessed May 3, 2025)

Lee, H. S., Sasaki, K., Higashitani, A., Ahn, S. N., and Sato, T. (2011). Mapping and characterization of quantitative trait loci for mesocotyl elongation in rice (*Oryza sativa* L.). *Rice* 5, 13. doi: 10.1186/1939-8433-5-13

Lee, H. S., Sasaki, K., Kang, J. W., Sato, T., Song, W. Y., and Ahn, S. N. (2017). Mesocotyl elongation is essential for seedling emergence under deep-seeding condition in rice. *Rice* 10, 32. doi: 10.1186/s12284-017-0173-2

Li, L., Liang, Q., Wang, C., and Cheng, W. F. (2013). Effect of auxin on weedy rice mesocotyl cell wall oxidase. *Appl. Mechanics Materials* 448-453, 69–73. doi: 10.4028/www.scientific.net/AMM.448-453

Li, L., Ma, D. R., Sun, J., Liang, Q., and Cheng, W. F. (2012). Observation of mesocotyl cell morphology of weed rice. *J. Shenyang Agric. Univ.* 43, 749–753. Available at: http://en.cnki.com.cn/Article_en/CJFDTOTAL-SYNY201206020.htm (Accessed May 3, 2025).

Liang, T. N., Zhao, Y., He, Z. B., Cao, Q. G., Zhang, Y. J., Zhang, Y. X., et al. (2018). Bioinformatics analysis and fluorescence quantitative PCR results of Ricinus communis PIP5Ks. *Chin. Traditional Herbal Drugs* 49, 5892–5900. Available at: http://en.cnki.com.cn/Article_en/CJFDTotal-ZCYO201824025.htm (Accessed May 3, 2025).

Liu, J. D., Wang, Y. M., and Liu, H. Y. (2023). Genome wide association study of candidate genes underlying rice mesocotyl elongation. *J. Plant Genet. Resour.* 24, 1702–1735. doi: 10.13430/j.cnki.jpgr.20230326001

Liu, H., Zhan, J., Li, J., Lu, X., Liu, J., Wang, Y., et al. (2019). Genome-wide Association Study (GWAS) for Mesocotyl Elongation in Rice (*Oryza sativa* L.) under Multiple Culture Conditions. *Genes* 11, 49. doi: 10.3390/genes11010049

Lu, Q., Zhang, M., Niu, X., Wang, C., Xu, Q., Feng, Y., et al. (2016). Uncovering novel loci for mesocotyl elongation and shoot length in indica rice through genome-wide association mapping. *Planta* 243, 645–657. doi: 10.1007/s00425-015-2434-x

Lv, Y. S., Shao, G. N., Jiao, G. A., Sheng, Z. H., Xie, L. H., Hu, S. K., et al. (2021). Targeted mutagenesis of POLYAMINE OXIDASE 5 that negatively regulates mesocotyl elongation enables the generation of direct-seeding rice with improved grain yield. *Mol. Plant* 14, 344–351. doi: 10.1016/j.molp.2020.11.007

Ma, B., He, S. J., Duan, K. X., Yin, C. C., Chen, H., Yang, C., et al. (2013). Identification of rice ethylene-response mutants and characterization of MHZ7/OsEIN2 in distinct ethylene response and yield trait regulation. *Mol. Plant* 6, 1830–1848. doi: 10.1093/mp/sst087

Mahender, A., Anandan, A., and Pradhan, S. K. (2015). Early seedling vigour, an imperative trait for direct-seeded rice: an overview on physio-morphological parameters and molecular markers. *Planta* 241, 1027–1050. doi: 10.1007/s00425-015-2273-9

Meng, Y., Zhan, J. H., Liu, H. Y., Liu, J. D., Wang, Y. M., Guo, Z., et al. (2023). Natural variation of OsML1, a mitochondrial transcription termination factor, contributes to mesocotyl length variation in rice. *Plant Journal: Cell Mol. Biol.* 115, 910–925. doi: 10.1111/tpj.16267

Mgonja, M. A., Ladeinde, T. A. O., and Aken'Ova, M. E. (1993). Genetic analysis of mesocotyl length and its relationship with other agronomic characters in rice (*Oryza sativa* L.). *Euphytica* 72, 189–195. doi: 10.1007/BF00034157

Muhammad, I., Nadeem, A., Shakeel, A. A., and Muhammad, A. I. H. (2020). Growth, yield and water productivity of dry direct seeded rice and transplanted aromatic rice under different irrigation management regimes. *J. Integr. Agric.* 19 (11), 2656–2673. doi: 10.1016/S2095-3119(19)62876-5

Niu, S. P., Lv, Y. S., Wu, Y. W., Wei, X. J., Sheng, Z. H., Jiang, G. A., et al. (2019). QTLs mapping for mesocotyl length in rice. *China Rice* 25, 55–59. Available at: https://link.cnki.net/urlid/33.1201.s.20191112.0913.022 (Accessed May 3, 2025).

Ohno, H., Banayo Nino, P. M. C., Bueno Crisanta, S., Kashiwagi, J.-i., Nakashima, T., Corales Aurora, M., et al. (2018). Longer mesocotyl contributes to quick seedling establishment, improved root anchorage, and early vigor of deep-sown rice. *Field Crops Res.* 228, 84–92. doi: 10.1016/j.fcr.2018.08.015

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Sagare, D. B., Abbai, R., Jain, A., Jayadevappa, P. K., Dixit, S., Singh, A. K., et al. (2020). More and more of less and less: Is genomics-based breeding of dry direct-seeded rice (DDSR) varieties the need of hour? *Plant Biotechnol. J.* 18, 2173–2186. doi: 10.1111/pbi.13454

Schillinger, W. F., Donaldson, E., Allan, R. E., and Jones, S. S. (1998). Winter wheat seedling emergence from deep sowing depths. *Agron. J.* 90, 582–586. doi: 10.2134/agronj1998.00021962009000050002x

Sun, S., Wang, T., Wang, L., Li, X., Jia, Y., Liu, C., et al. (2018). Natural selection of a GSK3 determines rice mesocotyl domestication by coordinating strigolactone and brassinosteroid signaling. *Nat. Communication* 9, 2523. doi: 10.1038/s41467-018-04952-9

Turner, F. T., Chen, C. C., and Bollich, C. N. (1982). Coleoptile and mesocotyl lengths in semi-dwarf rice seedlings. *Crop Sci.* 22, 43–46. doi: 10.2135/cropsci1982.0011183X002200010010x

Wang, Y., Liu, J., Meng, Y., Liu, H., Liu, C., and Ye, G. (2021). Rapid identification of QTL for mesocotyl length in rice through combining QTL-seq and genome-wide association analysis. *Front. Genet.* 12. doi: 10.3389/fgene.2021.713446

Wang, Y., Liu, H., Meng, Y., Liu, J., and Ye, G. (2023). Validation of genes affecting rice mesocotyl length through candidate association analysis and identification of the superior haplotypes. *Front. Plant sci.* 14. doi: 10.3389/fpls.2023.1194119

Wu, J., Feng, F., Lian, X., Teng, X., Wei, H., Yu, H., et al. (2015). Genome-wide Association Study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. *BMC Plant Biol.* 15, 1–10. doi: 10.1186/s12870-015-0608-0

Wu, M. G., Zhang, G. H., Lin, J. R., and Cheng, S. H. (2005). Screening for rice germplasms with specially elongated mesocotyl. *Rice Sci.* 12, 226–228. Available at: http://www.ricesci.org/EN/Y2005/V12/I3/226 (Accessed May 3, 2025)

Xiong, Q., Ma, B., Lu, X., Huang, Y. H., He, S. J., Yang, C., et al. (2017). Ethylene-inhibited jasmonic acid biosynthesis promotes mesocotyl/coleoptile elongation of etiolated rice seedlings. *Plant Cell.* 29, 1053–1072. doi: 10.1105/tpc.16.00981

Yan, J., Liao, X., He, R., Zhong, M., Feng, P., Li, X., et al. (2017). Ectopic expression of GA 2-oxidase 6 from rapeseed (Brassica napus L.) causes dwarfism, late flowering and enhanced chlorophyll accumulation in Arabidopsis thaliana. *Plant Physiol. Biochem.* 111, 10–19. doi: 10.1016/j.plaphy.2016.11.008

Yeo, I. K., and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika* 87, 954–959. doi: 10.1093/biomet/87.4.954

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom Proteom Bioinform* 19, 10. doi: 10.1016/j.gpb.2020.10.007

Zhan, J., Lu, X., Liu, H., Zhao, Q., and Ye, G. (2019). Mesocotyl elongation, an essential trait for dry-seeded rice (*Oryza sativa* l.): a review of physiological and genetic basis. *Planta* 251, 27. doi: 10.1007/s00425-019-03322-z

Zhang, X. J., Lai, Y. C., Meng, Y., Tang, A., Dong, W. J., Liu, Y. H., et al. (2023). Analyses and identifications of quantitative trait loci and candidate genes controlling mesocotyl elongation in rice. *J. Integr. Agric.* 22, 325–340. doi: 10.1016/j.jia.2022.08.080

Zhang, H., Ma, P., Zhao, Z., Zhao, G., Tian, B., Wang, J., et al. (2012). Mapping QTL controlling maize deep-seeding tolerance-related traits and confirmation of a major QTL for mesocotyl length. *Theor. Appl. Genet.* 124, 223–232. doi: 10.1007/s00122-011-1700-y

Zhao, H., Li, J. C., Yang, L., Qin, G., Xia, C. J., Xu, X. B., et al. (2021). An inferred functional impact map of genetic variants in rice. *Mol. Plant* 14, 1584–1599. doi: 10.1016/j.molp.2021.06.025

Zhao, Y., Zhao, W. P., Jiang, C. H., Wang, X. N., Xiong, H. Y., Todorovska, E. G., et al. (2018). Genetic architecture and candidate genes for deep-sowing tolerance in rice revealed by non-syn GWAS. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00332

Zheng, J., Hong, K., Zeng, L., Wang, L., Kang, S., Qu, M., et al. (2020). Karrikin signaling acts parallel to and additively with Strigolactone signaling to regulate rice mesocotyl elongation in darkness. *Plant Cell* 32, 2780–2805. doi: 10.1105/tpc.20.00123

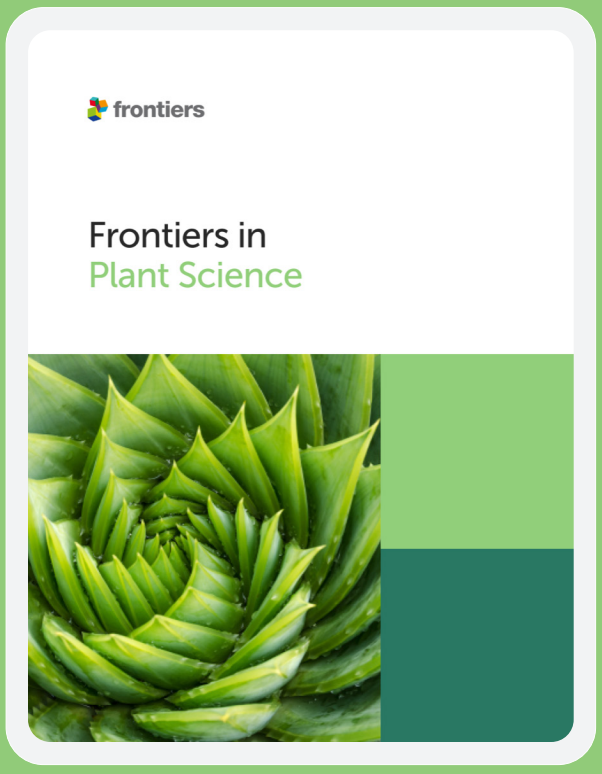
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact