

Emerging genomic technologies for agricultural biotechnology: Current trends and future prospects

Edited by

Md. Anowar Hossain and Hairul Roslan

Published in

Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

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

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

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

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

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

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

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

ISSN 1664-8714
ISBN 978-2-8325-3439-7
DOI 10.3389/978-2-8325-3439-7

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

Emerging genomic technologies for agricultural biotechnology: Current trends and future prospects

Topic editors

Md. Anowar Hossain — University of Rajshahi, Bangladesh

Hairul Roslan — University of Malaysia Sarawak, Malaysia

Citation

Hossain, M. A., Roslan, H., eds. (2023). *Emerging genomic technologies for agricultural biotechnology: Current trends and future prospects*.

Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3439-7

Table of contents

- 04 **Editorial: Emerging genomic technologies for agricultural biotechnology: current trends and future prospects**
Md. Anowar Hossain and Hairul Azman Roslan
- 07 **Combining genomic selection with genome-wide association analysis identified a large-effect QTL and improved selection for red rot resistance in sugarcane**
Anthony O'Connell, Jasmin Deo, Emily Deomano, Xianming Wei, Phillip Jackson, Karen S. Aitken, Ramaswamy Manimekalai, Krishnasamy Mohanraj, Govinda Hemaprabha, Bakshi Ram, Rasappa Viswanathan and Prakash Lakshmanan
- 23 **Genetically engineered crops for sustainably enhanced food production systems**
Mughair Abdul Aziz, Faical Brini, Hatem Rouached and Khaled Masmoudi
- 47 **Upcoming progress of transcriptomics studies on plants: An overview**
Parul Tyagi, Deeksha Singh, Shivangi Mathur, Ayushi Singh and Rajiv Ranjan
- 62 **Primary mapping of quantitative trait loci regulating multivariate horticultural phenotypes of watermelon (*Citrullus lanatus* L.)**
Sikandar Amanullah, Shenglong Li, Benjamin Agyei Osae, Tiantian Yang, Farhat Abbas, Meiling Gao, Xuezheng Wang, Hongyu Liu, Peng Gao and Feishi Luan
- 91 **Transcriptional dynamics of maize leaves, pollens and ovules to gain insights into heat stress-related responses**
Ashok Babadev Jagtap, Inderjit Singh Yadav, Yogesh Vikal, Umesh Preethi Praba, Navneet Kaur, Adeshpal Singh Gill and Gurmukh S. Johal
- 106 **Transcriptomic profiling reveals candidate allelopathic genes in rice responsible for interactions with barnyardgrass**
Most. Humaira Sultana, Md. Alamin, Jie Qiu, Longjiang Fan and Chuyu Ye
- 121 **Evaluate the guide RNA effectiveness via *Agrobacterium*-mediated transient assays in *Nicotiana benthamiana***
Zhibo Wang, Zachary Shea, Qi Li, Kunru Wang, Kerri Mills, Bo Zhang and Bingyu Zhao
- 130 **Identification of major QTLs for yield-related traits with improved genetic map in wheat**
Feifei Ma, Yunfeng Xu, Ruifang Wang, Yiping Tong, Aimin Zhang, Dongcheng Liu and Diaoguo An
- 145 **Development of new mutant alleles and markers for *KT11* and *KT13* via CRISPR/Cas9-mediated mutagenesis to reduce trypsin inhibitor content and activity in soybean seeds**
Zhibo Wang, Zachary Shea, Luciana Rosso, Chao Shang, Jianyong Li, Patrick Bewick, Qi Li, Bingyu Zhao and Bo Zhang



OPEN ACCESS

EDITED AND REVIEWED BY
Gulnihal Ozbay,
Delaware State University, United States

*CORRESPONDENCE
Md. Anwar Hossain
✉ mahossain95@hotmail.com

RECEIVED 19 July 2023
ACCEPTED 18 August 2023
PUBLISHED 24 August 2023

CITATION
Hossain MA and Roslan HA (2023) Editorial:
Emerging genomic technologies for
agricultural biotechnology: current trends
and future prospects.
Front. Plant Sci. 14:1263289.
doi: 10.3389/fpls.2023.1263289

COPYRIGHT
© 2023 Hossain and Roslan. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Editorial: Emerging genomic technologies for agricultural biotechnology: current trends and future prospects

Md. Anwar Hossain^{1*} and Hairul Azman Roslan²

¹Department of Biochemistry and Molecular Biology, University of Rajshahi, Rajshahi, Bangladesh,

²Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

KEYWORDS

CRISPR/Cas9, RNA-sequencing, quantitative trait loci, genome editing, marker assisted selection (MAS)

Editorial on the Research Topic

Emerging genomic technologies for agricultural biotechnology: current trends and future prospects

The current earth's population of 7.6 billion is expected to reach 8.6 billion by 2030. The increased population will need more food than it can currently produce. However, world agriculture is facing severe challenges such as global climate change, exhausted resources, reduction of arable lands and various pathogens. Advances in genomic technologies may offer potential solutions to these agricultural problems. New genomic technologies such as, Next generation sequencing (NGS), Ribonucleic acid sequencing (RNA-Seq), Clustered Regularly Interspaced Short Palindromic Repeat-Cas9 (CRISPR/Cas9), Transcription activator-like effector nucleases (TALENs) and Oligonucleotide-directed mutagenesis (ODM) as well as doubled haploids, molecular markers and mapping populations have been developed and utilized for increasing the crop production. Together with the rapidly expanding availability of genome sequence data, these technologies have the potential to transform plant breeding.

In this Research Topic we aim at the collection of articles on application of cutting-edge genomic technologies to improve various crops, vegetables and fruits which includes two reviews and seven original articles.

In first review, [Aziz et al.](#) explored the deployment of GM crops and their effects on sustainable food production systems which provided a comprehensive overview of the cultivation of GM crops and the issues preventing their widespread adoption, with appropriate strategies to overcome them. They also presented new tools for genome editing technology with special reference of CRISPR/Cas9 platform. They outlined the role of crops developed through CRISPR/Cas9 for sustainable development goals by 2030.

Nowadays in the post genomic era, transcriptomic or RNA-Seq technology focuses on functional studies of transcriptomes. It is a high-resolution, sensitive and high-throughput NGS approach used to study non-model plants and other organisms. RNA-Seq is an important technique for predictions and functional analysis of genes that improves gene ontology biological processes, molecular functions, and cellular components, but still now

there is limited information available on this topic. Tyagi et al. nicely presented the recent genomics technologies such as first generation, second generation (Next generation) and third generation for functional genomic studies. Their review article focused currently used technology in details and forthcoming strategies for improving transcriptome sequencing technologies for identifying the functional of genes in various plants using RNA-Seq technology, based on the principles, development, and applications.

In the first article, O'Connell et al. used genome wide association studies and genomic selection to identify the Single Nucleotide Polymorphism (SNP) markers for red rot resistance. *Colletotrichum falcatum* is a causal agent of red rot disease of sugarcane worldwide including India. They identified a single 14.6 Mb genomic region in Sorghum which in turned helps to map the Quantitative Trait Loci (QTLs). Finally genetic analysis nearby the SNP markers linked to the QTL revealed many biotic stress responsive genes within this QTLs, including a cluster of four chitinase A genes.

The primary mapping of QTLs is a molecular technique that has been widely used technique for the identification of candidate genomic regions controlling various crop traits based on different types of genetic markers. This article, Amanullah et al. performed the whole Genome Sequencing (WGS) for two watermelon varieties. They identified the 33 QTLs at different genetic positions across the eight chromosomes of watermelon for the ovary, fruit, and seed related phenotypes. Twenty-four QTLs were mapped as major-effect and 9 QTLs were identified as minor-effect QTLs across the flanking regions of Cleaved Amplified Polymorphic Sequences (CAPS) markers. We hoped that their detected QTLs provide critical genetic information concerning the watermelon phenotypes and fine genetic mapping could be followed to confirm them.

In the next article, Sultana et al. successfully identified 5,684 differentially genes expressed genes (DEGs) from transcriptome data that control the allelopathic interaction between banyardgrass and rice. Among them, 388 genes were transcription factors and associated with momilactone and phenolic acid biosynthesis which could play critical roles in allelopathy. Moreover, they identified the up-regulated and down-regulated genes at certain condition, which could be involved in the secondary metabolites' biosynthesis and developmental processes; respectively. They also detected some common DEGs between rice and banyardgrass which suggested different mechanisms underlying allelopathic interaction in these two species. Authors believed that their study could provide a valuable genetic resource for rice and banyardgrass associated candidate allelopathy genes and should be useful for controlling weeds, which would outcome in the development of agriculture.

In this article, to overcome the variable efficiency of the guide RNA (gRNA) in the CRISPR/Cas mediated genome editing system for the improvement of agronomic traits of many crop plant species, Wang et al. used the *Agrobacterium*-mediated transient assays to evaluate the effectiveness of gRNAs for editing genes in tobacco and soybean. They established a new transient assay system to validate the effectiveness of gRNAs before generating stable transgenic plants.

Trypsin inhibitor (TI) present in the soybean seeds is an anti-nutritional factor that hampers the digestibility of soybean meal in the human intestine tract. To develop the low TI soybeans, in their next article, Wang et al. identified two seed-specific TI genes such as

KTI1 and KTI3. They developed TI-mutant lines of soybeans through CRISPR/Cas9 mediated genome editing approach and also detected co-dominant selection markers for mutant alleles. Authors expected that the kti1/3 mutant soybean line and associated selection markers would help in accelerating the introduction of low TI trait into elite soybean cultivars in the future.

To discover the heat stress responsive genes of maize, Jagtap et al. employed the transcriptomic studies of two inbred lines, M 11 (sensitive to heat shock) and CML 25 (tolerant to heat shock), under intense heat stress at 42°C during reproductive stage from flag leaf, tassel, and ovule. They identified 1127 up-regulated and 1037 down-regulated differentially expressed genes (DEGs) with 1151, 451, and 562 DEGs in comparisons of LM 11 and CML 25, corresponding to a leaf, pollen, and ovule, respectively. Functional analysis revealed that DEGs associated with transcription factors (TFs), heat shock proteins (HSP20, HSP70, and HSP101/ClpB), as well as genes related to photosynthesis (PsaD & PsaN), antioxidation (APX and CAT) and polyamines (Spd and Spm). Their gene expression analysis found higher expression of HS-responsive genes in CML 25, which might explain why CML 25 is more heat tolerant in heat stress.

Due to the large genome size, hexaploidy nature and high percentage of repetitive regions of wheat, it is difficult to map yield-related QTLs. Ma et al. conducted high-density SNP arrays technology to construct genetic map and identify twelve environmentally stable QTLs for yield-related traits in wheat. Among them QTKw-1B.2 and QPh-4B.1 could be novel QTLs which could be used for wheat breeding programs. We hope that researchers and scientists working on crop improvement for sustainable food production will be benefited from this Research Topic in future.

The articles in this Research Topic provide a fascinating overview of the latest genomic technologies for agricultural biotechnology. These technologies, described in the articles, provide novel solutions to global agriculture concerns such as climate change, resource scarcity, and pathogen outbreaks. The papers discussed a variety of genomic technology applications, from developing GM crops to identifying QTLs and genes that respond to heat stress in different crops. The findings can help improve crop output, build sustainable food systems, and increase the nutritional content of agricultural products. We can expect significant advancements in plant breeding and agricultural practices as genomic technologies continue to advance and researchers and scientists collaborate, ultimately contributing to global food security and the achievement of sustainable development goals.

Author contributions

MH: Conceptualization, Data curation, Formal Analysis, Investigation, Supervision, Writing – original draft. HR: Writing – review & editing.

Acknowledgments

We acknowledge the contribution of all authors from all articles published in this Research Topic. We are also grateful to the

reviewers who spent their valuable time to critically review the manuscripts.

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.



OPEN ACCESS

EDITED BY

Hairul Roslan,
Universiti Malaysia Sarawak, Malaysia

REVIEWED BY

Prabina Kumar Meher,
Indian Agricultural Statistics Research
Institute (ICAR), India
Karansher Singh Sandhu,
Bayer Crop Science, United States

*CORRESPONDENCE

Prakash Lakshmanan
plakshmanan2018@outlook.com
Phillip Jackson
philjackson166@icloud.com

SPECIALTY SECTION

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

RECEIVED 17 August 2022

ACCEPTED 29 September 2022

PUBLISHED 31 October 2022

CITATION

O'Connell A, Deo J, Deomano E,
Wei X, Jackson P, Aitken KS,
Manimekalai R, Mohanraj K,
Hemaprabha G, Ram B, Viswanathan R
and Lakshmanan P (2022) Combining
genomic selection with genome-wide
association analysis identified a large-
effect QTL and improved selection for
red rot resistance in sugarcane.
Front. Plant Sci. 13:1021182.
doi: 10.3389/fpls.2022.1021182

COPYRIGHT

© 2022 O'Connell, Deo, Deomano, Wei,
Jackson, Aitken, Manimekalai, Mohanraj,
Hemaprabha, Ram, Viswanathan and
Lakshmanan. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Combining genomic selection with genome-wide association analysis identified a large-effect QTL and improved selection for red rot resistance in sugarcane

Anthony O'Connell¹, Jasmin Deo¹, Emily Deomano¹,
Xianming Wei¹, Phillip Jackson^{2*}, Karen S. Aitken²,
Ramaswamy Manimekalai³, Krishnasamy Mohanraj³,
Govinda Hemaprabha³, Bakshi Ram³, Rasappa Viswanathan³
and Prakash Lakshmanan^{1,4,5,6*}

¹Sugar Research Australia Limited, Brisbane, QLD, Australia, ²Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food, Queensland Bioscience Precinct, St Lucia, QLD, Australia, ³Sugarcane Breeding Institute, Coimbatore, India, ⁴Sugarcane Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China, ⁵Interdisciplinary Research Center for Agriculture Green Development in Yangtze River Basin (CAGD), College of Resources and Environment, Southwest University, Chongqing, China, ⁶Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, QLD, Australia

Red rot caused by the fungus *Colletotrichum falcatum* is the main disease limiting sugarcane productivity in several countries including the major producer India. The genetic basis for red rot resistance is unclear. We studied a panel of 305 sugarcane clones from the Australian breeding program for disease response phenotype and genotype using an Affymetrix® Axiom® array, to better understand the genetic basis of red rot resistance. SNP markers highly significantly associated with red rot response ($\leq 10^{-8}$) were identified. Markers with largest effect were located in a single 14.6 Mb genomic region of sorghum (the closest diploid relative of sugarcane with a sequenced genome) suggesting the presence of a major-effect QTL. By genomic selection, the estimated selection accuracy was ~0.42 for red rot resistance. This was increased to ~0.5 with the addition of 29 highly significant SNPs as fixed effects. Analysis of genes nearby the markers linked to the QTL revealed many biotic stress responsive genes within this QTL, with the most significant SNP co-locating with a cluster of four chitinase A genes. The SNP markers identified here could be used to predict red rot resistance with high accuracy at any stage in the sugarcane breeding program.

KEYWORDS

sugarcane, GWAS, genome wide association study, red rot, molecular breeding, genomic selection

Introduction

Cultivated sugarcane (*Saccharum* spp. inter-specific hybrids) is a major food and industrial crop grown in more than 110 countries in the tropics and sub-tropics (FAO, 2020). Globally, it is the fifth most valuable crop economically, providing >80% of the sugar and ~35% of the bioethanol in the world. Brazil and India combined account for more than 50% of sugarcane production in the world. Like other crops, maintaining resistance to important diseases is a major objective of sugarcane breeding programs worldwide (Heinz, 1987; Jackson, 2018). Sugarcane red rot disease caused by the fungus *Colletotrichum falcatum* Went has been reported in 77 countries and is the most damaging sugarcane disease in India, Pakistan, Thailand, Nepal, Myanmar and Vietnam (Viswanathan et al., 2018). The disease causes rotting of sugarcane stalk tissue, affecting cane yield and sugar quality. Inversion of stored sucrose by the pathogen affects sugar juice quality, causing reduced sugar recovery in sugar mills. In India, large-scale red rot epidemics have occurred every decade since its first appearance in 1901, resulting in large economic loss and the removal of highly productive and widely cultivated varieties from production (Viswanathan et al., 2018). By contrast in some other countries including Australia, red rot disease is present but is observed rarely and has only a very small impact on commercial cane production. Resolving the underlying reasons for the differing impact between countries is of interest and potential practical importance.

Selection and deployment of resistant varieties is the most common strategy used to manage red rot in affected sugarcane industries (Viswanathan et al., 2018). In India no variety is released for commercial production unless it has resistance to red rot. However, breakdown of red rot resistance is common (Viswanathan, 2021). In India, Co 205, the first hybrid sugarcane cultivar (i.e. first cultivar with a *Saccharum spontaneum* ancestor) that was released in 1918, and which quickly became dominant in northern India, succumbed to red rot within a few years after release (Chona, 1980). Over the next 100 years, nearly all sugarcane cultivars in India, which were resistant at the time of commercial release, became susceptible and succumbed to the disease within a period of 3 to 20 years following release (Chona, 1980; Viswanathan, 2021). Unlike other sugarcane pathogens, new *C. falcatum* pathotypes with varying degrees of virulence are frequently formed through mutations and parasexual recombination, causing resistance breakdown (Viswanathan et al., 2020). Gain and loss of virulence and occasional emergence of super-virulent pathotypes have been reported (Viswanathan et al., 2020). Although the underlying mechanism for host-resistance breakdown is unclear, development of new *C. falcatum* pathotypes and their adaptation to new varieties contributes to resistance breakdowns and disease epidemics (Viswanathan et al., 2021).

Studies on inheritance of red rot resistance have reported a moderate to high narrow-sense heritability and high broad sense heritability (Babu et al., 2010; Alarmelu et al., 2010). This indicates that both additive genetic variance (i.e. variation due to presence or absence of alleles) and non-additive (i.e. dominance variation due to combinations of alleles at particular loci, or epistasis variation due to interactions between alleles at different loci) are important. The high values (>0.90) reported for broad-sense also indicates potentially stronger genetic control of response to the disease compared with environmental factors (Ram et al., 2006). A major source of resistance in sugarcane cultivars is believed to be derived from *S. spontaneum* ancestors (Natarajan et al., 2001). It is also believed that a combination of vertical resistance (due to race specific large gene effects) and horizontal resistance (non-race specific resistance) contributes to overall resistance to red rot (Alarmelu et al., 2010; Babu et al., 2010). An association mapping study on red rot resistance in sugarcane by Singh et al. (2016) identified several markers explaining between 10–17% of variation in resistance scores which was independent of population structure. However, as noted by the authors, this study was limited in statistical power to some extent by the relatively small size of the association mapping panel used (116 clones) and because the majority of genotypes screened fell into the single category of being moderately resistant.

Determining the genetic basis of resistance to the disease through association mapping, and whether it is the same or different in other affected countries, could allow breeders to more effectively select for durable resistance. In Australia, despite red rot not currently being a serious disease, it is of interest to sugarcane breeders in Australia and India to better understand the genetic basis of resistance for two reasons. Firstly, this information may be used in future marker assisted breeding programs to eliminate susceptibility in parental or progeny populations. Secondly, this information may be coupled with future studies to determine the likely reaction of Australian germplasm to races of *C. falcatum* in India. If QTL identified as conferring resistance in Australian populations are not present in Indian breeding programs, these may provide a useful target for introduction by the latter. Conversely, if these QTLs are already present in clones susceptible to red rot in India, this would indicate a likely biosecurity vulnerability to guard against or address.

Genetic studies in sugarcane are usually more challenging in comparison to those with similar goals in other major crops. This is at least partly due to the large and complex genome of sugarcane, which is highly heterozygous and polyploid (frequently aneuploid). Genome wide association studies (GWAS) have been conducted in sugarcane research to identify specific QTL and associated DNA markers for a range of traits including fibre composition (Yang et al., 2019), yield traits (Gouy et al., 2015; Racedo et al., 2016; Barreto et al., 2019; Yang et al., 2020), yellow leaf virus resistance (Debibakas et al.,

2014; Pimenta et al., 2021), leaf angle (Chen et al., 2022) and red rot (Singh et al., 2016) using a combination of diversity array technology (DART), simple sequence repeats (SSR) amplified fragment length polymorphism (AFLP) and SNP markers in populations of 100–300 sugarcane genotypes (clones).

GWAS studies aim to identify individual markers correlated with traits of interest, but are constrained by limited statistical power to identifying only QTL with moderate to large effect, particularly when population sizes are small (eg. panels with <1000 genotypes). This is a potentially important limitation because it is clear that most commercially important traits in sugarcane are controlled by relatively small effects of large numbers of genes. Resolution of marker-trait association by some past studies may also be limited by the small number of available markers relative to the large size of the sugarcane genome.

More recently genomic selection (GS) has been advocated as a potentially useful approach that may be used in sugarcane breeding programs to more accurately select for traits controlled by large numbers of small effect QTL. In sugarcane this method has been applied with encouraging results (Gouy et al., 2013; Deomano et al., 2020; Yadav et al., 2020; Hayes et al., 2021; Islam et al., 2021). Where individual QTL of large effect are identified, GS can also be used in models with single QTL effects identified using genome wide association studies (GWAS) to maximise prediction accuracy (Bernardo, 2014). The development of a high density sugarcane Affymetrix® Axiom® array containing over 58K single nucleotide polymorphism (SNP) (Aitken et al., 2017) allows for low-cost screening of sugarcane germplasm with a far larger number of markers providing greater genome coverage and marker density. This improves the likelihood of identifying markers in close proximity to the gene or QTL underlying a trait of interest.

Here we screened a population of 305 clones representative of clones generated routinely in the Sugar Research Australia (SRA) sugarcane breeding program for response to red rot disease. These same clones were also genotyped using the Affymetrix® Axiom® SNP array to identify specific markers linked to resistance to this disease. We examined the results from analysing data using GWAS and GS for resistance to red rot. We evaluated the accuracy of GS using pedigree, markers and a combination of both to predict red rot resistance. The population of clones studied were also characterised for tonnes of cane per hectare (TCH) and commercial cane sugar (CCS) content, and results from these traits provided a comparison for red rot. In particular, GS prediction accuracies for red rot resistance were compared with TCH and CCS which have been studied previously (Deomano et al., 2020; Hayes et al., 2021). We also determined the genomic location of SNP markers that were strongly associated with red rot resistance in order to help identify candidate genes that may be related causally to the response to red rot. Highly significantly associated SNPs were located in close proximity to multiple stress responsive genes

which have previously been identified in transcription studies and this information may help resolve the genetic control of red rot resistance in sugarcane, and lead to more robust markers for clone selection in the future.

Materials and methods

Genetic materials, field experiment and yield measurements

Three hundred and five clones from a final stage regional selection trial of the SRA sugarcane were used for the study. These clones were representative of those routinely generated and evaluated in the Australian commercial sugarcane breeding program, apart from three clones which were commercial standard cultivars. The clones were derived from 186 different parent clones and 166 crosses, also representative of those used and generated in the Australian sugarcane breeding program.

A field trial was established at Kalamia mill estate in north Queensland, Australia to measure cane yield (tonnes/ha, TCH) and commercial cane sugar (% fresh weight, CCS) (BSES, 1984). About ten percent (30) of the 305 clones evaluated were replicated twice to measure error variance (see section 2.5), while other clones were planted in one replicate. Each individual plot had four rows, 10 m in length, and there was an interrow spacing of 1.6 m. The trial was planted at Kalamia on 6 May 2013, and then cultivated and harvested following recommended local commercial crop management practices. Cane yield (TCH) and CCS were determined at harvest (12-month-old crop) following standard methods used in the SRA breeding program (BSES, 1984), and data for these two traits from the plant crop (first year) and ratoon crop (second year) was collected and analysed.

Isolation and culturing of *Colletotrichum falcatum*

Sugarcane stalks showing typical red rot symptoms were sourced from a sugarcane farm near Mackay, Queensland, Australia. Stem cuttings were thoroughly cleaned with water, sprayed with 70% ethanol and split longitudinally with a sterile knife in a laminar-flow hood. Small pieces of infected tissues were isolated under sterile condition and cultured on potato dextrose agar in Petri dishes and stored at room temperature in dark conditions for several days. Fungal colonies with typical *C. falcatum* morphology were sub-cultured regularly to produce pure isolates. Pathogen identity of isolates was confirmed by conidia, culture morphology, and red rot symptoms in sugarcane. Sporulating fungi were suspended in deionised water and mixed with a kitchen blender to produce inoculum

for red rot screening trials. Conidial concentration of the inoculum was adjusted to 1 million conidia per ml.

Screening of breeding trial clones for red rot resistance

The level of resistance to red rot was observed for all 305 clones in the trial following methods developed by the Sugarcane Breeding Institute, Coimbatore, India and described by Mohanraj et al. (1997). These procedures were briefly as follows. Six-month-old cane stalks with an intact shoot top were used for the screening trial. After stripping off the older leaves and trimming the remaining ones to half-length, the upper part of the stalk with seven visible nodes was cut and inoculated by wrapping cotton swabs moistened with 5 mL of *C. falcatum* conidial suspension or water (control) around the second and third visible nodes from the top. Inoculated stalks were positioned upright with cut end inside wet sand and maintained in a growth chamber set at 30°C, > 90% humidity and constant light for two weeks. The experiment design was a randomised complete block with four replicated stalks inoculated for each clone. A highly resistant standard (negative control) and a highly susceptible standard (positive control) were also included in this experiment.

After two weeks the stalks were split longitudinally and disease symptoms were scored following the metrics described previously (Srinivasan and Bhat, 1961): shoot top condition (0 = healthy, 1 = dry/yellow), lesion width above inoculated node (0, 1, 2, or 3), nodal transgression of lesion (0, 1, 2, or 3 nodes transgressed) and occurrence of white spots (0, 1 = restricted, 2 = progressive) to give a total score from 0 (highly resistant) to 9 (highly susceptible).

Genotyping and SNP marker screening

All clones were genotyped using an Affymetrix Axiom SNP array developed for sugarcane with 58,028 SNPs, previously screened and chosen based on quality parameters and polymorphism in Australian and Brazilian parental clones (Aitken et al., 2017). High-quality DNA was extracted from leaf tissues using a standard CTAB method, treated with proteinase K and purified on a Qiagen column. The Axiom assay was performed on 96-sample Axiom array following the procedure described by Affymetrix (http://media.affymetrix.com/support/downloads/manuals/axiom_2_assay_auto_workflow_user_guide.pdf). DNA samples that had a dish quality control (DQC) measure of less than 0.82 or a quality control (QC) call rate of less than 97% were excluded from the analysis. Allele calling was performed using generated CEL files with Axiom Analysis Suite (1.1.0.616) (http://media.affymetrix.com/support/downloads/manuals/axiom_genotyping_solution_

[analysis_guide.pdf](#)). For each polymorphic marker, all genotypes were given a marker score of 1 if only the most frequent allele was present (i.e. homozygous for this allele), 0 if both alleles were present (i.e. heterozygous), and – 1 if only the minor allele was present. Markers in which one of these three classes occurred for > 98% of the clones were deleted. A total of 56,788 polymorphic markers were retained for further analysis after the above filtering. For each marker, missing values were replaced with the most frequent allele within a marker.

Analysis of phenotypic data

TCH, CCS, red rot rating were analysed under a mixed model framework using a commercial R package, Asreml-R.

For TCH and CCS, an optimal linear mixed model was first determined for each trial crop class data from fitting different fixed, random and residual effects (Butler et al., 2017). For fixed effects replicate, linear row and linear column were considered. Clone is fitted as a random factor as well as spline row and spline column. Spatial variation along the row and along the column was also accounted for. The best model for each trial crop data were then used to fit a model to the multiple trial crop class data. Trial crop is added as a fixed effect in the model. Also, a correlated genetic variance was fitted to the G matrix to account for genotypic variance heterogeneity and correlated measurements between trial crop classes. On the other hand, an unstructured general correlation model and heterogenous variance form was used for the R matrix to account for heterogenous and correlated error variances of trial crop classes.

For red rot, the linear mixed model can be described as follows:

$$y = \mu + \text{date} + \text{rep}/\text{date} + \text{clone} + e$$

where y is the measurement of total score from each plot, μ is a grand mean and e is the residual effects. Date, replicates within date (rep/date) and clone were set as random effects. All the random effects were assumed to follow iid $N(0, I\sigma^2)$. Similarly, BLUPs of clonal effects and the broad-sense heritability were estimated from the model, which were used as observed values in genomic selection.

Genome wide association studies

The association between each individual SNP marker and Red Rot resistance, TCH, and CCS was analysed using ASReml-R package based on the following mixed model:

$$y = a + d + r/d + \text{SNP} + u + e$$

where y is BLUP of a trait obtained from the above analysis, a is the intercept, SNP is a fixed effect of SNP, u is a polygenic

effect $\sim N(0, A\sigma^2\alpha)$ where A is the numerator relationship matrix and $\sigma^2\alpha$ is the polygenic variance; e = random residual effects $\sim \text{iid } N(0, I_d \otimes \sigma^2_{ed})$

Linkage disequilibrium between highly significant SNPs ($P < 10^{-7}$) was calculated using Haploview software (Barrett et al., 2005). SNPs with an LD score of ≥ 0.8 were allocated to the same linkage group for the purpose of examining potentially causal linked genes. These linkage groups are putative only as resolution for this test is limited by our population size. For a comparison with results for red rot, we performed a GWAS for TCH and CCS for all clones in the same population as screened and reported for red rot resistance.

Highly significant SNPs for red rot resistance were aligned to the sorghum genome in order to identify genes that were physically located next to or on top of these SNPs. Some details of the methods used are provided below.

Identification of nearby genes

The genomic location of SNPs most significantly associated with red rot resistance, TCH and CCS was determined by searching the sugarcane and sorghum genomes on the CSIRO public genome browser (<http://gbrowse-ext.bioinformatics.csiro.au/>). Sorghum was used as a reference genome as it is the closest diploid relative to sugarcane with a high level of synteny between the two genomes, but is smaller with more detailed gene annotation (Garsmeur et al., 2018). Genes located at or within 5 kb of a significant SNP were investigated to determine their likely biological function. This was determined by observing their sequence homology to characterised genes, and their transcriptional regulation in sorghum using the Morokoshi sorghum transcriptome database (<http://matsui-lab.riken.jp/morokoshi/Home.html>).

Genomic selection

Methods for genomic selection and prediction mostly followed those detailed by Deomano et al. (2020). For each trait, four Bayesian models and two Machine Learning methods were fitted to the data. The Bayesian models used were BayesA, BayesB, Bayesian Lasso (BL), and Genomic BLUP (GBLUP) (Crossa et al., 2010; Pérez and Campos, 2014). The Bayesian models differ on the assumed distribution of the clone effects. Three sets of explanatory variables per model were used, with these being pedigree (A), marker (M) and pedigree + marker (AM) information (Deomano et al., 2020). Pedigree data over 3–10 generations was retrieved from information on ancestors in a database owned by SRA. What was previously considered to be a semi-parametric method (Gianola et al., 2006), the Reproducing

Kernel Hilbert Spaces (RKHS) is now also included in the Machine Learning (ML) group (Gonzalez-Recio et al., 2014). The A type of explanatory variable was used on RKHS. For both the M and AM type, the Reproducing Kernel Hilbert Spaces-Kernel Averaging (RKHS-KA) model was used (Gonzalez-Camacho et al., 2012). Random Forest (RF) (Breiman, 2001), one of the popular ML methods, which is a tree-based ensemble method for regression was fitted to the data using the A, M and AM explanatory variables. For all three sets of explanatory variables and 6 models, the full SNP data was fitted.

For the Bayesian models including RKHS the BGLR R-package (Pérez and Campos, 2014) was used with default values provided by the software. The number of MCMC iterations, burn in and thinning were 10K, 1K and 10, respectively. The ranger R-package was used for the RF model with $\text{ntree} = 500$, $\text{nodesize} = 5$ and other parameters set to default (Gonzalez-Camacho et al., 2018). Each model per type per trait was cross-validated on 50 replicates of a randomised 80 training:20 test dataset. The prediction accuracy of a model was calculated as the Pearson's product-moment correlation coefficient between the observed trait and predicted trait for the test dataset. Accuracy was calculated for each replicate for each model per type per trait. Accuracies were then averaged across 50 replicates per model per type per trait.

The above methods were also applied for the M type models except with some markers with large effects identified from the GWAS analysis considered as fixed effects, and all other markers considered as random effects. Except for RFR, the second set of SNP data consisting of the remaining non-significant SNPs were fitted as random effects. For the RKHS model, only one kernel was used in the model.

Results

Analysis of variance and distribution of red rot resistance

The broad-sense heritability for the scores for response to red rot was estimated to be 0.89, indicating nearly 90% of variation in measured phenotype was attributable to genetic effects (with the remainder due to experimental or environmental effects). The majority of clones appeared resistant or highly resistant to red rot in response to the screening method (Figure 1).

Genome-wide association studies

A summary of the results of association tests between individual SNP markers and red rot resistance are given in

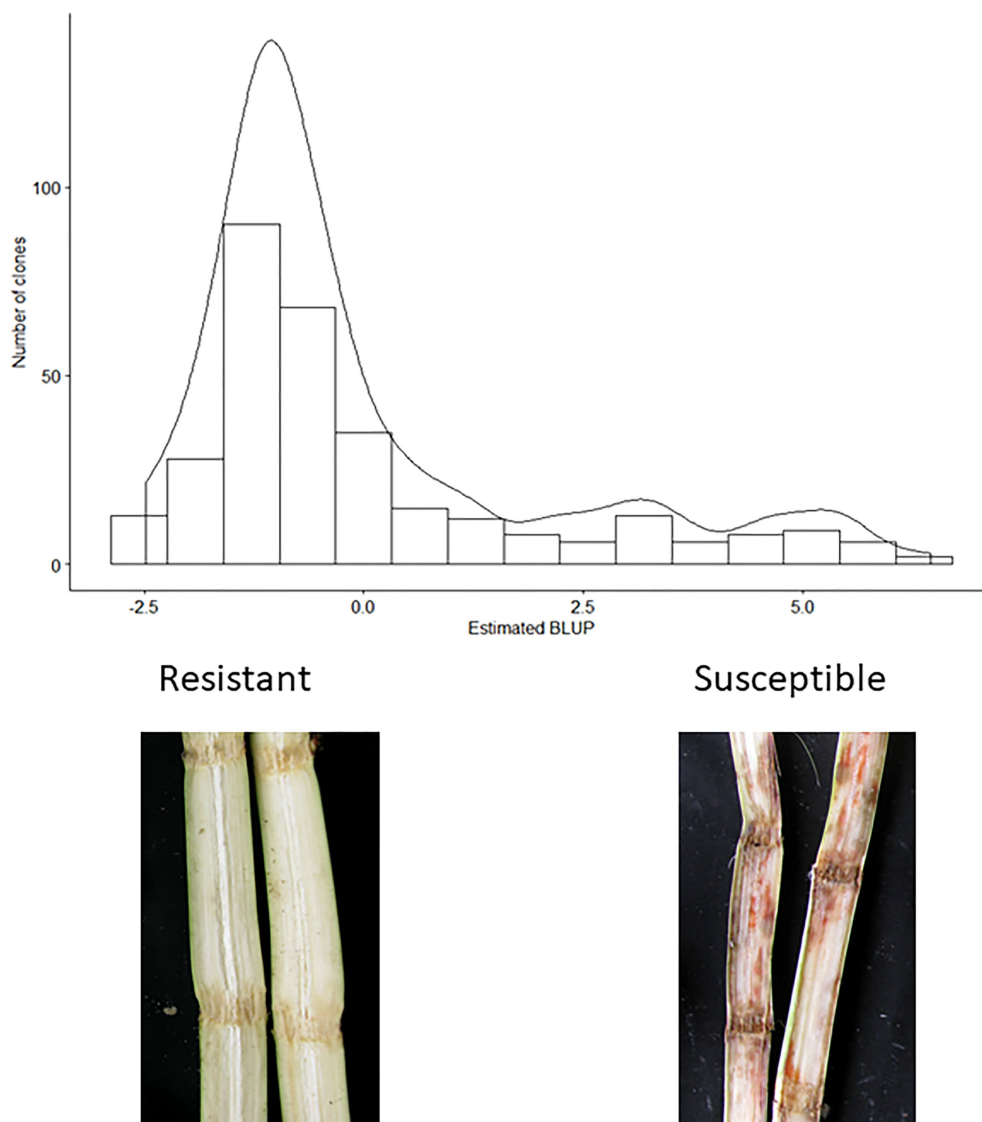


FIGURE 1
Distribution of estimated red rot BLUPs. Lowest BLUP values correspond to greater resistance levels. Photographs show typical symptoms observed in resistant and susceptible clones in the assay.

Table 1. Figure 2 shows a quantile-quantile plot of p-values for all SNPs in the red rot association mapping population. The observed p-values in our study deviated significantly from a normal distribution that would be expected by chance if there was no association with the trait. This is consistent with markers declared significant at the lowest P values having low false discovery rates (i.e. ratio of number of markers expected at the P value to be declared by random chance compared with number observed if the null hypothesis of no markers linked to the trait was valid, Table 1), meaning there is a high level of certainty that

none of these markers are being declared as significant due to Type 1 statistical errors (random chance).

Thirty-five markers were found to be significantly associated with red rot at $P \leq 10^{-7}$, and this level of P corresponded to a near zero false discovery rate. These markers were all found to be in LD, with most in strong LD ($r > 0.8$), suggesting all were associated with a common QTL for resistance to red rot.

The number of markers associated with TCH and CCS at low P values ($<10^{-4}$) was considerably less than for red rot, with no markers associated at P values less than 10^{-6} . However, a

TABLE 1 GWAS results for all traits.

| P value | By random chance | Red Rot resistance | Tonnes cane per hectare | | Commercial cane sugar | |
|----------------|------------------|--------------------|-------------------------|------|-----------------------|------|
| | | | P | 1R | P | 1R |
| $\leq 10^{-2}$ | 352 | 714 | 907 | 1083 | 497 | 1037 |
| $\leq 10^{-3}$ | 35 | 243 | 184 | 232 | 63 | 134 |
| $\leq 10^{-4}$ | 4 | 184 | 42 | 47 | 3 | 23 |
| $\leq 10^{-5}$ | 0 | 141 | 10 | 4 | 0 | 0 |
| $\leq 10^{-6}$ | 0 | 99 | 0 | 0 | 0 | 0 |
| $\leq 10^{-7}$ | 0 | 35 | 0 | 0 | 0 | 0 |
| $\leq 10^{-8}$ | 0 | 10 | 0 | 0 | 0 | 0 |
| $\leq 10^{-9}$ | 0 | 1 | 0 | 0 | 0 | 0 |

The number of SNPs associated with Red Rot resistance, cane yield (TCH) and commercial cane sugar content (CCS) at different p values is shown, along with the number of SNPs expected to be associated with the trait by random chance (i.e. assuming the null hypothesis of no markers linked to red rot resistance). For TCH and CCS, the crop cycle in which the trait was measured is indicated where P, plant crop; 1R, first ratoon.

higher number of markers than expected by random chance were associated with TCH and CCS in both the plant and first ratoon crops for p values between 0.001 and 0.01, with the ratoon crop having higher numbers. For TCH, a higher number of SNPs were also associated with the trait than expected by random chance for P values ≤ 0.0001 and ≤ 0.00001 . This was also the case for CCS for $P \leq 0.0001$ in the ratoon crop only. SNP

associations for each of TCH and CCS traits were strongly correlated between plant and ratoon crop measurements, consistent with performance for each of these traits being correlated between crop cycles (data not shown). However, there was no correlation between SNP marker effects for TCH and CCS, indicating that these two traits appear governed by different genes.

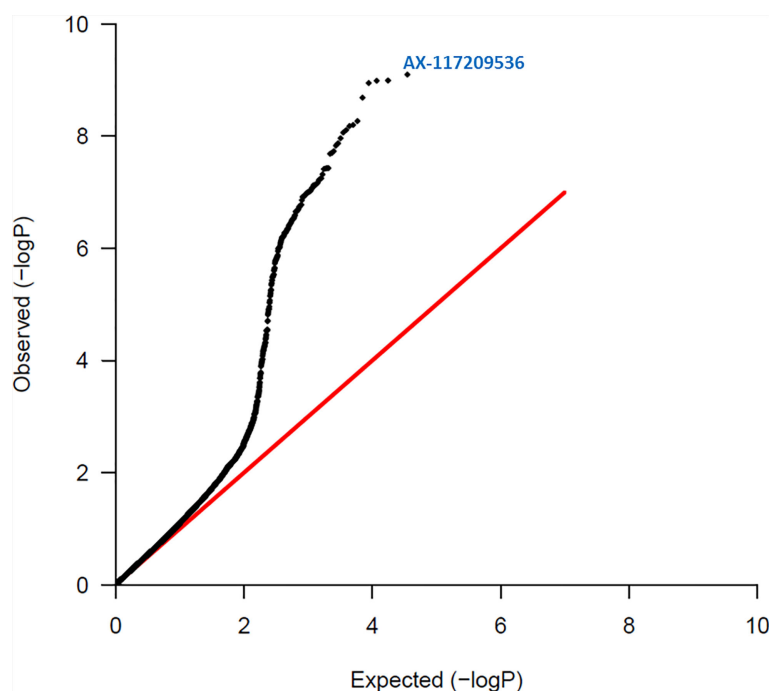


FIGURE 2

Quantile – Quantile plot of red rot SNPs. Expected normal distribution of p-values assuming no associations (null hypothesis) is given on the x axis, and the observed p-values on the y axis. The red line indicates the expected trend if observed p-values were normally distributed. The most significant SNP (AX-117209536) is indicated on the figure and was associated with red rot disease at $p = 8.00 \times 10^{-10}$.

TABLE 2 Genes co-located with red rot resistance associated SNPs.

| SNP (AX#) | P value | Effect size | Nearby genes (Sobic.#) | kb from SNP | Gene family/function |
|-----------|----------|-------------|------------------------|-------------|---|
| 117163812 | 3.73E-08 | -0.8772 | 001G000400 | 0.0 | Pleiotropic drug resistance protein. |
| 117962959 | 1.13E-09 | -0.9654 | 002G141200 | 0.0 | DNA binding protein. |
| 117995732 | 1.39E-07 | 0.8357 | 003G302400 | 0.0 | Unknown protein. |
| 117891751 | 8.65E-09 | 0.8897 | 005G126600 | 0.0 | Auxin signalling F-box 2. |
| 118058341 | 3.78E-08 | 0.8504 | 005G142900 | 0.0 | Pentatricopeptide repeat (PPR) superfamily protein. |
| 118121450 | 8.50E-08 | 0.8353 | 005G143300 | 0.0 | Histone chaperone domain CHZ domain containing protein. |
| 117949242 | 6.62E-09 | -0.9233 | 005G149600 | 2.5 | Cytochrome P450, family 76, subfamily C, polypeptide 2. Biotic stress inducible in Arabidopsis. |
| 117315834 | 1.03E-07 | -0.8287 | 005G153000 | 0.0 | Agenet domain containing protein. |
| 118032095 | 1.09E-08 | -0.8979 | 005G154800 | 0.0 | OsWAK receptor-like cytoplasmic kinase. |
| 118019715 | 9.90E-08 | 0.8439 | 005G160600 | 0.5 | OsFBO15 – F-box and other domain containing protein. |
| 117191093 | 8.34E-08 | -0.8368 | 005G162900 | 0.0 | DNAJ heat shock N-terminal domain-containing protein. |
| 117891837 | 1.01E-07 | 0.8397 | 005G163400 | 0.0 | DNAJ heat shock N-terminal domain-containing protein. Biotic stress responsive in Arabidopsis. |
| 117992398 | 1.67E-07 | 0.8141 | 005G176600 | 1.0 | Eukaryotic aspartyl protease. |
| 117209536 | 8.00E-10 | 0.9071 | 005G177800 | 0.0 | Basic helix-loop-helix (bHLH) DNA-binding superfamily protein. |
| 117272856 | 1.21E-07 | 0.8837 | 005G180900 | 1.0 | DUF630/DUF632 domains containing protein, putative, expressed. bZIP transcription factor. |
| 117155383 | 5.83E-08 | 0.8824 | 005G181000 | 0.0 | Acyl-coenzyme A oxidase. Biotic stress responsive in Arabidopsis. |
| 117133021 | 7.78E-09 | -0.9088 | 005G182200 | 0.0 | NB-ARC domain-containing disease resistance protein. |
| 117177631 | 1.02E-09 | -0.9732 | 005G183600 | 0.0 | Mannose-binding lectin superfamily protein. Similar to Jasmonate-induced protein. |
| 118058750 | 3.87E-08 | -0.9076 | 005G186200 | 0.5 | PATATIN-like protein 4. |
| 117168779 | 7.38E-08 | 0.9070 | 005G189700 | 0.0 | Expressed protein. |
| 117870219 | 2.06E-09 | 0.9613 | 005G200300 | 1.0 | Tyrosine aminotransferase. |
| 118126726 | 1.99E-08 | 0.9474 | 005G204700 | 0.0 | Pectin lyase fold/virulence factor domain containing protein. |
| 118011427 | 4.81E-08 | 0.9321 | 005G209600 | 0.0 | P-loop nucleoside triphosphate hydrolases superfamily protein with CH (Calponin Homology) domain. |
| 117927589 | 1.03E-09 | -0.9799 | 005G212800 | 2.0 | Carboxyl-terminal peptidase, unknown function. |
| 117301744 | 1.34E-08 | 0.8960 | 009G216700 | 0.0 | WD domain, G-beta repeat domain containing protein. |
| 117874758 | 9.39E-08 | 0.8523 | 010G054400 | 0.0 | Leucine-rich repeat protein kinase family protein. Biotic stress responsive in Arabidopsis. |
| 117154432 | 2.00E-07 | 0.8357 | 010G214500 | 0.0 | Purple acid phosphatase. |

(Continued)

TABLE 2 Continued

| SNP (AX#) | P value | Effect size | Nearby genes (Sobic.#) | kb from SNP | Gene family/function |
|-----------|----------|-------------|------------------------|-------------|--|
| 117133579 | 5.39E-09 | -0.9333 | K004600 | 1.0 | NADH:ubiquinone/plastoquinone oxidoreductase, chain 3 protein. |
| 117879610 | 7.51E-08 | -0.8435 | K031500.1 | 0.0 | DNA binding, ATP binding. |

A distance from SNP of 0.0 kb indicates that the SNP occurs within the gene. Only highly significant SNPs ($p \leq 10^{-6}$) with closely co-located genes on the sorghum genome are listed in the table. Effect size indicates the difference between clones with the SNP and the clones that are homozygous and lack the SNP (in rating units).

Location of significant SNPs on sorghum and sugarcane genomes, and nearby genes

SNPs associated with resistance to red rot at $P < 10^{-7}$ were aligned with the sorghum genome, and those SNPs (29) with closely located genes ($< 5\text{kb}$ from the SNP) were listed in Table 2. Many of the SNPs occurred within the coding sequence of genes, reflected in Table 2 by the high proportion of genes located 0.0 kb from significant SNPs. Out of these 29 SNPs, most (21) were aligned to sorghum chromosome 5, and a smaller number were aligned to several other chromosomes (Figure 3). Of these, 17 SNPs corresponded to a 14.6 Mb section of sorghum chromosome 5 (Figure 4). All of these SNPs were in strong LD in the sugarcane population in this study apart from one, AX117209536, which was in weaker LD (although still statistically significant) with the others in this group (Figure 4). A weak LD in our breeding population among SNPs closely physically located on the genome

could arise because of recombination in one or more key ancestors. In addition, sugarcane sorghum chromosomes 5, 6 and 7 are rearranged in some of the chromosomes inherited from *S. spontaneum* which could lead to incomplete synteny to sorghum in this region (Garsmeur et al., 2018).

Genomic prediction

Accuracies of genomic prediction for red rot resistance, CCS and cane yield are presented in Table 3. The accuracies attained for red rot (up to 0.50) were greater than for the other traits. This is consistent with what may be expected considering that a higher number of markers were observed as associated with red rot at low p values than for the other traits.

For red rot, the A type model (i.e. using just pedigree data without marker data) gave lower accuracy than the M and AM models (Table 3). This indicated the inclusion of DNA marker

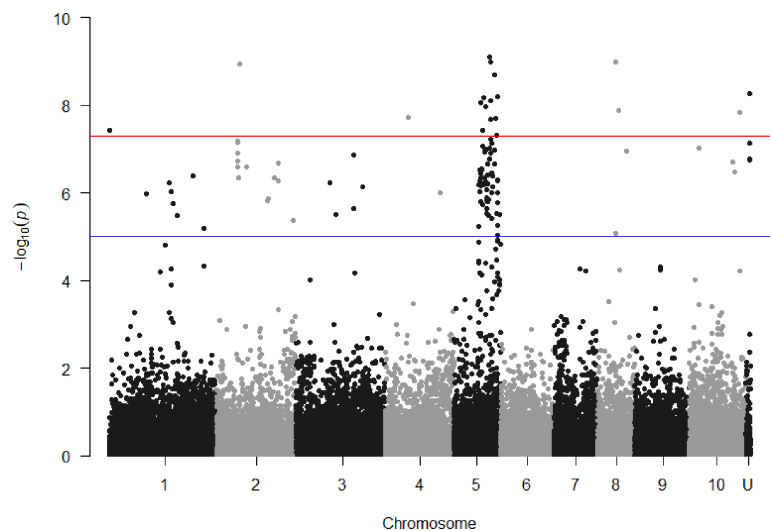


FIGURE 3
Manhattan plots showing P value of SNPs obtained from association analysis for red rot resistance score versus locational alignment to the sorghum genome (chromosomes numbered 1 to 10, U indicating no alignment found). For guidance, the red line indicates $p = 5 \times 10^{-8}$ and the blue line $p = 1 \times 10^{-5}$.

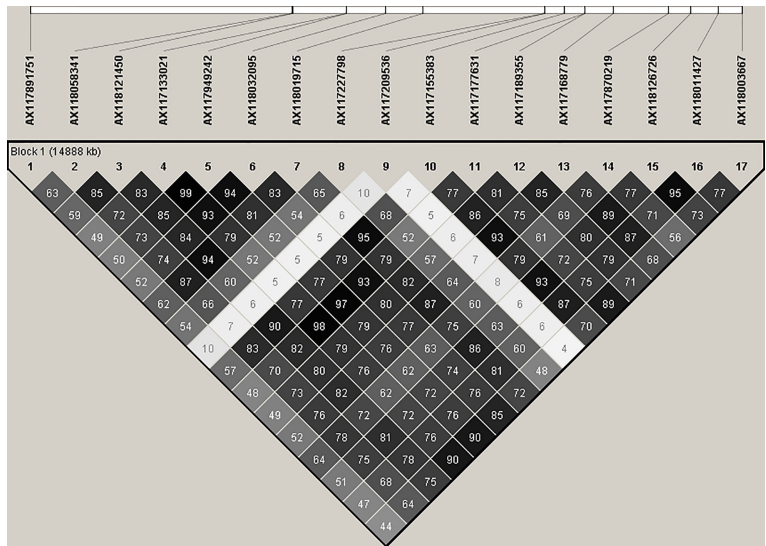


FIGURE 4
Linkage disequilibrium. r^2 (expressed as %) observed between 17 significant SNP markers, and physical alignment to the sorghum genome. These SNPs are closely linked on the sorghum genome and are significantly ($P < 10^{-7}$) associated with resistance to red rot.

TABLE 3 Genomic prediction accuracies per type, per model, per crop class and per trait.

| Type | Model | Red Rot resistance | | CCS | | TCH | |
|------|--------|--------------------|-------|------|------|------|------|
| | | Random | Mixed | P | 1R | P | 1R |
| A | BayesA | 0.25 | | 0.20 | 0.19 | 0.11 | 0.16 |
| A | BayesB | 0.26 | | 0.19 | 0.18 | 0.11 | 0.16 |
| A | BL | 0.25 | | 0.18 | 0.18 | 0.11 | 0.16 |
| A | GBLUP | 0.24 | | 0.22 | 0.22 | 0.11 | 0.16 |
| A | RKHS | 0.25 | | 0.22 | 0.21 | 0.12 | 0.16 |
| A | RFR | 0.18 | | 0.18 | 0.17 | 0.11 | 0.14 |
| M | BayesA | 0.40 | 0.15 | 0.27 | 0.27 | 0.20 | 0.22 |
| M | BayesB | 0.44 | 0.48 | 0.18 | 0.20 | 0.17 | 0.19 |
| M | BL | 0.41 | 0.50 | 0.28 | 0.28 | 0.20 | 0.22 |
| M | GBLUP | 0.41 | 0.50 | 0.28 | 0.28 | 0.20 | 0.22 |
| M | RKHS | 0.41 | 0.51 | 0.29 | 0.29 | 0.19 | 0.22 |
| M | RFR | 0.46 | 0.49 | 0.28 | 0.28 | 0.31 | 0.33 |
| AM | BayesA | 0.40 | | 0.27 | 0.28 | 0.20 | 0.23 |
| AM | BayesB | 0.45 | | 0.23 | 0.23 | 0.17 | 0.19 |
| AM | BL | 0.42 | | 0.28 | 0.29 | 0.20 | 0.22 |
| AM | GBLUP | 0.40 | | 0.28 | 0.28 | 0.18 | 0.22 |
| AM | RKHS | 0.40 | | 0.28 | 0.28 | 0.17 | 0.21 |
| AM | RFR | 0.46 | | 0.27 | 0.28 | 0.30 | 0.32 |

CCS, commercial cane sugar; TCH, tonnes cane per hectare; RR, red rot. The types of model refer to models (Bayes A, Bayes B, Bayesian Lasso, Ridge regression (GBLUP), Kernel Hilbert spaces (RKHS), Random Forest Regression (RFR), with pedigree data only (A), marker data only (M) and pedigree and marker data combined (AM). For Red Rot resistance, accuracies are given for models assuming random marker effects only (random) and models assuming a mixed model (Mixed), with 29 markers given in Table 3 designated as fixed effects and the remaining markers as random effects.

data in the model improved prediction of trait performance over and above that which can be attained through pedigree data alone.

In addition to “standard” genomic prediction models where markers are assumed as random effects, the 29 red rot SNPs listed in Table 2 at $P < 10^{-7}$ were added as fixed effects. Except for one case, the accuracies of prediction in most cases for the model with the fixed effects added were higher (0.48 to 0.51) than without (0.41–0.46). This result indicates adding large effect markers separately in the genomic prediction models can improve prediction capacity. However, there was a single exception to this result with the Bayes A model gave a very low accuracy when the fixed effects were added, and reasons for this were unclear.

Effect of the QTL

Average resistance scores for clones homozygous for each allele, and heterozygous for the alleles were determined for each of 29 markers listed in Table 2. Effects for each of the individual 29 markers were very similar (data not shown), as expected given the strong LD (correlation) among this group of markers, and thus similar to the overall average of all markers (Figure 5). Clones with the resistance allele were more resistant by about 1.5 resistance rating units than clones without the resistance allele. A strong dominance effect of the resistance allele is also apparent, with the average resistance of the heterozygous clones being similar to that for clones homozygous for the allele conferring

resistance (Figure 5). This result suggests that the QTL (or multiple closely linked QTL) linked with this group of SNP markers has both an additive and dominance variation component.

Clones were classed into those putatively with and without the QTL (or multiple QTLs) characterised by the presence of alleles of the 29 linked SNP markers in Table 2, as follows. The allelic composition of each clone was firstly determined in terms of the number of the 29 SNP markers for which it had at least one copy of the allele found positively associated with resistance to red rot. The number of clones with different numbers (ranging from zero to 29) of SNP markers with at least one copy of the positively associated allele is shown in Figure 6. For example, this shows there are 82 clones for which only one of the 29 SNP markers presented with the allele positively associated with red rot resistance. Because of the high linkage disequilibrium among this set of markers, the clones are distributed as two contrasting and distinct groups. This consisted of one group of 170 clones (on the left side of Figure 6) having most of the 29 SNP markers not presenting with the allele associated with resistance (i.e. most markers SNP presenting as homozygous for the alternative alleles to the alleles associated with resistance), and another group (on the right side of Figure 6) having most of the SNP markers presenting with the allele positively associated with resistance (either as being heterozygous for the two alternative SNP alleles or homozygous for the SNP allele positively associated with resistance). Based on these results, the clones on the right-hand side of Figure 6 (with most of the 29 SNP markers) presenting with at least one copy of the resistance allele) were

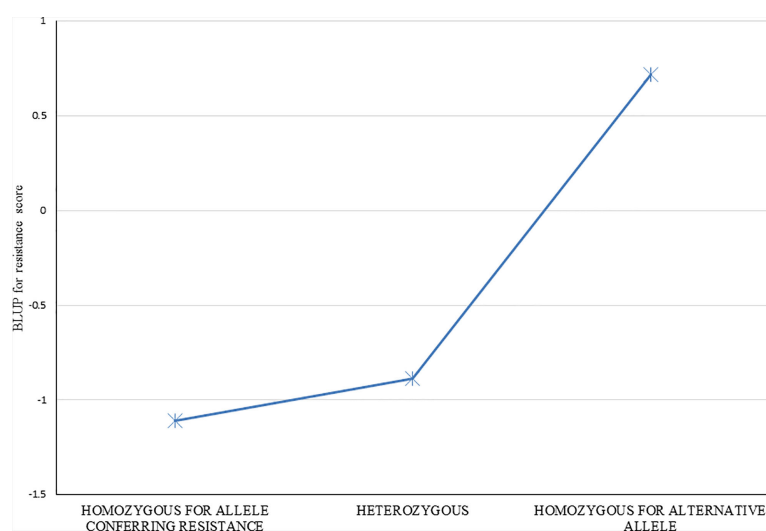


FIGURE 5

Resistance to red rot of groups of clones with the three different observed SNP marker genotypes (i.e. Each of two homozygotes, and the heterozygote) for the markers found associated with red rot resistance at $P < 10^{-7}$. Average resistance score BLUPs for genotypes within each of the three groups were determined for each of the 35 markers individually, and the average of all 35 markers is shown.

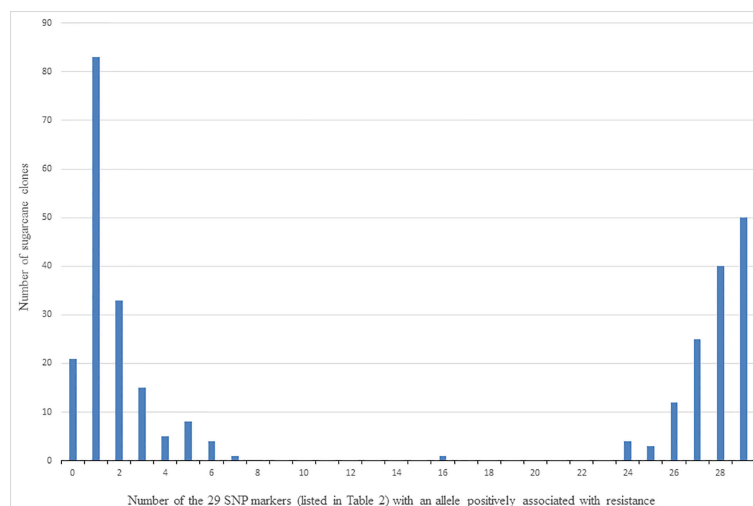


FIGURE 6

Number of clones (out of 305) versus number of the 29 SNP markers listed in Table 2 that detected the allele (for each marker) associated with resistance to red rot.

arbitrarily classed for the purpose of further investigation as having the QTL associated with red rot resistance, while those on the left-hand side were considered to not have the QTL.

As shown in Figure 5, clones were arbitrarily classed into three resistance groups, as (i) susceptible, (ii) intermediate, and (iii) resistant, based on having a BLUP for resistance score of >2 , 2 to 0 , and <0 , respectively. These classifications were cross-tabulated with presence or absence of the putative QTLs for resistance as defined above (Table 4). Of the one hundred and thirty-five clones with the resistance QTLs, 126 were resistant to red rot, and only two were classed as susceptible (Table 4). By contrast, there were 51 clones that were classed as susceptible, 49 of these did not have the resistance QTLs (Table 4). However, there were also 91 clones that were classed as being resistant but without the QTLs, which indicates that some other genetic factors could also contribute to resistance. In summary, these results are consistent with the resistance QTL (linked to the cluster of 21 (16 high LD) markers in Table 2 having a penetrative, dominant effect on resistance, but with some other

genetic factors present on other chromosomes also independently imparting resistance.

Discussion

The results indicate that a major QTL affects resistance to red rot within germplasm generated in the Australian sugarcane breeding programs. The presence of this effect, indicated by a cluster of closely linked SNP markers, appears to provide a high chance of resistance, with only a very small proportion (2% or less) of clones with resistance alleles linked to this QTL showing susceptibility. However, the presence of resistance in over 50% of clones without this QTL also indicates that other genetic factors also contributed to red rot resistance in the germplasm studied, reducing somewhat the overall difference in resistance levels of clones with the QTL versus those without, across the whole population.

The set of clones sampled for this study represented those routinely generated in the commercial sugarcane breeding program in Australia. The observation that less than 20% of the sampled clones were susceptible to red rot, with most clones exhibiting resistance, is consistent with the situation seen for many years in the Australian sugarcane breeding program and industry, where susceptibility to red rot is not generally an important problem. However, this situation is clearly different to other countries, such as southern Asian countries where cultivar susceptibility to red rot is a critical problem, and it is of interest to understand possible reasons for the difference. It is at present unclear if the potentially important QTL conferring resistance to red rot in the germplasm, or other genetic effects,

TABLE 4 Numbers of clones at different levels of resistance and whether a QTL for resistance to red rot was putatively present.

| Resistance level | Resistance QTL present? | | Grand Total |
|------------------|-------------------------|-----|-------------|
| | No | Yes | |
| Susceptible | 49 | 2 | 51 |
| Intermediate | 30 | 7 | 37 |
| Resistant | 91 | 126 | 217 |
| Grand Total | 170 | 135 | 305 |

contributes to the high proportion of resistant clones relative to the situation in other countries, or if pathogen variation is the major cause of differences. It is of interest to understand if the major QTL in the Australian germplasm exists in sugarcane cultivars and breeding programs in countries such as India. At this stage, the value of this QTL in other countries is unknown, and its effectiveness (or vulnerability) against different evolved strains of red rot in other countries should be investigated. The use of common SNP markers available for easily comparing germplasm in different programs will facilitate this investigation in the future. For the Australian sugarcane breeding program, based on the results in this study, it would be possible to screen for several of the SNP markers (listed in Table 2) linked to the major resistance QTL to ensure elimination of red rot susceptible material from any selection populations. In addition, selected SNPs for red rot resistance could possibly be usefully included in a targeted, low-cost marker platform used to screen sugarcane clones for multiple disease traits.

The pattern of results shown in Table 4 suggests some degree of non-additive genetic effects may arise, which could limit prediction of resistance using models based on only additive genetic effects, including the standard GWAS methods and most of the well-established genomic prediction models used in this study. In this situation, models based on decision trees may be more effective in predicting resistance. For example, a decision tree that is based on an initial branch that predicts a clone is relatively resistant if it had a set of alleles associated with resistance for the majority of the SNP loci indicated in Table 2, may be appropriate. This may be one reason why the random forest method of genomic prediction produced slightly better accuracy levels than the other methods (where the QTL effect was not included as a separate fixed effect) (Table 3).

A cluster of SNPs strongly associated with resistance to red rot were located within a 14.6 Mb region of sorghum chromosome 5, suggesting the identification of a novel QTL. There is an indication from the SNP effect that minor QTL are also present on alternative copies of this chromosome in the sugarcane genome. This QTL region has not to our knowledge been previously reported. Singh et al. (2016) used association mapping to identify putative red rot responsive sugarcane QTLs homologous to regions of chromosome 2 and 7 in sorghum. These effects appeared smaller than those identified in our study, although resolution of these was also limited by small population size. Sathyabhama et al. (2015) identified differentially expressed EST clusters in red rot resistant sugarcane variety Co 93009 with sequences that were homologous to regions of sorghum chromosomes 1, 3, 4, 7, 8, and 9.

Many of the genes co-located with SNPs linked to the proposed QTL in this study (and listed in Table 2) encode proteins involved in plant response to pathogens. These include jasmonate induced proteins, pleiotropic drug resistance proteins (a general defence protein) (Sasabe et al., 2002), pectin lyase fold/virulence factor domain containing proteins, DUF630/DUF632

domains containing bZIP transcription factors (involved in pathogen response) (Alves et al., 2013), auxin signalling F-box 2 proteins (Navarro et al., 2006), and NB-ARC domain-containing disease resistance proteins. We have corroborated the expression of some co-located genes at the transcript and protein level. A significant SNP (AX 117962959) which was found on a DNA binding protein gene, and its peptide was present in the proteome developed during *C. falcatum* sugarcane interaction. An RNA binding/nucleic acid binding/zinc ion binding protein was associated with red rot resistance and found in this proteome (Kumar et al., 2020). Another significant SNP (AX 117891751) was located in Auxin signalling F-box 2 gene. A previous study showed differential expression of a F-box domain containing protein in a subtractive library (Sathyabhama et al., 2015) and in the proteome (Kumar et al., 2020). Prathima et al. (2013) also found its expression through differential display (DD-RT-PCR) in a resistant variety after pathogen inoculation. Specific expression of a set of chitinases was demonstrated in a resistant variety as well as in a susceptible variety tolerating red rot development due to plant growth-promoting rhizobacteria (PGPR) mediated induced systemic resistance (Viswanathan et al., 2003; Viswanathan et al., 2009). Subsequently, Rahul et al. (2016) demonstrated that sugarcane chitinase genes were upregulated in sugarcane cells when they are challenged with *C. falcatum* elicitors, and some success has been achieved in suppressing red rot symptoms in sugarcane by transgenic overexpression of chitinases (Tariq et al., 2018) or by application of biocontrol agents that produce chitinases (Joshi et al., 2019). Further investigation of this region of the sugarcane genome may provide greater insights into the mechanisms underlying genetic control of red rot resistance in sugarcane.

Significant SNPs were identified in a cytochrome P450 gene which showed up regulation in a resistant variety (Prathima et al., 2013). The SNPs (AX118019715, AX117191093, AX117891837) are located at genes which showed specific expression in a resistant variety (Co 93009) after pathogen inoculation in proteomic studies (Kumar et al., 2020). The SNP (AX117133021) was co-located on a disease resistance gene and specific expression of disease resistant protein RPM1 and RPS5 was established in a resistant sugarcane variety Co 93009 (Viswanathan et al., 2016). For SNP (AX117874758), increased expression of its co-located gene has been found in a resistant variety (Sathyabhama et al., 2015). Using differential display (DD)-RT-PCR, LRR family protein expression was found to increase in resistant variety (Co 93009) after pathogen inoculation and in sugarcane suspension cultures treated with *C. falcatum* elicitors (Prathima et al., 2013; Rahul et al., 2016). The SNP co-located to the gene NADH ubiquinone oxidoreductase (SNP ID AX117133579) was also identified in a resistant variety in a gene expression study (Rahul et al., 2016).

The 14.6 Mb region homologous to the sorghum genome in Figure 3 also contains genes which have orthologues reportedly

transcriptionally regulated in response to stress when tested in the model plant *Arabidopsis thaliana*. This includes a number of genes associated with fungal pathogen defence including a cluster of chitinase A genes (Sobic.005G177100, Sobic.005G177400, Sobic.005G177500, Sobic.005G177600) whose orthologues are transcriptionally activated in response to abiotic stress in *Arabidopsis* (Berardini et al., 2015). Interestingly, the single most significant SNP in our study (AX117209536, $p = 8.00E-10$) was also the closest SNP to this cluster, located just 13 kb away.

At this stage the ancestral origin of the major QTL identified in this study is not known. Markers with $LD \geq 0.8$ are considered to be in strong LD and Figure 4 shows clusters of markers in strong LD which probably correspond to ancestral haplotypes segregating as large blocks in the population of sugarcane cultivars screened. This cluster appears to correspond to a homolog of sorghum chromosome 5. It is known from other mapping studies and cytogenetic analysis there are from 10–12 homologous copies of every chromosome in sugarcane (Aitken et al., 2005; Piperidis and D'Hont, 2020). Wei et al. (2007) detected LD in a collection of sugarcane cultivars and identified haplotype blocks that contained from 2 to 10 markers although high significance levels were needed to reduce spurious associations. They identified significant LD between markers up to 40 cM apart but the majority of LD occurred between 0 and 30 cM. In this study, when four of the markers associated with red rot were also mapped to a single linkage group aligned to sorghum chromosome 5 of a genetic map of variety Q208, they covered 33.7 cM (K Aitken pers comm.). This is consistent with the LD previously identified in sugarcane and its breeding history which has a strong foundation bottleneck (Wei et al., 2007).

Results for cane yield and CCS on the same materials in this study provided an interesting benchmark for the results for red rot resistance. The number of markers observed for association with red rot resistance was much higher, and P values lower, than for cane yield and CCS, consistent with larger additive marker effects for red rot resistance. The genomic prediction accuracies for TCH and CCS in the population reported here were low (< 0.25) because of the relatively small number of clones observed (307). These accuracies are lower than those found with a larger number of clones (about 2500) measured in prior work (with accuracies of >0.35) (Deomano et al., 2020). However, a significantly higher number of markers were observed as being significantly associated with both cane yield and CCS than expected by random chance, indicating that markers are explaining a proportion of variation observed and indicative of the value of the extra data collected. It is likely that greater accuracy and resolution of smaller marker effects linked to red rot resistance could be attained with a larger population than the 305 clones used in the current study.

Data availability statement

The data presented in the study are deposited in the GitHub repository, accession number <https://github.com/SugarAus/data.git>.

Author contributions

AO'C, JD, PL, ED, XW, PJ, and KA performed the experiments and/or analysed data. PJ, PL, and RM conceived and designed the project. HG, KM, RV, and BR contributed to methodology development and related technical expertise. All authors contributed to the article and approved the submitted version.

Funding

This project was supported by the Australian Government under the Australia-India Strategic Research Fund grant number 48454, and by Sugar Research Australia, Brisbane and the ICAR Sugarcane Breeding Institute, Coimbatore, India.

Conflict of interest

Authors AO'C, JD, ED, XW and PL were employed by the company Sugar Research Australia Limited.

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

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.2022.1021182/full#supplementary-material>

References

- Aitken, K., Farmer, A., Berkman, P., Muller, C., Wei, X., Demano, E., et al. (2017). Generation of a 345K sugarcane SNP chip. *Int. Sugar J.* 119, 816–820.
- Aitken, K., Jackson, P., and McIntyre, C. (2005). A combination of AFLP and SSR markers provides extensive map coverage and identification of homo(eo)logous linkage groups in a sugarcane cultivar. *Theor. Appl. Genet.* 110, 789–801. doi: 10.1007/s00122-004-1813-7
- Alarmelu, S., Nagarajan, R., Shanthi, D., Mohanraj, D., and Padmanaban, P. (2010). A study on genetics of red rot resistance in sugarcane. *Electron. J. Plant Breed* 1, 656–659.
- Alves, M., Dadalto, S., Gonçalves, A., De Souza, G., Barros, V., and Fietto, L. (2013). Plant bZIP transcription factors responsive to pathogens: a review. *Int. J. Mol. Sci.* 14, 7815–7828. doi: 10.3390/ijms14047815
- Babu, C., Natarajan, U., Shanthi, R., Govindaraj, P., Sundar, A., and Viswanathan, R. (2010). Inheritance of red rot resistance in sugarcane (*Saccharum* sp. hybrids). *Sugar Tech.* 12, 167–171. doi: 10.1007/s12355-010-0032-6
- Barreto, F., Rosa, J., Balsalobre, T., Pastina, M., Silva, R., Hoffmann, H., et al. (2019). A genome-wide association study identified loci for yield component traits in sugarcane (*Saccharum* spp.). *PLoS One* 14 (7), e0219843. doi: 10.1371/journal.pone.0219843
- Barrett, J., Fry, B., Maller, J., and Daly, M. (2005). HAPLOVIEW: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Berardini, T., Reiser, Z., Li, D., Mezheritsky, Y., Muller, R., Strait, E., et al. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* 53, 474–485. doi: 10.1002/dvg.22877
- Bernardo, R. (2014). Genomewide selection when major genes are known. *Crop Sci.* 54, 68. doi: 10.2135/cropsci2013.05.0315
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- BSES (1984). "The standard laboratory manual for Australian sugar mills," in *Volume 1. principles and practices* (Brisbane, Australia: Bureau of Sugar Experiment Stations).
- Butler, D., Cullis, B., Gilmour, A., Gogel, B., and Thompson, R. (2017). *ASReml-r reference manual version 4* (Hemel Hempstead, UK: VSN International Ltd).
- Chen, X., Huang, Z., Fu, D., Fang, J., Zhang, X., Feng, X., et al. (2022). Identification of genetic loci for sugarcane leaf angle at different developmental stages by genome-wide association study. *Front. Plant Sci.* 13, 841693. doi: 10.3389/fpls.2022.841693
- Chona, B. (1980). Red rot of sugarcane and sugar industry – a review. *Indian Phytopathol.* 33, 191–207
- Crossa, J., Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186, 713–724. doi: 10.1534/genetics.110.118521
- Debibakas, S., Rocher, S., Garsmeur, O., Toubi, L., Roques, D., D'Hont, A., et al. (2014). Prospecting sugarcane resistance to sugarcane yellow leaf virus by genome-wide association. *Theor. Appl. Genet.* 127, 1719–1732. doi: 10.1007/s00122-014-2334-7
- Deomano, E., Jackson, P., Wei, X., Aitken, K., Kota, R., and Pérez-Rodríguez, P. (2020). Genomic prediction of sugar content and cane yield in sugar cane clones in different stages of selection in a breeding program, with and without pedigree information. *Mol. Breed.* 40, 38. doi: 10.1007/s11032-020-01120-0
- FAO (2020). *Food and agriculture society of the united nations*. Available at: <https://www.fao.org/faostat/en/#data/QC> (Rome, Italy).
- Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 9, 2638. doi: 10.1038/s41467-018-05051-5
- Gianola, D., Fernando, R., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173, 1761–1776. doi: 10.1534/genetics.105.049510
- Gonzalez-Camacho, J., Campos, G., Pérez, P., Gianola, D., Cairns, J., Mahuku, G., et al. (2012). Genome-enabled prediction of genetic values using radial basis function neural networks. *Theor. Appl. Genet.* 125, 759–771. doi: 10.1007/s00122-012-1868-9
- Gonzalez-Camacho, J., Ornella, L., Pérez-Rodríguez, P., Gianola, D., Dreisigacker, S., and Crossa, J. (2018). Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. *Plant Genome* 11, 170104. doi: 10.3835/plantgenome2017.11.0104
- Gonzalez-Recio, O., Rosa, G., and Gianola, D. (2014). Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits. *Livest. Sci.* 166, 217–231. doi: 10.1016/j.livsci.2014.05.036
- Gouy, M., Rousselle, Y., Bastianelli, D., Lecomte, P., Bonnal, L., Roques, D., et al. (2013). Experimental assessment of the accuracy of genomic selection in sugarcane. *Theor. Appl. Genet.* 126, 2575–2586. doi: 10.1007/s00122-013-2156-z
- Gouy, M., Rousselle, Y., Chane, A., Anglade, A., Royaert, S., Nibouche, S., et al. (2015). Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane. *Euphytica* 202, 269–284. doi: 10.1007/s10681-014-1294-y
- Hayes, B., Wei, X., Joyce, P., Atkin, F., Deomano, E., Yue, J., et al. (2021). Accuracy of genomic prediction of complex traits in sugarcane. *theor. Appl. Genet.* 134, 4414–4423. doi: 10.1007/s00122-021-03782-6
- Heinz, D. (1987). *Sugarcane improvement through breeding* (Elsevier: Amsterdam).
- Islam, M., McCord, P., Olatoye, M., Qin, L., Sood, S., Lipka, A., et al. (2021). Experimental evaluation of genomic selection prediction for rust resistance in sugarcane. *Plant Genome* 14, e20148. doi: 10.1002/tpg2.20148
- Jackson, P. (2018). "Advances in conventional sugarcane breeding programmes," in *Achieving sustainable cultivation of sugarcane*, vol. 2. Ed. P. Rott (Cambridge: Burleigh Dodds Science), 75–98.
- Joshi, D., Singh, P., Holkar, S., and Kumar, S. (2019). Trichoderma-mediated suppression of red rot of sugarcane under field conditions in subtropical India. *Sugar Tech* 21, 496–504. doi: 10.1007/s12355-018-0624-0
- Kumar, V., Viswanathan, R., Malathi, P., Sundar, A., Prasanth, C., and Nandakumar, M. (2020). Identification of differential expressed proteins and establishing a defense proteome of sugarcane in response to colletotrichum falcatum infection. *J. Plant Pathol.* 102, 685–702. doi: 10.1007/s42161-020-00577-4
- Mohanraj, D., Padmanaban, P., Viswanathan, R., and Alexander, K. (1997). Sugarcane screening for red rot resistance. *Sugar Cane* 3, 18–23.
- Natarajan, U., Balasundaram, S., Ramana Rao, T., Padmanaban, P., Mohanraj, D., Karthigeyan, S., et al. (2001). Role of saccharum spontaneum in imparting stable resistance against sugar cane red rot. *Int. Sugar J.* 27, 17–20.
- Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., et al. (2006). A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. *Science* 312, 436–439. doi: 10.1126/science.1126088
- Pérez, P., and Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442
- Pimenta, R., Aono, A., Burbano, R., Coutinho, A., da Silva, C., Anjos, I., et al. (2021). Genome-wide approaches for the identification of markers and genes associated with sugarcane yellow leaf virus resistance. *Sci. Rep.* 11, 15730. doi: 10.1038/s41598-021-95116-1
- Piperidis, N., and D'Hont, A. (2020). Sugarcane genome architecture decrypted with chromosome-specific oligo probes. *Plant J.* 103, 2039–2051. doi: 10.1111/tpj.14881
- Prathima, P., Raveendran, T., Kumar, K., Rahul, P., Kumar, V., Viswanathan, R., et al. (2013). Differential regulation of defense-related gene expression in response to red rot pathogen colletotrichum falcatum infection in sugarcane. *Appl. Biochem. Biotechnol.* 171, 488–503. doi: 10.1007/s12010-013-0346-4
- Racedo, J., Gutierrez, L., Francisca Perera, M., Ostengo, S., Mariano Pardo, E., Ines Cuenya, M., et al. (2016). Genome-wide association mapping of quantitative traits in a breeding population of sugarcane. *BMC Plant Biol.* 16, 142. doi: 10.1186/s12870-016-0829-x
- Rahul, P., Ganesh Kumar, R., Viswanathan, R., Ramesh Sundar, A., Malathi, P., Naveen Prasanth, C., et al. (2016). Defense transcriptome analysis of sugarcane and *Colletotrichum falcatum* interaction using host suspension cells and pathogen elicitor. *Sugar Tech* 18, 16–28. doi: 10.1007/s12355-014-0356-8
- Ram, B., Singh, N., and Sahi, B. (2006). Selection for juice quality and red rot disease index in sugarcane (*Saccharum officinarum* L.). *Indian J. Genet. Plant Breed.* 66, 151–152.
- Sasabe, M., Toyoda, K., Shiraishi, T., Inagaki, Y., and Ichinose, Y. (2002). cDNA cloning and characterization of tobacco ABC transporter: NtPDR1 is a novel elicitor-responsive gene. *FEBS Lett.* 518, 164–168. doi: 10.1016/S0014-5793(02)02697-2
- Sathyabhama, M., Viswanathan, R., Nandakumar, M., Malathi, P., and Ramesh Sundar, A. (2015). Understanding sugarcane defence responses during the initial phase of *Colletotrichum falcatum* pathogenesis by suppression subtractive hybridization (SSH). *Physiol. Mol. Plant Pathol.* 91, 131–140. doi: 10.1016/j.pmpp.2015.07.003

- Singh, R., Banerjee, K., Khan, M., Yadav, S., Kumar, S., Duttamajumder, S., et al. (2016). Identification of putative candidate genes for red rot resistance in sugarcane (*Saccharum* species hybrid) using LD-based association mapping. *Mol. Genet. Genom.* 291, 1363–1377. doi: 10.1007/s00438-016-1190-3
- Srinivasan, K., and Bhat, N. (1961). Red rot of sugarcane – criteria for grading resistance. *J. Indian Bot. Soc.* 40, 566–577.
- Tariq, M., Khan, A., Tabassum, B., Toufiq, N., Bhatti, M., Riaz, S., et al. (2018). Antifungal activity of chitinase II against *Colletotrichum falcatum* causing red rot disease in transgenic sugarcane. *Turk. J. Biol.* 42, 45–53. doi: 10.3906/biy-1709-17
- Viswanathan, R. (2021). Sustainable sugarcane cultivation in India through threats of red rot by varietal management. *Sugar Tech* 23, 239–253. doi: 10.1007/s12355-020-00882-3
- Viswanathan, R., Amalraj, R., Selvakumar, R., and Malathi, P. (2018). “Progress in understanding fungal diseases affecting sugarcane: red rot,” in *Achieving sustainable cultivation of sugarcane, volume 2: Breeding, pests and diseases*. Ed. P. Rott (Cambridge: Burleigh Dodds Science Publishing), 201–219. Available at: <https://www.jstor.org/stable/43215546>
- Viswanathan, R., Nandakumar, R., and Samiyappan, R. (2003). Role of pathogenesis-related proteins in rhizobacteria-mediated induced systemic resistance against *Colletotrichum falcatum* in sugarcane. *J. Plant Dis. Prot.* 110, 524–534.
- Viswanathan, R., Padmanaban, P., and Selvakumar, R. (2020). Emergence of new pathogenic variants in *Colletotrichum falcatum*, stalk infecting ascomycete in sugarcane: role of host varieties. *Sugar Tech* 22, 473–484. doi: 10.1007/s12355-019-00780-3
- Viswanathan, R., Sathyabhama, M., Malathi, P., and Sundar, A. (2016). Transcriptome analysis of host-pathogen interaction between sugarcane and *Colletotrichum falcatum* by suppression subtractive hybridization and illumina sequencing. *proc. Int. Soc. Sugarcane Technol.* 29, 1639–1644.
- Viswanathan, R., Selvakumar, R., Malathi, P., and Prakasam, N. (2021). Controlled condition testing (CCT): An ideal high-throughput method for screening of pre-release clones and progenies for red rot resistance in sugarcane. *Sugar Tech* 23, 1045–1055. doi: 10.1007/s12355-021-00970-y
- Viswanathan, R., Sundar, A., Malathi, P., Rahul, P., Ganesh Kumar, V., Banumathy, R., et al. (2009). Interaction between sugarcane and *Colletotrichum falcatum* causing red rot: Understanding disease resistance at transcription level. *Sugar Tech* 11, 45–50. doi: 10.1007/s12355-009-0008-6
- Wei, X., Jackson, P., McIntyre, C., Aitken, K., and Croft, B. (2007). Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor. Appl. Genet.* 114, 155–164. doi: 10.1007/s00122-006-0418-8
- Yadav, S., Jackson, P., Wei, X., Ross, E., Aitken, K., Deomano, E., et al. (2020). Accelerating genetic gain in sugarcane breeding using genomic selection. *Agronomy* 10, 585. doi: 10.3390/agronomy10040585
- Yang, X., Luo, Z., Todd, J., Sood, S., and Wang, J. (2020). Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (*Saccharum* spp.). *Plant Genome* 13, e20006. doi: 10.1002/tpg2.20006
- Yang, X., Todd, J., Arundale, R., Binder, J., Luo, Z., Islam, M., et al. (2019). Identifying loci controlling fiber composition in polyploid sugarcane (*Saccharum* spp.) through genome-wide association study. *Ind. Crops Prod.* 130, 598–605. doi: 10.1016/j.indcrop.2019.01.023



OPEN ACCESS

EDITED BY

Hairul Roslan,
Universiti Malaysia Sarawak, Malaysia

REVIEWED BY

Rasappa Viswanathan,
Indian Council of Agricultural Research
(ICAR), India
Krishan Kumar,
Indian Institute of Agricultural
Biotechnology (ICAR), India

*CORRESPONDENCE

Khaled Masmoudi
khaledmasmoudi@uaeu.ac.ae

SPECIALTY SECTION

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

RECEIVED 25 August 2022

ACCEPTED 18 October 2022

PUBLISHED 08 November 2022

CITATION

Abdul Aziz M, Brini F, Rouached H and
Masmoudi K (2022) Genetically
engineered crops for sustainably
enhanced food production systems.
Front. Plant Sci. 13:1027828.
doi: 10.3389/fpls.2022.1027828

COPYRIGHT

© 2022 Abdul Aziz, Brini, Rouached and
Masmoudi. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Genetically engineered crops for sustainably enhanced food production systems

Mughair Abdul Aziz¹, Faical Brini², Hatem Rouached³
and Khaled Masmoudi^{1*}

¹Department of Integrative Agriculture, College of Agriculture and Veterinary Medicine, United Arab Emirates University, Al-Ain, Abu-Dhabi, United Arab Emirates, ²Biotechnology and Plant Improvement Laboratory, Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia,

³Michigan State University, Plant and Soil Science Building, East Lansing, MI, United States

Genetic modification of crops has substantially focused on improving traits for desirable outcomes. It has resulted in the development of crops with enhanced yields, quality, and tolerance to biotic and abiotic stresses. With the advent of introducing favorable traits into crops, biotechnology has created a path for the involvement of genetically modified (GM) crops into sustainable food production systems. Although these plants heralded a new era of crop production, their widespread adoption faces diverse challenges due to concerns about the environment, human health, and moral issues. Mitigating these concerns with scientific investigations is vital. Hence, the purpose of the present review is to discuss the deployment of GM crops and their effects on sustainable food production systems. It provides a comprehensive overview of the cultivation of GM crops and the issues preventing their widespread adoption, with appropriate strategies to overcome them. This review also presents recent tools for genome editing, with a special focus on the CRISPR/Cas9 platform. An outline of the role of crops developed through CRISPR/Cas9 in achieving sustainable development goals (SDGs) by 2030 is discussed in detail. Some perspectives on the approval of GM crops are also laid out for the new age of sustainability. The advancement in molecular tools through plant genome editing addresses many of the GM crop issues and facilitates their development without incorporating transgenic modifications. It will allow for a higher acceptance rate of GM crops in sustainable agriculture with rapid approval for commercialization. The current genetic modification of crops forecasts to increase productivity and prosperity in sustainable agricultural practices. The right use of GM crops has the potential to offer more benefit than harm, with its ability to alleviate food crises around the world.

KEYWORDS

GM crops, genome editing, sustainable agriculture, food production, environmental constraints

1 Introduction

Agriculture faces severe challenges for delivering food and maintaining nutritional security through sustainable practices. In relation to the concept of sustainability, sustainable agriculture is defined as a system of growing crops for the short and long-term period without damaging the environment, society, and the economy for the present and future generations (Tripathi et al., 2022). The main goals of sustainable agriculture are to produce high yield of healthy crop products, efficiently use the environmental resources with minimal damages, enhance the quality of life within the society through the just distribution of food, and provide economic benefits for the farmers (Tseng et al., 2020). These goals have become a prominent issue of discussion in agriculture in the past few years and have been recognized widely in scientific communications, since it is difficult to produce large amounts of food with minimal environmental degradation.

However, there has been a remarkable breakthrough in the field of agriculture through plant genetic modification. Plant biotechnology has generated products that helped agriculture sector to achieve enhanced yields in a more sustainable manner. It has witnessed an increase in the production capacity that is as huge as it was during the period of the green revolution in the early 70's (Raman, 2017). A genetically modified (GM) crop is defined as any plant whose genetic material has been manipulated in a particular way that does not occur under natural conditions, but with the aid of genetic techniques (Sendhil et al., 2022). Agriculture is the first sector that invested heavily in the use of genetic modifications (Raman, 2017). The massive experiments in agricultural biotechnology have enabled the development of suitable traits in plants for food production. The employment of genetic tools for the introduction of a foreign gene, as well as the silencing and expressing of a specific gene in plants, have brought a dramatic expansion of GM crops (Kumar et al., 2020). It has led to the propagation of crops that are disease resistant, environmental stress tolerant, and have an improved nutrient composition for consumers (Batista et al., 2017).

The techniques for the improvement of plants for food production have been undertaken since the humankind stopped migration and relied on agriculture for their survival. At present, more advanced molecular tools are developed for specific genetic manipulation of crops than the conventional methods. Genome editing is the process of making targeted improvements to a plant's genome, specifically within plant's own family (Kaur et al., 2022). Its precision in changing almost any desired location in the genome makes it discrete from other breeding methods. Most of the changes that are made through genome editing occurs naturally within the plants, through traditional breeding or evolution (Graham et al., 2020). However, through genome editing such results are obtained within years rather than decades. With this method, there is no addition of foreign genes, and it is more accurate and predictable than earlier techniques of plant genetic modification (Kaur et al., 2022).

In the twenty-first century, the genetic modification of crops is considered a potential solution for achieving the goals of sustainable agriculture (Oliver, 2014). However, the use of GM crops has raised complex issues and dilemmas related to their safety and sustainability. There have been several debates which have led some countries to contest the use, cultivation, and commercialization of GM crops (Kikulwe et al., 2011). Specifically, the majority of European and Middle Eastern countries have imposed full or partial limitations on the commercialization of GM crops. Regulatory approval for the commercialization of GM crops is hampered by poor communication and awareness brought about by consumer mistrust (Mustapa et al., 2021). Moreover, the difficult process of completing risk assessments and meeting biosafety regulations, has only compounded the existing mistrust of GM crops, based on ethics, history and customs.

Nevertheless, because the GM crops are considered as good candidates for sustainable food production, it is imperative to perform the risk assessment of any developed GM crop, exploring their negative and positive consequences for the current agricultural developments. In this regard, the goal of the present study is to evaluate the use of genetic manipulation and genome editing of crops for overcoming the global food challenges in a sustainable manner. It aims to review current knowledge of GM crops, the concerns and dilemmas associated with them and provides appropriate solutions to overcome them. The study further delivers several perspectives on their incorporation into sustainable food production systems and eliminate the mistrust placed on GM crops for the achievement of Sustainable Development Goals (SDGs).

2 Developmental pathway of GM crops over the years

The genetic modification of plants dates back approximately 10,000 years with the practice of artificial selection and selective breeding. The selection of parents with favorable traits and their utilization in breeding programs has facilitated the introgression of these traits into their offspring's (Raman, 2017). For instance, artificial selection of maize out of weedy grasses having smaller ears and less kernels, has resulted in the generation of edible maize cultivars (Doebley, 2006). In 1946, the advancement leading to contemporary genetic modification took place, with the scientist's discovery of genetic material being moveable between various species (Figure 1) (James, 2011). This was accompanied with the identification of the double helical DNA structure and concept of the central dogma in 1954 by Watson and Crick (Cobb, 2017). Successive advances in the experiments by Boyer and Cohen in 1973 that included the extraction and introduction of DNA between various species resulted in the engineering of the World's first GM organism (Cohen et al., 1973). In 1983, antibiotic resistant tobacco and petunia, first GM crops, were auspiciously developed by three independent scientists (Fraley, 1983).

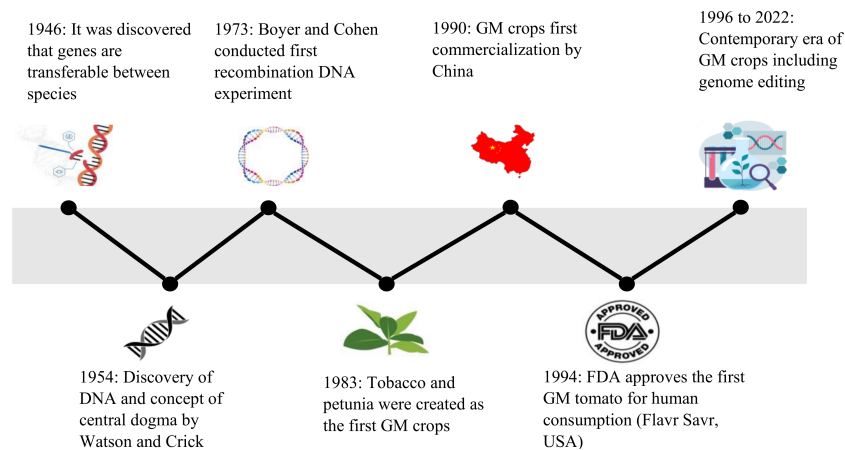


FIGURE 1
Timeline of various events from the discovery of genes being transferable during 1946 leading to the contemporary era of advanced tools for developing GM crops.

In 1990, GM tobacco plants that were resistance to tobacco mosaic virus (TMV) were first commercialized by China (Wu and Butz, 2004). In 1994, Food and Drug Administration (FDA) approved the Flavr Savr tomato (Calgene, USA) as the first GM crop for human consumption (Vega Rodríguez et al., 2022). The antisense technology was used to genetically modify this tomato plant by interfering the production of the enzyme polygalacturonase, major enzyme responsible for pectin disassembly in ripening fruit, that retarded its ripening and protected it from rot (Bawa and Anilakumar, 2013). Several transgenic plants were approved since then for expansive production in 1995 and 1996. For instance, transgenic cantaloupe Charentais melons expressing an antisense ACC oxidase gene were developed to block their ripening process (Ayub et al., 1996). Some of the GM crops that received initial FDA-approval included cotton, corn and potatoes (modification of *Bacillus thuringiensis* (Bt) gene, Monsanto), Roundup Ready soybeans (resistance to glyphosate, Monsanto), and canola (increased oil production, Calgene) (Bawa and Anilakumar, 2013). At present, the genetic modifications are performed on various cereals, fruits, and vegetables that includes rice, wheat, strawberry, lettuce, and sugarcane. The genetic modifications are also carried out to increase vaccine bioproduction in plants, improved nutrients in animal feed, and for conferring environmental stresses such as salinity and drought (Kurup and Thomas, 2020).

2.1 Method of genetic modification of crops

The creation of a GM crop is a complex phenomenon that involves several steps, from the identification of the target gene to the regeneration of transformed plants (Figure 2).

2.1.1 Target gene identification

Developing a GM plant requires the determination of the gene of interest for a particular trait such as drought tolerance gene that is already present in a specific plant species (Snow and Palma, 1997). The genes are identified using the available data and knowledge about their sequences, structures, and functionalities. In case of an unknown gene, a much laborious method will be used, such as map-based cloning. The gene of interest is isolated and amplified using the Polymerase Chain Reaction (PCR). It allows the desired gene to be enlarged into several million copies for the gene assembly (Schouten et al., 2006).

2.1.2 Cloning of the gene of interest and its insertion into a transfer vector

After several copies of genes are attained, it is inserted into a construct downstream a strong promoter and upstream a terminator. This complex is then transferred into bacterial plasmid (manufacturing vectors), allowing for the duplication of gene of interest within the bacterial cell (Zupan and Zambryski, 1995). The DNA construct with the gene of interest is introduced into the plants *via Agrobacterium tumefaciens* or gene gun (particle bombardment) (Lacroix and Citovsky, 2020).

2.1.3 Modified plant cells selection and plant regeneration

When using antibiotic resistance as a selectable marker gene, only transformed plant cells survive and will be regenerated to entire plant using different regeneration techniques (Ibáñez et al., 2020). Several genetic analyses are performed for the determination of insertion and activation of the gene of interest and its interaction with different plant pathways that may cause unintended

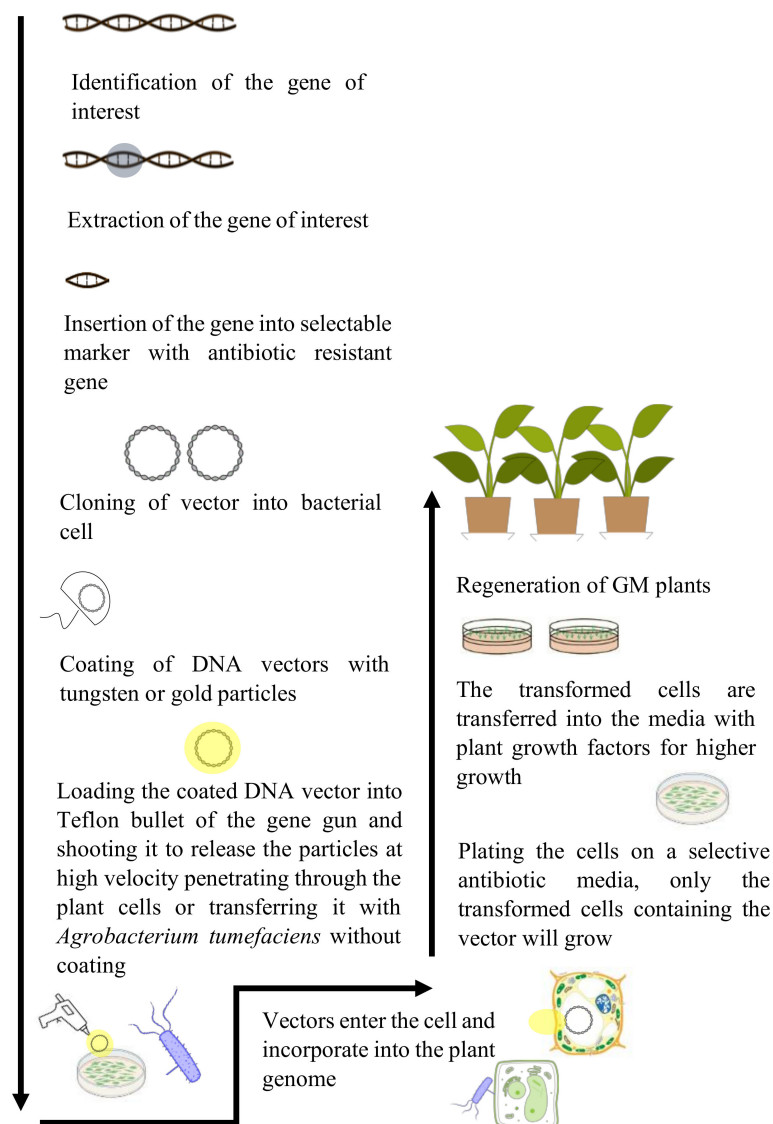


FIGURE 2

Illustration of the process of genetic modification of crops. It involves the identification of gene of interest, its isolation, and insertion into the genome of a desired plant species. The modified plants are regenerated and used for commercialization.

changes in the final traits within the plants (Shrawat and Armstrong, 2018).

The transformed plants are introduced into the field conditions and risk assessments are performed for their environmental and health impacts (Giraldo et al., 2019). Nonetheless, plants with foreign genes have remained in the scrutiny of society for crop production. To overcome these concerns related to transgenic crops, newer biotechnological techniques, such as cisgenesis and intragenesis, are developed as alternatives to transgenesis (Holme et al., 2013; Kumar et al., 2020). In these methods, genetic material used for trait

enhancement are from identical or related plant species with sexually compatible genes.

Besides these techniques, genome editing tools has enabled the plant transformation with ease, accuracy, and specificity. Some of these methods including Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system, were directed towards the concerns related to the unpredictability and inefficiency of traditional transgenesis (Bhardwaj and Nain, 2021). These tools are set for developing enhanced plant varieties through accurate

modification of endogenous genes and site-specific introduction of target genes.

2.2 Status of GM crops

The global production status of GM crops has increased between the year 1996 to 2019, from 1.7 to 190.4 million ha with approximately 112-fold increase (Table 1; Figure 3) (ISAAA, 2019). Subsequently, a large increase occurred in the commercialization of GM crops at an elevated rate in the history of present-day agriculture. Currently, the world's largest GM crops producer is USA with 71.5 Mha (37.5%), with GM cotton, maize, and soybean accounting for 90% of its production (ISAAA, 2019). Brazil was the second largest GM crops producer with 52.8 Mha (27.7%) and Argentina was the third largest producer with 24 Mha. Canada and India were

fourth and fifth largest producers with 12.5 and 11.9 Mha, respectively (ISAAA, 2019).

In 2019, the largest area of GM crops was possessed by soybean 48%, GM maize occupied an area of 60.9 million hectares globally, around 32% of the global maize production (Figure 3) (Turnbull et al., 2021). GM cotton covered 14% of the global area of cotton production in 2019 with 25.7 Mha of area. While GM canola occupied 5% from its 27% of global production in 2019 (Turnbull et al., 2021). In contrast to GM maize, soybean, canola and cotton, some of the GM crops that were planted in different countries also included sugarcane, papaya, alfalfa, squash, apples and sugar beets.

There has been a sharp increase in the approval of the number of plant species with GM varieties. Around 44 countries have provided regulatory acceptance to 40 GM crops and to 509 events of genetic modification since January 2022 (ISAAA,

TABLE 1 The proportion of area covered and common GM crops in various parts of the world.

| No. | Continent | Country | Area (Mha) | Common GM crops |
|-----|---------------|---------------|------------|---|
| 1 | North America | United States | 71.5 | Cotton, papaya, alfalfa, sugar beet, rapeseed, soybean, maize, and squash |
| 2 | South America | Brazil | 52.8 | Soybean, cotton, and maize |
| 3 | South America | Argentina | 24 | Cotton, soybean, and maize |
| 4 | North America | Canada | 12.5 | Soybean, sugar beet, rapeseed, and maize |
| 5 | Asia | India | 11.6 | Cotton |
| 6 | South America | Paraguay | 3.8 | Maize, soybean, and cotton |
| 7 | Asia | China | 2.9 | Tomato, sweet pepper cotton, papaya, and poplar |
| 8 | Asia | Pakistan | 2.8 | Cotton |
| 9 | Africa | South Africa | 2.7 | Cotton, soybean, and maize |
| 10 | South America | Bolivia | 1.3 | Soybean |
| 11 | South America | Uruguay | 1.3 | Maize and soybean |
| 12 | Asia | Philippines | 0.6 | Maize |
| 13 | Australia | Australia | 0.8 | Rapeseed and cotton |
| 14 | Asia | Myanmar | 0.3 | Cotton |
| 15 | Africa | Sudan | 0.2 | Cotton |
| 16 | North America | Mexico | 0.2 | Soybean and cotton |
| 17 | Europe | Spain | 0.1 | Maize |
| 18 | South America | Colombia | 0.1 | Cotton and maize |
| 19 | Asia | Vietnam | 0.1 | Maize |
| 20 | North America | Honduras | < 0.1 | Maize |
| 21 | South America | Chile | < 0.1 | Rapeseed, soybean, and maize |
| 22 | Africa | Malawi | < 0.1 | Cotton, cowpea, and banana |
| 23 | Europe | Portugal | < 0.1 | Maize |
| 24 | Asia | Indonesia | < 0.1 | Cotton |
| 25 | Asia | Bangladesh | < 0.1 | Eggplant |
| 26 | Africa | Nigeria | < 0.1 | Cowpea |
| 27 | Africa | Eswatini | < 0.1 | Cotton |
| 28 | Africa | Ethiopia | < 0.1 | Cotton |
| 29 | North America | Costa Rica | < 0.1 | Soybean and cotton |
| | | Total | 190.4 | |

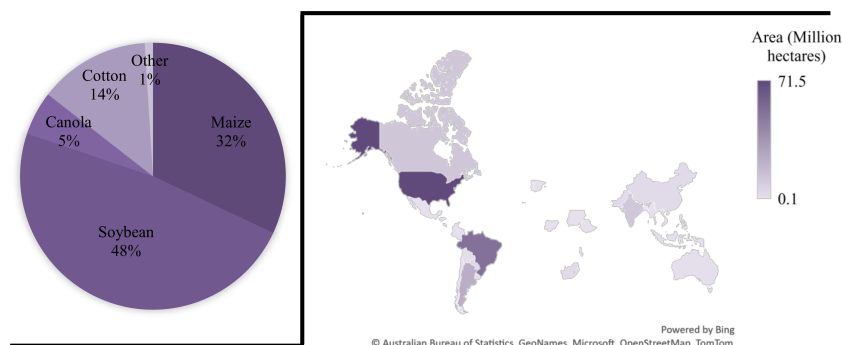


FIGURE 3
Percentage of Globally adopted GM crops and their production area (hectares) in various countries. The largest proportion of GM crops grown are soybeans (48%) and the USA covers a substantial area of 71.5 Mha with different GM crops.

2020). This manipulation includes 41 commercial traits for use in cultivation, food, and feed.

3 Concerns and related issues of GM crops production

The inception of GM crops has been controversial mainly due to the ethical concerns and issues of sustainability surrounding the negative impacts of GM crops. These issues range in different forms such as the detrimental effects of GM crops on the environment and human health, the ideology of creating new life forms within the society, and the intellectual property ownership of GM crops that provides economic benefits to specific people (Oliver, 2014). Most of these issues arise due to the arguments that farmers and seed companies attain the benefits of the GM crops rather than the consumer (Raman, 2017).

3.1 In relation to the environment

The introduction of GM crops may cause adverse impacts on the environmental conditions, which has been raised ethically by certain sections of the society (Figure 4). It has been argued that the GM crops pose a threat to the decline of crop biodiversity due to the hybridization of GM crops with related non-GM crops through the transfer of pollen (Fernandes et al., 2022). The GM crops may become invasive over time and affect the population of local wild crop species. The use of specific chemical herbicides for controlling weeds that grow in the fields with GM crops tolerant to that chemical herbicide will lead to the appearance of highly resistant weeds that will be difficult to control. Due to the high use of chemicals to control those weeds, soil and water degradation

can also occur (Sharma et al., 2022). The use of GM crops can have negative impacts on non-target organisms such as predators and honeybees (Roberts et al., 2020). For instance, the spread of the genetically manipulated herbicide tolerant corn and soybean, with the use of chemical herbicides has damaged the habitat and population of the monarch butterfly in North America (Boyle et al., 2019). It is considered that such environmental risk raised by the GM crops are difficult to be eliminated.

3.2 In relation to the human health

The biggest ethical concern for the genetic modification of crops is their harmful effects on the human beings (Figure 5). It is assumed that consumption of the GM crops can result in the development of certain diseases that can be immune to antibiotics (Midtvedt, 2014). This immunity develops through the transfer of antibiotic resistant gene from the GM crops into humans after the consumption (Midtvedt, 2014). The long-term effects of GM crops are not known, which decreases their consumption rate. It is also found that a number of cultural and religious communities are against these crops and considers them detrimental for humans. It is believed that GM crops can trigger allergic reactions in human beings. In a study conducted for enhancing nutritional quality of soybeans (Glycine max), a methionine-rich 2S albumin from the Brazilian nut was transferred into transgenic soybeans. Since the Brazil nut is a common allergenic food, the allergenicity testing of transgenic soybean indicated allergenic reaction on three subjects through skin-prick testing. This allergenicity was associated to the introduction of 2S albumin gene of Brazil nut into the soybeans (Nordlee et al., 1996; EFSA et al., 2022). There are also assumptions that GM crops can cause the development of cancerous cells in human beings (Touyz, 2013). It is argued that

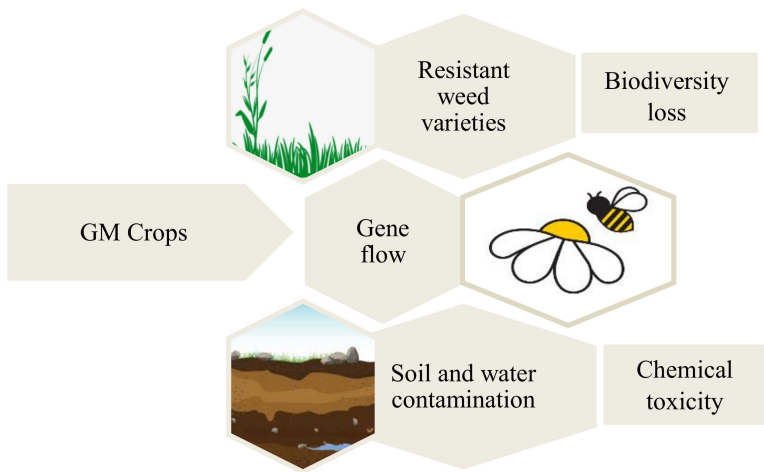


FIGURE 4
Major environmental concern related to the GM crops. The manipulated crops are widely prevented for their gene flow and its detrimental effect on the natural resources and biodiversity.

cancer diseases are caused due to the mutations in the DNA, and the introduction of new genes into human body may cause such mutations (Mathers, 2007). Antibiotic resistance genes from genetically modified plants, used as selectable marker genes can get transferred to bacteria in the gastro-intestinal tract of humans (Karalis et al., 2020). However, the risk of such occurrence is very low, but it has to be considered when assessing the biosafety of the transgenic plants during field trials or commercialization approvals. The health risk of foods

derived from genetically engineered crops are still being debated for rigorous evidence among the scientific community.

3.3 In relation to the development and intellectual property rights of GM crops

In the ethical debate of GM crops for sustainability, the philosophical reasons are fundamental against the development

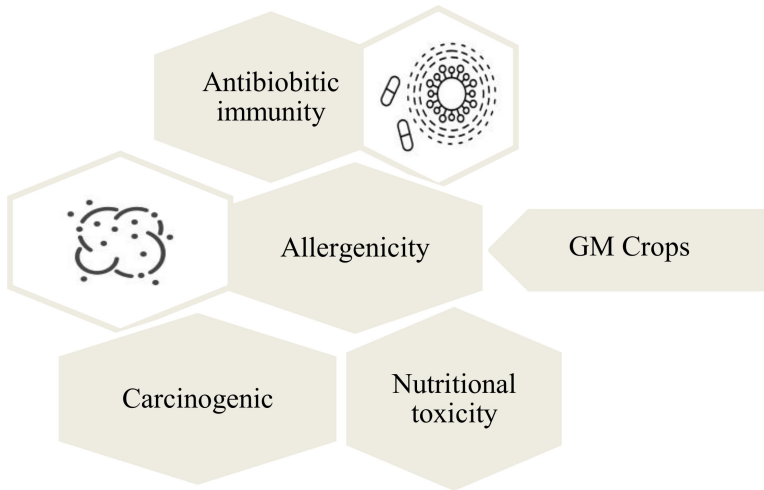


FIGURE 5
Human health related threats of GM crops. The consumption of GM crops is widely associated with toxicity and allergenicity of human beings.

of these crops. It is viewed that genetic modifications of crops are inappropriate interference in the life of an organism (Evanega et al., 2022). The gap in this ethical ideology is aggravated in developing countries due to the prominent role of large Biotech companies in deciding how life forms are to be altered to make benefit from them. The concerns of the intellectual property rights, patents of these crops and their ownerships are at the heart of the ethical issues (Xiao and Kerr, 2022).

The private sector provides the majority of agricultural inputs such as the fertilizers, pesticides and seeds of improved crop varieties that farmers stored and reused season to season (Lencucha et al., 2020). This practice of seed reuse has made it difficult to gain benefits from the investments in artificial breeding. Nonetheless, production of hybrid species and advances in genetic technologies, it became possible to protect the new crop varieties that were developed, especially the larger-volume crops, such as the soybean and maize plants (Liu and Cao, 2014). This is particularly true for the genetic modification tools, which provide producers a stronger intellectual property right for their plants (Brookes and Barfoot, 2020). The patent rights provide monopoly power to the seed companies, which require the farmers to purchase the seeds from the patent owners during each year of plantation (Maghari and Ardekani, 2011). These seeds are known as terminator seeds that develop into infertile crops. The terminator technology was used for developing such seeds that prevented the diversion of genetic modifications to other plants, but limited farmers seed propagation (Niiler, 1999). This made farmers to purchase new seeds during each growing season, giving seed producers larger authority over the utilization of their seeds. It is considered to be ethically wrong to develop plants whose seeds are sterile that farmers cannot use for the second year of plantation (Bawa and Anilakumar, 2013). However, terminator seeds that produced infertile crops were temporarily terminated. The intellectual property rights for the GM crops provided protection to the crop varieties and limited farmers in using the seeds of GM crops for another cycle (Rodrigues et al., 2011). Moreover, intellectual property rights created a barrier for innovation as it provided a limited access to GM crops for several purposes (Redden, 2021).

Despite of these concerns of the GM crops, they are considered as one of the tools for achieving the sustainable food production. However, it needs to be evaluated for possible solutions for their negative impacts in securing their benefits.

4 Potential solutions for growing, commercializing, and incorporating GM crops into sustainable food production systems

The detrimental effects of the GM crops can be reduced or eliminated through appropriate measures that needs to be taken

at different stages of incorporation, marketing and human consumption for ensuring that the GM plants are as harmless as the non-GM crops. This will lead to meeting the goals of sustainability and allow for the incorporation of GM crops into sustainable food production (Figure 6).

4.1 Towards the negative impacts on the environment

One of the major concerns of GM crops is their potential damage to the environment. It affects the environment through the gene flow that occurs from the GM crops to the neighboring non-GM crops *via* the pollen, a phenomenon known as the genetic pollution (Fitzpatrick and Ried, 2019). It is stated that genetic pollution will result in decline of biodiversity. However, the transfer of genes can occur through the pollen of plants at a distance between 50 m to 100 m (Carrière et al., 2021). Therefore, a feasible solution suggested towards the use of GM crops is the practice of growing the GM crops at a distance farther away from the non-GM crops that will lower the chances of gene flow. In addition, such a solution will contribute towards the lowering of crop pollen viability and competitiveness after moving through long distance between the plants (Nishizawa et al., 2010). It was reported in a study that a gene resistant to herbicide from a field of genetically modified oilseed rape moved to the neighboring non-genetically modified oilseed rape (Nishizawa et al., 2010). The investigation from this study indicated that one out of ten thousand oilseed rape contained the modified gene at a distance of 50 m (Nishizawa et al., 2010). Therefore, it is suggested to grow the GM crops at a distance of 50 m away from the non-GM crops during their use in sustainable agriculture, as this practice will reduce the percentage of gene flow (Carrière et al., 2021).

4.2 Towards the negative impacts on the societal and community health

The sustainable agriculture is focused on the health effects of GM crops on the current and future generations. The health effects of GM crops remain an ethical issue that needs to be investigated due to lack of direct studies on the human health effects and the consumption of GM crops (Garcia-Alonso et al., 2022). The possible solutions towards the health effects of GM crops are the constant regulation of these crops through different biosafety testing and risk assessment by health authorities before consumption (Akinbo et al., 2021). The biosafety testing of GM crops should consider the standard that foods developed from GM plants are intended to be as safe as genetically similar varieties of non-GM plants. To date there is no solid evidence that GM crops approved in the US and other countries have harmed humans or animals that had consumed them. This

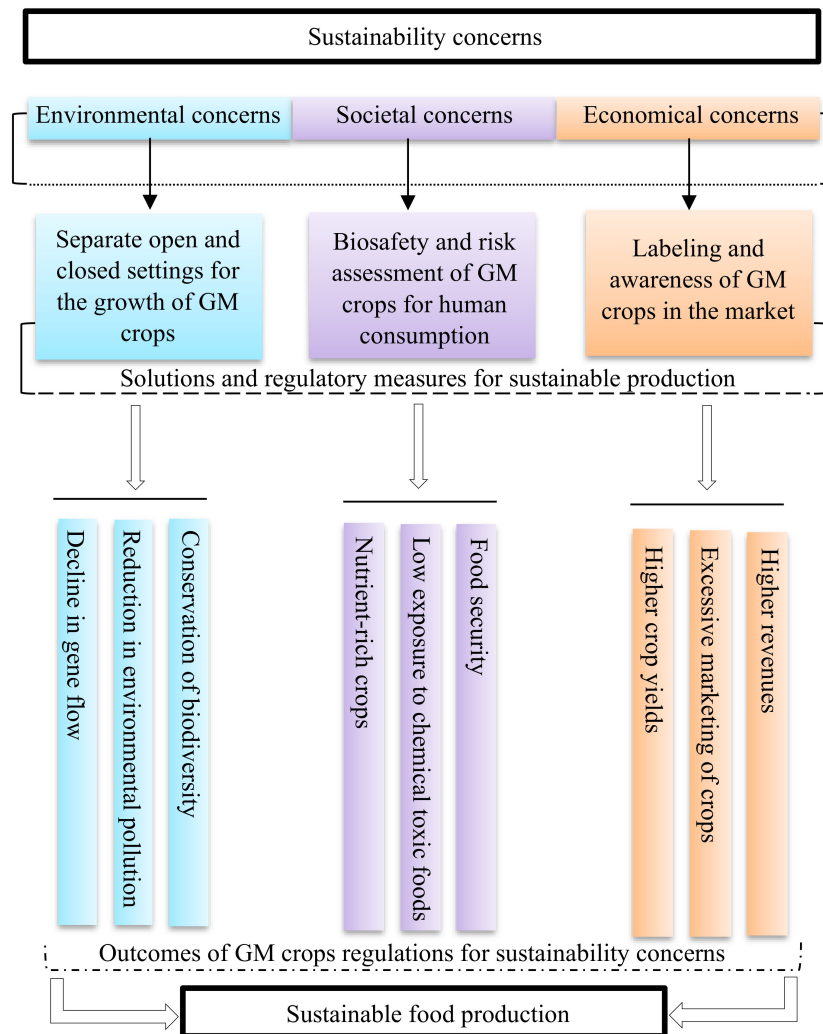


FIGURE 6

Schematic representation of the pathway for countering the concerns of GM crops. The development of separate settings, regulatory framework for biosafety and risk assessments, and commercialization continuum for GM crops will lead to their beneficial impacts, which will result in meeting the goals of sustainability.

highlights that the safety assessment of GM crops is quite robust. However, to predict the adverse effect of GM crops consumption on human health, scientifically sound and long-term studies need to be conducted under controlled and validated experimental conditions on animals such as rats, cows, pigs, etc.

4.3 Towards the negative impacts on the economy

Since the GM crops are passing through various regulatory measures and meeting the testing standards, these crops are still prevented from release to the market (Davison and Ammann, 2017). For instance, with the introduction of new drugs, people

are always given a choice to be the first users or second users, but after certain stages of testing, the drugs are released to the market for the use of everyone, in that ground it would be unethical to prevent the release of GM crops after testing and meeting the regulatory measures (Teferra, 2021). The holding of the release of GM crops to the market prevents the economic benefits that countries can attain through their production. However, a solution has been developed towards such issue is by labelling of the GM crops for the market sales (Delgado-Zegarra et al., 2022). Such labelling's allow for the consumer's sovereignty as the people have the fundamental right to know what food they are consuming and about the processes involved in its production (Yeh et al., 2019). Positive information about GM crops needs to be brought into the public in comparison to the

negative assumptions for improving their marketability. The surveys conducted on the public opinions in a study indicated that majority of the people in USA supports labelling of GM crops (Wunderlich and Gatto, 2015). According to the Food and Drug Administration (FDA), the labelling of GM crops is not to indicate if they are harmful, but rather to describe the attributes of these crops to the public (Borges et al., 2018).

5 Genome editing in the new era as a promising solution for crop manipulation

The scientists have developed advanced molecular tools for the precise modification of plants. Zinc finger nucleases (ZFN) was developed in 2005 with *Nicotiana tabacum* plants for plant trait improvements (Raza et al., 2022). A ZFN is a synthetic endonuclease that is composed of a designed zinc finger protein (ZFP) joined to the cleavage domain of a restriction enzyme (FokI) (Paschon et al., 2019). It can be redesigned to cleave new targets by creating ZFPs with new selected sequences. The process of cleavage event instigated by the ZFN causes cellular repair processes that in turn mediate efficacious manipulation of the desired locus. Within a passage of five years, transcription activator-like nucleases (TALENs) were developed as a new genome editing technique (Raza et al., 2022). Transcription activator-like effector nucleases (TALENs) introduces specific DNA double-strand breaks (DSBs), as an alternative method to ZFNs for genome editing (Forner et al., 2022). TALENs are identical to ZFNs and consists of a non-specific domain of FokI nuclease fused to a changeable DNA-binding domain. This DNA-binding domain possesses highly conserved repeats acquired from transcription activator-like effectors (TALEs) (Tsuboi et al., 2022). These are proteins synthesized by the bacteria *Xanthomonas* to prevent the transcription of genes in host plant cells.

Although these two techniques have modernized plant genomics, each had its own limitations. However, in 2013 emerged the new editing technique named CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats), which provided the plant breeders a widespread ability to make targeted sequence variations, resulting in rapid improvement of crops (Nekrasov et al., 2013). This technique of genome editing uses site-directed nucleases (SDN) to make exceptionally precise incisions at a particular region of DNA (Metje-Sprink et al., 2019; He and Zhao, 2020). SDN techniques are classified into three categories: SDN-1, SDN-2, and SDN-3 (Lusser et al., 2012). The SDN-1 technique instigates a single or double stranded break to remove a part of the DNA, while SDN-2 technique utilizes a small donor DNA template to induce a desired mutation sequence. The third technique, SDN-3 uses a much longer donor DNA template that is introduced into the

target DNA region, which makes this technique similar to traditional recombinant DNA technology (Podevin et al., 2013).

5.1 CRISPR/Cas9 tool for plant genome editing

The CRISPR/Cas system is composed of CRISPR repeat-spacer arrays and Cas proteins. It is a bacterial RNA mediated adaptive immune system that safeguards against bacteriophages and other harmful genetic components by breaking the foreign nucleic acid genome (Hu and Li, 2022). The CRISPR system is based on the RNA-guided interference (RNAi) with DNA (Koonin et al., 2017). This system is divided into two classes based on their Cas genes and the type of the interference complex. Class 1 CRISPR/Cas systems utilize multi-complexes of Cas proteins for interference, while class 2 systems use construct having single effector polypeptides with CRISPR RNAs (crRNAs) to perform interference (Hu and Li, 2022).

In comparison to TALENs and ZFN, CRISPR system can target multiple sites using several single guide RNA's (SgRNAs) with a single Cas9 protein expression (Figure 7). This kind of multiplex editing has sophisticated its use in genome engineering and pyramid breeding (Chen et al., 2019). It can create multigene knockouts and knock-ins, chromosomal translocations and deletions (Salsman and Dellaire, 2016). Various approaches have been employed for multiplex guide RNA (gRNA) expression with one cassette in plants. The editing efficiency can be maintained with one promoter to attain consistent expression of each gRNA by placing it into a small vector (Chen et al., 2019). This has been attained by utilizing a polycistronic gene, having interspersed of gRNA within Csy4 recognition sites, transfer RNA sequences, and ribozyme sites, refined in the cell to produce mature gRNAs for modification (Gao and Zhao, 2013; Xie et al., 2015; Cermák et al., 2017). Moreover, the potential of the discovered new generation of CRISPR nuclease termed as Cpf1 that initiates its own crRNA, has been an efficient system for complex genome editing in crops (Wang et al., 2017).

5.2 CRISPR/Cas9 application for SDGs

In the year 2015, the SDGs were launched. It consisted of 17 SDGs, with enhanced human health, poverty eradication and improved food security being the three important goals (Aftab et al., 2020). All the SDGs have been set for an achievement date of 2030. The successful achievement of these essential and valuable goals requires substantial adoption of technology and innovation. Advancements in plant breeding have resulted in efficient food production systems since the middle 20th century (Smyth, 2022). With further improvement in the current era of agricultural biotechnology through CRISPR/Cas9 system, crop

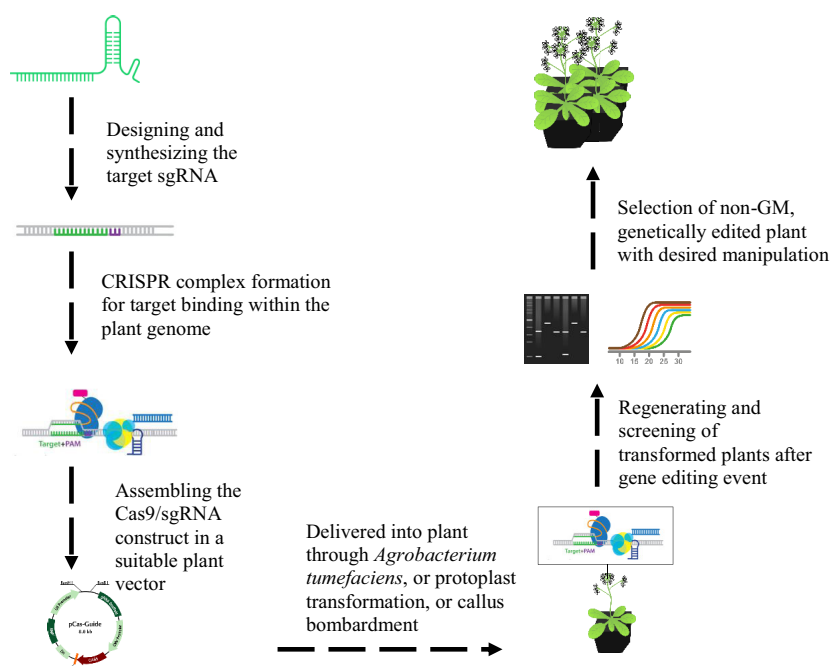


FIGURE 7

Illustration of CRISPR CAS9/sgrNA plant genome editing system. The designing of sgRNA is performed using the available online resources for the target gene. The CRISPR complex is formed with target sgRNA and suitable Cas9 variant, which will be cloned into a plant vector for the target plant species transformation with a suitable technique of transformation. The putative transformed plants will be selected after identifying the Cas9 and target sgRNA based on the screening through PCR or RE genotyping and DNA sequencing. The plants with edited genome will be selected and regenerated.

yield improvements, nutritional enhancements, and reduced environmental impacts are possible (Tripathi et al., 2022). This indicates the potential role of genome editing technologies and highlight their important role in achieving the three essential SDGs. CRISPR/Cas9 technique enhances the sustainability and improves global food security in various ways.

5.2.1 Abiotic stress tolerance

Under abiotic stresses, CRISPR/Cas9 based genome editing is a propitious tool for developing resilient cultivars for sustainable agricultural production. Drought, salinity, and temperature stresses have been significant abiotic stresses studied using CRISPR/Cas9 (Wang et al., 2020). In tomato, two drought stress responsive genes, such the non-expressor of pathogenesis-related gene 1 (*SINPR1*) and mitogen-activated protein kinase 3 (*SLMAPK3*), were knocked out by CRISPR/Cas9 system, with no improvement in drought tolerance observed (Wang et al., 2017b). Salt tolerant alleles were identified in functional analyses of genes related to the reception of salt stress. In tomato, blocking the activity of Salt Overly Sensitive 1 (*SOS1*) gene, which is a Na^+/H^+ antiporter for controlling Na^+ levels in root epidermal cells, resulted in reduced salt tolerance (Wang et al., 2020). Moreover, the precise manipulation of protein domains of tomato hybrid

proline-rich protein1 (HyPRP1), a negative regulator of salt stress responses, provided enhanced salinity tolerance to the edited tomato plants (Tran et al., 2021).

Climatic changes accompanied with large variations in temperature affects cropping. The function of genes in temperature stress reactions are essential for developing and breeding temperature tolerating crops. In relation to this, CRISPR/Cas9 system was used to knockout a chilling related gene and a heat responsive factor, tomato C-repeat binding factor 1 (*SICBF1*) and Brassinazole Resistant 1 (*SIBZR1*). It was shown that these genes were firmly associated with temperature tolerance as the altered alleles of *SICBF1* and *SIBZR1* displayed lessened chilling and heat stress tolerance, respectively (Yin et al., 2018).

In another study, CRISPR/Cas9 triple knockout of *pyl1*, *pyl4*, and *pyl6*, *OsPYL* abscisic acid receptor gene family displayed high-temperature tolerance, increased grain yield, and lower preharvest sprouting in comparison to wild type (Miao et al., 2018). Genome-edited crops contribute towards improved water and nitrogen use efficiency as demonstrated through field trial and experimental data (Wen et al., 2021). Research on enhanced drought tolerance of cotton developed through genome editing tools demonstrated environmental footprint of cotton production (Peng et al., 2021). Arable lands are also heavily

contaminated with metal toxicity. Rice varieties with reduced level of arsenic, radioactive cesium, and cadmium were obtained *via* CRISPR/Cas9 knockout of *OsHAK1*, *OsARM1*, and *OsNramp5* (Nieves-Cordones et al., 2017; Tang et al., 2017; Wang et al., 2017a).

This application of genome editing can aid towards the SGD-13 and SDG-15, which are to promote more environmentally sustainable agriculture. The building of sustainable environment improves the life of organisms on earth. The progression in enhancing the sustainability of present agricultural systems is vital due to the pressures of climatic changes, clearing of forest lands and utilization of arable lands for non-agricultural activities. Without research focuses on genome editing, decline in yields due to the impacts of climatic changes can severely damage food security. Hence, abiotic stress tolerance can also help towards SDG-2 for reducing hunger through food production under various climatic conditions.

5.2.2 Biotic stress tolerance

Crop yields and quality are largely affected by biotic stress factors. Several plants are made resistance to insects, bacterial, viral and fungal diseases through CRISPR/Cas9 knockout. For instance, with CRISPR/Cas9, wheat varieties resistance to one of common fungal disease, powdery mildew has been created *via* knocking out of all six TaMLO alleles responsible for powdery mildew (Wang et al., 2014). Furthermore, CRISPR/Cas9 knockout of *OsERF922*, an ethylene responsive gene, generated blast-resistant rice, resistant to a devastating fungal rice disease (Wang et al., 2016). Crops are also affected by bacterial blight generated by *Xanthomonas oryzae*. In rice plants, excision of *OsSWEET13* promoter resulted in the development of blight resistance plants (Zhou et al., 2015). In relation to viral diseases, CRISPR/Cas9 technique has produced several resistance plants such as tungro disease-resistant rice (Macovei et al., 2018), cotton leaf curl disease-resistant (Iqbal et al., 2016), and broad potyvirus-resistant cucumber (Chandrasekaran et al., 2016). Lately, Lu et al. (2018a) demonstrated that altering *OsCYP71A1* gene resulted in serotonin biosynthesis blockage that heavily increased the level of salicylic acid, resulting in resistance to two destructive plant pests, plant hoppers and stem borers.

Pseudomonas syringae is the causal agent of bacterial speck disease, a major threat to tomato productions (Cai et al., 2011). In an early application, CRISPR/Cas9 was utilized to knockout a positive regulator of downy mildew disease in tomato, which generated tomato mutant alleles ortholog of downy mildew disease resistance in Arabidopsis 6 (*DMR6*). It was found that the mutant lines showed resistance against *P. syringae*, spp. *Xanthomonas* spp. and *Phytophthora capsica* (Paula et al., 2016). The mutant lines were highly useful resources for breeding tomato plants. In another common biotic stress,

susceptibility to *Oidium neolycopersici* infection was associated to few members of the transmembrane protein Mildew Locus O (MLO). It was identified that among the 16 MLOs in tomato, the profound gene was *SIMLO1*, and its innate mutants with loss-of function displayed resistance towards powdery mildew disease (Zheng et al., 2016). The mutant strains generated *via* CRISPR/Cas9 containing homozygous *SIMLO1* alleles, 48-bp truncated versions of the wild *SIMLO1*, exhibited resistance towards *O. neolycopersici* infection. Similarly, Nekrasov et al. demonstrated that CRISPR/Cas9 derived knockout of MLO provided powdery mildew resistance to tomatoes (Nekrasov et al., 2017). It was also found that the *SIMLO1* plants produced through CRISPR/Cas9 technique were devoid of any foreign T-DNA sequence, which made them indistinguishable from natural *SIMLO1* mutant plants (Nekrasov et al., 2017).

In addition to major bacterial, viral, and fungal diseases, CRISPR/Cas9 was applied for other biotic stresses of oomycete infections. In papaya, *Phytophthora palmivora* is a devastating agent of oomycete disease. A papaya mutant plant was developed with a functional cysteine protease inhibitor (*PpalePIC8*) that resulted in enhanced *P. palmivora* resistance (Gumtow et al., 2018). Similarly, cocoa beans have been made resistance towards another oomycete pathogen, *Phytophthora tropicalis*, *via* the CRISPR/Cas9 system (Fister et al., 2018).

Similar to abiotic stress tolerance generating biotic stress tolerance can also lead to SDG-15, as it will lead enhanced living condition for the plants. This will also result in creation of a sound environment for different organisms that depend on plants for their survival and growth.

5.2.3 Crop yield enhancement

The genome editing tools are employed primarily for improving crop yield which is a composite characteristic that relies on various components. CRISPR/Cas has been used to knock-out negative regulators that influences yield controlling factors such as grain weight (*TaGW2*, *OsGW5*, *OsGLW2*, or *TaGASR7*), grain number (*OsGn1a*), panicle size (*OsDEP1*, *TaDEP1*), and tiller number (*OsAAP3*) for achieving the contemplated traits in plants with loss-of-function alteration in these genes (Li et al., 2016; Li et al., 2016; Zhang et al., 2016; Liu et al., 2017; Zhang et al., 2018a; Lu et al., 2018b). In rice, using CRISPR system, simultaneous knockout of various grain weight related genes (*GW2*, *GW5*, and *TGW6*) led to trait pyramiding that efficiently increased grain weight (Xu et al., 2016). Huang et al. (2018) recently combined CRISPR/Cas9 with pedigree analysis and whole-genome sequencing for the large identification of genes that were responsible for composite quantitative traits, including yield. The study analyzed 30 cultivars of the Green Revolution miracle rice cultivar (IR8) and identified 57 different genes in all high-yielding lines, to be used for gene editing *via* Cas9 knockout or knockdown system. Phenotypic trait analysis indicated the role of most of these

genes in determining yield of rice. It laid insight on yield improvement and facilitated the molecular breeding of improved rice varieties.

A high yielding commercial corn was produced by DuPont Pioneer through the CRISPR/Cas9 knockout in waxy corn line (Waltz, 2016a). Genome editing techniques are also used to develop semi-dwarf corn varieties having higher production and low height, in order to lower moisture and nutrient requirements of the corn (Bage et al., 2020). Moreover, in maize, multiple grain yield traits were enhanced by creating fragile promoter alleles of *CLE* genes, and a null allele of a recently spotted partially redundant recompensing *CLE* gene, utilizing CRISPR/Cas9 technique. Considerable gene editing research is being undertaken on wheat for increased yield, seed sizes, and seed weight (Li et al., 2021b). Although the future of plant genome editing remains uncertain in Europe, researchers of Vlaams Instituut voor Biotechnologie (VIB) of Belgium have currently applied to undertake three genetically edited corn varieties for field trials that have higher yields and enhanced digestibility (VIB, 2022).

Most of the genome edited crops for yield enhancement will lead to increased farm and household revenues. This results in reducing poverty. Although few studies are conducted to date on this discrete goal measurement. It has been reported in one study that the acquisition of Bt cotton developed through transgenesis approach in India, has increased the income by 134% for farmers living with less than 2 USD/day (Subramanian and Qaim, 2010). This was mainly due to improved yields and lowered inputs costs. The potential of genome editing for yield increases indicates that, similar to GM crop adoption, genome edited crops can also improve the incomes of the farmers. The early evidence related to possible increase in the household income due to yield increases, indicates that genome editing makes significant contributions to SDG-1 for eradicating poverty. The significant genome editing studies for increasing the yield of major food staple crops and other essential crops indicate the substantial potential of GMs in contributing towards SDG-2, which aims to end hunger and achieve food security.

5.2.4 Quality improvement

The quality of crops may differ depending on the various breeding techniques used. The genome editing has impacted several quality traits such as nutrition, fragrance, starch content and storage quality of crops. Using CRISPR/Cas9, the knockout of *Waxy*, resulted in enhanced rice eating and cooking quality with low amylose content (Zhang et al., 2018b). Resistant starch rich varieties with elevated amylose were developed by altering the starch connecting enzyme gene, *SBEIIb*, by CRISPR/Cas9. Consuming food with increased amylose content is essential for the patients with diet-related to noninfectious chronic diseases (Sun et al., 2017). Another important quality for the commercial

and edible rice varieties is the fragrance. The biosynthesis of a major rice fragrant compound, 2-acetyl-1-pyrroline is due to a variation in the betaine aldehyde dehydrogenase 2 (*BADH2*) gene. With TALEN genome editing tool, specific alteration of *OsBADH2* resulted in a fragrant rice variety with low 2-acetyl-1-pyrroline content identical to the innate fragrant rice variant (Shan et al., 2015).

In Western countries, celiac disease is triggered due to cereal crops Gluten protein in more than 7% of individuals. Wheat plant consists of nearly 100 genes or pseudogenes, α -gliadin gene family, for gluten-encoding. CRISPR/Cas9 system allows for newer pathways to modify traits governed by massive gene families with unessential properties. At present, researchers have created low-gluten wheat by simultaneous knockout of most conserved domains of α -gliadin family (Sanchez-Leon et al., 2018). Furthermore, other high-quality plants produced by CRISPR/Cas9 includes *Camelina sativa* (Jiang et al., 2017) and *Brassica napus* (Okuzaki et al., 2018) plants with high oleic acid oil seeds, long shelf-life varieties of tomato (Li et al., 2018a), enhanced lycopene tomatoes (Li et al., 2018) or γ -aminobutyric acid content in tomatoes (Li et al., 2018b), and resulted in low levels of toxic steroidal glycoalkaloids in potatoes (Nakayasu et al., 2018). The increased lycopene production acts as an antioxidant for lowering the risk of cancer and heart diseases (Zaraska, 2022). Recently in UK, Rothamsted Research has received acceptance for field trials of genetically edited wheat that synthesizes lower asparagine, a potential cancer producing compound in the toasted breads (Case, 2021).

Genome editing applications surrounding the quality improvement has the prospective to make considerable contributions to SDG-3. Quality improvements in crops promote human health and well-being. Moreover, the capability of genome editing in producing food that may avert specific diseases are directly associated with beneficial health implications.

5.2.5 Nutritional enhancement

One of the applications of genome editing is to enhance the nutritional metabolism and decrease the undesirable substances from the crops through gene expression regulation. In 2021, Japan launched the first genome-edited tomato Sicilian Rouge High GABA (gamma-aminobutyric acid). The edited variety has around four to five times higher amount of GABA than the ordinary tomatoes. The increase was the result of CRISPR/Cas9 genome editing that targeted the autoinhibitory domain (AID) of GAD3 on the C-terminal side, an enzyme involved for the GABA biosynthesis (Nonaka et al., 2017). A frameshift mutation was induced in this autoinhibitory domain that caused the early termination of translation, and the excision of autoinhibitory domain of GAD3 (Nonaka et al., 2017). This strategy eliminated the inhibitors of GAD3 and increased the enzymatic activity involved in the GABA biosynthesis, whose activity is generally suppressed without manipulating the expression level of GAD3.

Furthermore, CRISPR/Cas9 system was also utilized to improve the total wheat protein content with enhanced grain weight with the knockout of *GW2* gene that encodes for a RING-type E3 ubiquitin ligase, known to govern the cell numbers of spikelet hulls (Zhang et al., 2018a). Moreover, genome editing was applied to lettuce that has produced a new variant synthesizing enhanced levels of thiamine, β -carotene, and vitamin C (Southy, 2022).










Research is additionally focused on enhancing corn vitamin A content and provitamin A (Maqbool et al., 2018; Xiao et al., 2020). In the US, a genome editing study was targeted on increasing wheat fiber content. The research is underway for the field trials of this new enhanced fiber wheat (Knisley, 2021). Ensuring sufficient nutrient content in human diets enhances life-long health benefits and prevents the

debilitating diseases. The genome editing tools promising results in broad applications of nutritional enhancements is essential for food insecure developing countries. With this application, the genetic editing underpins the other portions of SDG-2 and SDG-3, which are to achieve and consume fortified nutritional food.

5.2.6 Enhancing hybrid breeding

Hybrid breeding is an appropriate method for enhancing crop productivity. A male-sterile maternal line is essential for producing an improved-quality hybrid variety. Through CRISPR/Cas9 technique, tremendous progress has been made to produce male-sterile lines, which includes photosensitive genic male-sterile rice (Li et al., 2016), heat sensitive male-sterile lines in rice (Zhou et al., 2016), wheat (Singh et al.,

TABLE 2 Overview of the recent CRISPR/Cas9 applications for the SDGs.

| Applications | Plants | Target genes | Traits | SDGs | References |
|---------------------------------------|-------------|-------------------------|---|---|----------------------------|
| Abiotic and Biotic stresses tolerance | Arabidopsis | HSFA6a and HSFA6b | ABA and osmotic tolerance |  | (Wenjing et al., 2020) |
| | Arabidopsis | AITR | Drought tolerance | | (Chen et al., 2021) |
| | Rice | OsSAP | Drought tolerance | | (Park et al., 2022) |
| | Rice | OsBHLH024 | Salinity tolerance |  | (Alam et al., 2022) |
| | Rice | OsERA1 | Drought stress | | (Ogata et al., 2020) |
| | Rice | OsSWEET14 | Bacterial blight resistance | | (Zafar et al., 2020) |
| | Soybean | F3H1, F3H2, and FNSII-1 | Mosaic virus resistance |  | (Zhang et al., 2020) |
| | Soybean | GmAIR | Salinity tolerance | | (Wang et al., 2021) |
| | Chickpea | 4CL and REV7 | Drought tolerance | | (Razzaq et al., 2022) |
| | Tomato | SIMAPK3 | Heat tolerance | | (Yu et al., 2019) |
| Enhanced yield | Barley | HvARE1 | Nitrogen use efficiency | | (Karunaratne et al., 2022) |
| | Banana | DMR6 | Downy mildew resistance | | (Tripathi et al., 2021) |
| | Potato | StDND1 and StCHL1 | Late blight resistance | | (Kieu et al., 2021) |
| | Rapeseed | BnaMAX1 | Improved productivity |  | (Zheng et al., 2020) |
| | Wheat | GW7 | Increase in grain size and weight | | (Wang et al., 2019b) |
| | Rice | GS3 | Improved productivity | | (Huang et al., 2022) |
| | Soybean | GmFT2a and GmFT5a | Increased pod and seed size |  | (Cai et al., 2020) |
| | | | | | |
| | | | | | |
| | | | | | |
| Enhanced Quality | Rice | GW2 | Enhanced protein |  | (Achary and Reddy, 2021) |
| | Rice | OsGAD3 | Increased GABA | | (Akama et al., 2020) |
| | Rice | GBSS | Low amylose | | (Huang et al., 2020) |
| | Wheat | TaSBEIIa | High amylose |  | (Li J et al., 2021a) |
| | Wheat | CM3 and CM16 | Reduced allergens | | (Camerlengo et al., 2020) |
| | Wheat | TaASN2 | Reduced asparagine | | (Raffan et al., 2021) |
| | Potato | SBE1 | Reduced asparagine |  | (Tuncel et al., 2019) |
| | Barley | D-hordein | Lowered prolamine and enhanced glutenin | | (Yang et al., 2020) |
| | Rapeseed | BnITPK | Reduced phytic acid | | (Sashidhar et al., 2020) |
| | | | | | |
| Hybrid breeding | Rice | Zep1 | Improves the frequency of genetic recombination |  | (Liu et al., 2021) |
| | Wheat | ZIP4-B2 | Recombination of homologous chromosomes | | (Martín et al., 2021) |

2018), and corn (Li et al., 2017a). Heterosis in breeding faces hybrid sterility as an obstacle. Reproductive barriers were disrupted in hybrids between japonica and indica, SaF/SaM (sterility locus Sa) (Xie et al., 2017a) and African rice (*Oryza glaberrima* Steud) OgTPR1 (sterility at the S1 locus) (Xie et al., 2017b). It was found that knockout of the indica Sc gene in the allele Sc-I protected the male fertility in japonica-indica hybrids (Shen et al., 2017). Likewise, it was shown that knockout of the toxin gene *ORF2*, improved the fertility of japonica-indica hybrids (Yu et al., 2018). Furthermore, in rice plants, genetic editing was utilized to replace mitosis for meiosis, through the knockout of three important meiotic genes, PAIR1, OSD1, and REC8 (Mieulet et al., 2016). Moreover, simultaneous activation of BBM1 in egg cells or knockout of MTL, by two independent research groups resulted in asexual propagation lines that fixed the hybrid heterozygosity through seed propagation (Khanday et al., 2018; Wang et al., 2019a). In addition, gene editing is also a constructive method for enhancing haploid breeding (Yao et al., 2018), shortening growth periods (Li et al., 2017b), improving resistance to silique shatter (Braatz et al., 2017), and countering the self-incompatibility of diploid potatoes (Ye et al., 2018), that meets the requirements of breeders. The enhanced breeding of hybrid plants results in the developing of novel plant varieties that supports the SDG-15, enhancing life on land through diverse plant species. Therefore, the successful application of genome editing technologies have modified and improved many essential traits in diverse crops for the achievement of different SDGs (Table 2).

5.3 Regulatory concerns of crop genome editing

The recent developments in biotechnology in the form of genome editing has made it viable for food products to get into the market quicker in a feasible rate. The latest genome editing tools are essential for the future production of crops. This is due to their robustness, process precision and timely regulation in comparison to conventional GM crops. Several products are now developed through CRISPR/Cas9 system that are not considered as GMO in several countries. It was stated by the US Department of Agriculture (USDA) that the crops edited *via* CRISPR/Cas9 platform can be grown and marketed without regulatory processes and risk assessments that are mandatory on GMOs biosafety regulations (Waltz, 2016b). Such a step will save millions of dollars spent on investigating GM crops through field tests and data collections, reduces the time required for introduction of improved crop varieties into the market, and removes the uncertainty associated with the consumption of GM crops within the public. To date, five crops developed through CRISPR/Cas9 system were accepted by the USDA without the regulatory measures of GMOs. These includes browning-resistant mushrooms, created through CRISPR/Cas9 technique

by the knockout of polyphenol oxidase (*PPO*) gene (Waltz, 2016b). Likewise, waxy corn plants with enhanced amylopectin have been developed by CRISPR/Cas9 system with the inactivation of an endogenous waxy gene (*Wx1*) and introduced without regulations (Waltz, 2016a). *Setaria viridis* with delayed flowering period was attained through the deactivation of the *S. viridis* homolog of the corn *ID1* gene (Jaganathan et al., 2018), camelina altered for improved oil content (Waltz, 2018), and soybean with modified *Drb2a* and *Drb2b* for drought tolerant, were not subjected to GMO regulatory measures (Cai et al., 2015; Kumar et al., 2020).

6 Perspectives on the criticisms of GM crops incorporation into sustainable food production systems

Agriculture plays an important role towards the SDGs achievement, such as for reducing hunger and malnutrition, alleviating poverty, implementing a sustainable production and consumption system, countering climatic changes, ensuring gender equality, improving energy use, and maintaining healthy ecosystem services (Viana et al., 2022). It acts as a basis for economic development in several countries. Global agriculture has successfully provided sufficient food for meeting the rising demand and varied consumption patterns of humans over the recent decades (da Costa et al., 2022). This has been possible largely due to the agricultural intensification at the expense of environmental resource degradation, biodiversity loss, harmful gas emissions, and land clearing (Liu et al., 2022). However, it has been shown that the advancement of the biotechnological tools for genetic modification of crops will allow agricultural practices to achieve SDGs in a sustainable manner. Nonetheless, GM crops faces the moral and ethical dilemma of their incorporation into the sustainable agricultural practices, which can be negotiated through the appropriate balance of benefits and negative impacts of GM crops by encompassing all the three relational aspects of sustainability, such as the environment, society, and the economy (Figure 8).

6.1 GM crops for sustainable environment

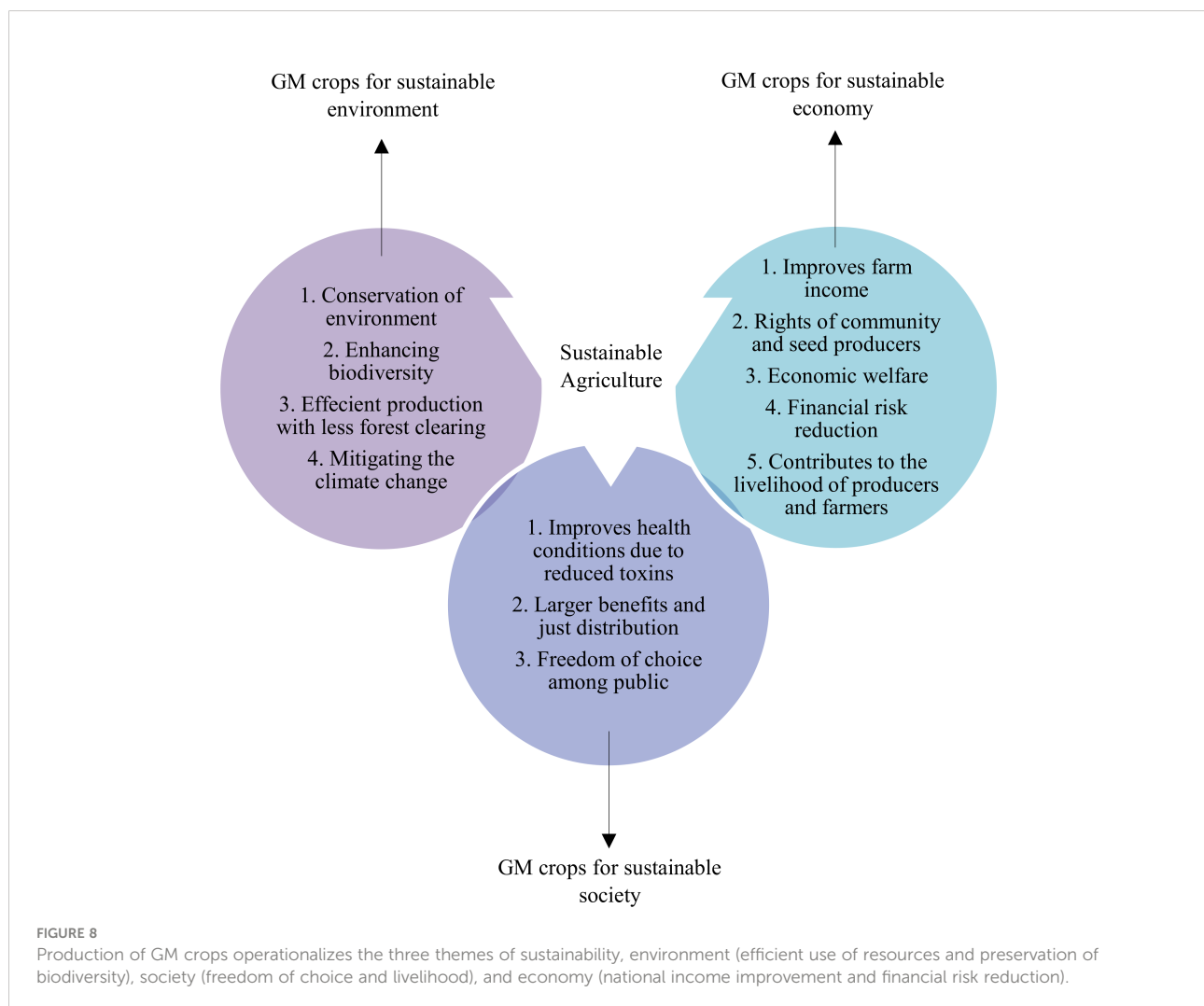
GM crops are scrutinized for the environmental safety. More than 300 million EUR were invested by the European Union (EU) in 130 research projects. It covered a research period of more than 25 years specifically to reach at the interpretation that GM crops are not riskier compared to conventional bred plants (European Commission, 2010). In fact, GM crops that were developed for input traits such as insect resistance and herbicide

tolerance have resulted in a reduction to agriculture's environmental footprint by enhancing sustainable farming practices (Brookes and Barfoot, 2015). Moreover, the genetic modification of crops is a logical continuation of selective plant breeding that humans have developed for thousands of years. It results in the conservation of environment and the plant biodiversity allowing their incorporation into sustainable food production systems. Klumper and Qaim undertook a meta-analysis of the initial data obtained from the farm surveys or field trials in various parts of the world (Klumper and Qaim, 2014). It indicated that the insect resistance of GM crops has lowered the pesticide application by 36.9%.

In addition, GM seeds contribute towards the adoption of conservation tillage, which are sown straight into the fields without early ploughing. This practice conserves the essential soil microorganisms, preserves soil moisture, and maintains carbon in the soil. A meta-analysis was performed by Abdalla et al. (2016) to compare the CO₂ emissions of entire season from the tilled and untilled soils. It was found that on average, 21%

more carbon was emitted from the tilled soils than the untilled soils. Furthermore, the use of powered agricultural machines was lowered due to less pesticide use and no/less field ploughing. This provides indirect benefits to sustainable agriculture by conserving fossil fuels and decreasing the emission of CO₂ into the atmosphere. In the United States, the land area for the soybean production increased by approximately 5 million hectares between 1996 and 2009, and 65% of those fields were that of no-tillage practices field due to the adoption of GM soybeans (Brookes and Barfoot, 2016). This resulted in a decline of fuel utilization of 11.8%, from 28.7 to 25.3 liters per hectare, and an approximate reduction of greenhouse gas emissions of more than 2 Gt between 1996 and 2009. The genetically modified soybean fields presented similar impacts of reduction in greenhouse gas emission in various countries such as, Uruguay, Argentina and Paraguay (Brookes and Barfoot, 2016).

The cultivation of GM crops has also increased the biodiversity of non-target beneficial insects due to the lack of chemical use in the fields for the control of harmful insects (Karalis et al., 2020;



Talakayala et al., 2020). The pest resistant traits of GM crops allow the restoration of crop species that were discontinued due to harmful insect pressure. In addition, it improved the crops adaption to several environmental conditions, allowing for a diversified production practice (Anderson et al., 2019). Despite the argument that GM crops threaten biodiversity, it is found that different practices of agriculture affect the biodiversity, and the GM crops do not broaden this threat.

Agriculture causes significant clearance of natural habitat for the food production (Mrówczyńska-Kamińska et al., 2021). However, it was indicated that the high yields of GM crops were achieved at lower land areas (Burney et al., 2010). The improved productivity also reduces the pressure of converting additional land for agriculture (Bouët and Gruère, 2011). Genetic modification reduces habitat destruction, which is a common practice of intensive farming that poses a large threat to biodiversity. For instances, without the use of GM crops, an additional 22.4 Mha would have been needed for maintaining the global production at 2016 levels (Brookes and Barfoot, 2018).

GM crops are considered as unique species that pose a threat through movement of their genes (Raman, 2017). At present there are no scientific manifestation of hazards associated with the transfer of genes between unrelated organisms developed through genetic alterations. Different scientific corporations such as the U.S. National Academy of Sciences, World Health Organization (WHO), and the British Royal Society have stated that consumption of GM foods is not as harmful as consuming the same foods that were modified using conventional crop improvement techniques. Therefore, the GM crops cannot be prevented for use in sustainable food production systems.

6.2 GM crops for sustainable society

The adoption of GM crops has significant health benefits. It reduces the exposure to harmful chemical pesticides that are used with non-GM crops (Smyth, 2020b). Two decades analysis of GM corn consumption by Pellegrino et al. (2018) indicated that it posed no threat to the health of human or livestock. It showed a substantive positive impact on health due to the presence of lower mycotoxins in crops (Pellegrino et al., 2018). The emergence of new genetic modification technologies enabled the production of crop varieties with enhanced flavors and reduced allergens (Mathur et al., 2015). Moreover, the prospective production of edible vaccine in GM crops can result in low-cost vaccine production and allow for their accessibility to a larger section of the society. The pre-testing for the safety of GM crops in several areas has indicated no evidence of any adverse reactions (Kamle et al., 2017). Although the negative health consequences of GM crops consumption are reported on rats, analyses of most of the studies about the safety of GM crops, indicated no human health consequences (Szymczyk et al., 2018; Giraldo et al., 2019).

The sustainable food production systems need to ensure food security for the growing population. Since most of the countries depend on the food imports for their supply's due to the climatic constraints and the insect pests, food security appears difficult (Xiao et al., 2020). However, GM crops climatic stress tolerance and higher yields will ease the process of achieving food security (Evanega et al., 2022; Keiper and Atanassova, 2022). Therefore, including the GM crops into the sustainable food production systems will enable different communities to produce their own food. Moreover, the GM crops are developed with improved shelf-life that can be stored for longer periods without wastage. Such practices appeal to the ethical principles of beneficence and justice, which means to have fair and equitable food supply that will benefit the larger society (Smyth, 2020a; MatouskovaVanderberg, 2022; Vega Rodríguez et al., 2022).

The genetic modification of crops further provides the different nutrients required for healthy human living. Kettenburg et al. (2018) depicted the evidence of health gains from the Bt maize crops and Golden Rice that produces Vitamin A for human beings. It has been reported that around 1 million children die annually due to the Vitamin A deficiency (Swamy et al., 2019). Therefore, the production of Golden Rice plays an important role in preventing these deaths of children. Hence, the introduction of GM crops can save human lives. The potential risks of GM crops that are not proven remain insignificant for people who are starving or having severe nutrient deficiencies (Vega Rodríguez et al., 2022). People with life threatening disease deploy themselves to experimental drugs, which is considered ethical after a consent, the same could be applied to the GM crops.

The experts from governmental and non-governmental agencies in some of the developing countries have increasingly included the GM crops into the wider approaches of sustainability (Hartline-Grafton and Hassink, 2021). However, there are certain people within different communities who still resist the GM crops because of the personal and religious beliefs (Bawa and Anilakumar, 2013). It includes the concern over the right to "play God", as well as the introduction of any foreign gene into crops that are abstained for religious reasons (Omobowale et al., 2009). It is believed that it is intrinsically wrong to tamper with nature, and others consider inserting new genes into plant genome as unethical (Daunert et al., 2008). However, such an issue can be addressed through genome editing techniques and with the contrasting view that the genetic modification is simply one more step in the processes of modification of the physical world. It is similar to the manufacturing of novel chemicals in industries and to natural breeding of plants and animals (Yang et al., 2022). As people are having a choice to use different novel chemicals, similarly a right to choose can be developed for the GM crops consumption. Moreover, the science and technology have advanced humans in putting adequate measures to evaluate and monitor scientific innovations to prevent potential risks to the society (John and Babu, 2021). Therefore, it is of support to use GM crops in

sustainable food production systems, as the development of GM crops is identical to any another scientific invention.

6.3 GM crops for sustainable economy

The economical aspect of GM crops faces the issue of intellectual property rights (Delgado-Zegarra et al., 2022). The producers of GM crops have used terminator technology to protect their seeds and reduce the gene flow. The seeds and pollens of these crops are made sterile (Turnbull et al., 2021). After the completion of harvest, farmers have to re-purchase the seeds from the seed producers. It has been argued that such a technique provides seed companies more control over what the farmers should grow, and it is considered to be unethical by the society (Delgado-Zegarra et al., 2022). But, from an innovation perspective, it is ethical to protect intellectual properties, because these seeds are the obtention of biotechnological companies, they need to have the same intellectual property protection rights as any other potential product, such as the protection of a new software developed by an IT company (Muehlfeld and Wang, 2022). However, it is due to the negative publicity of GM crops that they are held back by the public. There are also very few farmers that depend on second-generation seeds. Hence, the introduction of sterile seeds does not affect the farmer's seed choices (Addae-Frimpomaah et al., 2022). Many of the GM seed manufacturers developed a solution towards sterile seeds through the creation of seed contract with the farmers. The seed contract is an agreement that states that the GM seeds are sterile and are used by farmers on their own choice and that the seeds shouldn't be distributed for any other purposes. This has resulted in economic benefits to the seed producers in an ethical way through the farmer contract agreement.

The farmers have also benefited economically with the adoption of GM crops (Raman, 2017). With the introduction of GM crops, there would be a major increase in the farmer's production efficiency which in turn results in higher revenue (Oliver, 2014). Since the GM crops are made resistant to pests, the cost spent on chemical pesticides will decline, as less chemicals will be required for the GM crops (Buiatti et al., 2013). Furthermore, the use of farm machineries declines as well due to no-tillage practices with GM crops that reduces the cost spent on the fuel of machineries. In addition to this, the land cost for the growers can decline using GM crops as these crops produce high yield in small spaces. Moreover, the poor farmers are mostly engaged in subsistence farming, but the adoption of GM crops would enable such farmers to market their products due to surplus yields from GM crops (Azadi et al., 2016), which would improve their quality of life within the society (Lucht, 2015). The farmers have well-adopted GM crops into the food production systems. Since mid-1990s, GM crops were planted by 18 million farmers (ISAAA, 2017). The track records indicated logistical and economic advantages to the farmers. A net economic benefit of USD 186.1 billion within the twenty-one years of GM crop use was recorded in various farms. It was

found by Brookes and Barfoot (2018) that 52% of these benefits were reaped by farmers from developing country. The majority of these gains (65%) were mainly due to yield and productivity increases, while the remaining (35%) resulted from the cost savings.

7 Conclusion

The practice of sustainable agriculture has become challenging due to the changes in climate, the rising population, and shrinkage of arable lands. There is a need to develop modified crops having higher productivity, quality, and tolerance to various biotic and abiotic stresses. The genetic modification of crops has enabled the development of efficient production systems that provided substantial benefits to the producers and the community based on the three principles of sustainable agriculture such as protecting environment, enhancing human health, and improving the economy.

Even when there are strict assessments of environmental and health safety, and these crops are granted regulatory approval, concerns are still raised over the involvement of the genetic modification tools and their long-term unknown disadvantages on environment and health. The potential negative consequences of GM crops have caused to their lesser implementation in various countries. To overcome and address some of these concerns, new advanced alternative molecular techniques are developed, such as genome editing, particularly CRISPR/Cas9 system that improves crop traits without introducing foreign genes. The expansion of plant breeding to genetic modification through genome editing would further produce more per unit of land that makes them essential in achieving the SDGs, especially for eradicating hunger, improving food security and human health.

The present review indicates that it would be imprudent to dismiss GM crops as a tool for meeting the goals of sustainable development. With the increasing global challenges, GM crops can help humanity. However, it is imperative that the scientific community and agricultural industries invest in better communications and regulations to counter the misinformation and unethical research associated with GM crops. Moreover, this review suggests that GM crops can be broadly adopted by improving the already present regulations, efficient monitoring, and practice implementation through government agriculture bodies. In addition, developing a global risk alleviation strategy and communication with growers, will ensure a substantial acceptance and adoption of GM crops in several countries to bring global profitability and productivity.

Finally, the sustainability of GM crops should be determined based on their role in sustainable agriculture and human development within the next 30 years. It is not only GM crops that pose certain risks and concerns, but all the methods of food production are associated with some drawbacks. However, the use of genome editing tools and regulation of GM crops ensure

that these crops are as safe as conventionally bred crops and can act as the drivers of sustainable food security.

Author contributions

KM and MA conceptualized and wrote the original manuscript. KM, MA, FB, and HR. reviewed and edited the draft paper. All the authors have read and approved to the submitted version of the manuscript.

Funding

This research work was supported by funding from the United Arab Emirates University, the Research Office to KM under grant number 31R203.

References

- Abdalla, K., Chivenge, P., Ciais, P., and Chaplot, V. (2016). No-tillage lessens soil CO₂ emissions the most under arid and sandy soil conditions: Results from a meta-analysis. *Biogeosciences*. 13, 3619–3633. doi: 10.5194/bg-13-3619-2016
- Abdul Aziz, M., Brini, F., Rouached, H., and Masmoudi, K. (2022) Genetically engineered crops for sustainably enhanced food production systems. *Front. Plant Sci.* 13, 1027828. doi: 10.3389/fpls.2022.1027828
- Achary, V., and Reddy, M. (2021). CRISPR-Cas9 mediated mutation in GRAIN WIDTH and WEIGHT2 (GW2) locus improves aleurone layer and grain nutritional quality in rice. *Sci. Rep.* 11, 21941. doi: 10.1038/s41598-021-00828-z
- Addae-Frimpomaah, F., Amenorpe, G., Denwar, N., Amiteye, S., Adazebra, G., Sossah, F., et al. (2022). Participatory approach of preferred traits, production constraints and mitigation strategies: Implications for soybean breeding in Guinea savannah zone of Ghana. *Heliyon*. 8 (5), e09497. doi: 10.1016/j.heliyon.2022.e09497
- Aftab, W., Siddiqui, F., Tasic, H., Perveen, S., Siddiqi, S., and Bhutta, Z. (2020). Implementation of health and health-related sustainable development goals: Progress, challenges and opportunities - a systematic literature review. *BMJ Glob Health* 5 (8), e002273. doi: 10.1136/bmjgh-2019-002273
- Akama, K., Akter, N., Endo, H., Kanesaki, M., Endo, M., and Toki, S. (2020). An *in vivo* targeted deletion of the calmodulin-binding domain from rice glutamate decarboxylase 3 (OsGAD3) increases γ -aminobutyric acid content in grains. *Rice*. 13, 20. doi: 10.1186/s12284-020-00380-w
- Akinbo, O., Obukosia, S., Ouedraogo, J., Sinebo, W., Savadogo, M., Timpo, S., et al. (2021). Commercial release of genetically modified crops in Africa: Interface between biosafety regulatory systems and varietal release systems. *Front. Plant Sci.* 12, 605937. doi: 10.3389/fpls.2021.605937
- Alam, M., Kong, J., Tao, R., Ahmed, T., Alamin, M., Alotaibi, S., et al. (2022). CRISPR/Cas9 mediated knockout of the OsbHLH024 transcription factor improves salt stress resistance in rice (*Oryza sativa* L.). *Plants (Basel)* 11 (9), 1184. doi: 10.3390/plants11091184
- Anderson, J., Ellsworth, P., Faria, J., Head, G., Owen, M., Pilcher, C., et al. (2019). Genetically engineered crops: Importance of diversified integrated pest management for agricultural sustainability. *Front. Bioeng Biotechnol.* 7, 24. doi: 10.3389/fbioe.2019.00024
- Ayub, R., Guis, M., Ben Amor, M., Gillot, L., Roustan, J., Latché, A., et al. (1996). Expression of ACC oxidase antisense gene inhibits ripening of cantaloupe melon fruits. *Nat. Biotechnol.* 14 (7), 862–866. doi: 10.1038/nbt0796-862
- Azadi, H., Samiee, A., Mahmoudi, H., Jouzi, Z., Kachak, P., Maeyer, P., et al. (2016). Genetically modified crops and small-scale farmers: Main opportunities and challenges. *Crit. Rev. Biotechnol.* 36 (3), 434–446. doi: 10.3109/07388551.2014.990413
- Bage, S., Barten, T., Brown, A., Crowley, J., Deng, M., Fouquet, R., et al. (2020). Genetic characterization of novel and CRISPR-Cas9 gene edited maize brachytic 2 alleles. *Plant Gene*. 21, 100198. doi: 10.1016/j.plgene.2019.100198
- Batista, R., Fonseca, C., Planchon, S., Negrão, S., Renaut, J., and Oliveira, M. (2017). Environmental stress is the major cause of transcriptomic and proteomic changes in GM and non-GM plants. *Sci. Rep.* 7, 10624. doi: 10.1038/s41598-017-09646-8
- Bawa, A., and Anilakumar, K. (2013). Genetically modified foods: safety, risks and public concerns—a review. *J. Food Sci. Technol.* 50 (6), 1035–1046. doi: 10.1007/s13197-012-0899-1
- Bhardwaj, A., and Nain, V. (2021). TALENs—an indispensable tool in the era of CRISPR: A mini review. *J. Genet. Eng. Biotechnol.* 19 (1), 125. doi: 10.1186/s43141-021-00225-z
- Borges, B., Arantes, O., Fernandes, A., Broach, J., and Fernandes, P. (2018). Genetically modified labeling policies: Moving forward or backward? *Front. Bioeng Biotechnol.* 6, 181. doi: 10.3389/fbioe.2018.00181
- Bouët, A., and Gruère, G. (2011). Refining opportunity cost estimates of not adopting GM cotton: An application in seven sub-Saharan African countries. *Appl. Econ Perspect. Policy*. 33, 260–279. doi: 10.1093/aep/prr010
- Boyle, J., Dalgleish, H., and Puzey, J. (2019). Monarch butterfly and milkweed declines substantially predate the use of genetically modified crops. *PNAS U.S.A.* 116 (8), 3006–3011. doi: 10.1073/pnas.1811437116
- Braatz, J., Harloff, H., Mascher, M., Stein, N., Himmelbach, A., and Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* 174, 935–942. doi: 10.1104/pp.17.00426
- Brookes, G., and Barfoot, P. (2015). Environmental impacts of genetically modified (GM) crop use 1996–2013: Impacts on pesticide use and carbon emissions. *GM Crops Food*. 6, 103–133. doi: 10.1080/21645698.2015.1025193
- Brookes, G., and Barfoot, P. (2016). Global income and production impacts of using GM crop technology 1996–2014. *GM Crops Food*. 7 (1), 38–77. doi: 10.1080/21645698.2016.1176817
- Brookes, G., and Barfoot, P. (2018). Farm income and production impacts of using GM crop technology 1996–2016. *GM Crops Food*. 9, 59–89. doi: 10.1080/21645698.2018.1464866
- Brookes, G., and Barfoot, P. (2020). GM crop technology use 1996–2018: Farm income and production impacts. *GM Crops Food*. 11 (4), 242–261. doi: 10.1080/21645698.2020.1779574
- Buiatti, M., Christou, P., and Pastore, G. (2013). The application of GMOs in agriculture and in food production for a better nutrition: Two different scientific points of view. *Genes Nutr.* 8 (3), 255–270. doi: 10.1007/s12263-012-0316-4
- Burney, J., Davis, S., and Lobell, D. (2010). Greenhouse gas mitigation by agricultural intensification. *PNAS U.S.A.* 107, 12052–12057. doi: 10.1073/pnas.0914216107
- Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., et al. (2015). CRISPR/Cas9-mediated genome editing in soybean hairy roots. *PloS One* 10, e0136064. doi: 10.1371/journal.pone.0136064
- Cai, R., Lewis, J., Yan, S., Liu, H., Clarke, C., Campanile, F., et al. (2011). The plant pathogen *Pseudomonas syringae* pv. tomato is genetically monomorphic and under strong selection to evade tomato immunity. *PloS Pathog.* 7, e1002130. doi: 10.1371/journal.ppat.1002130

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.

- Cai, Y., Wang, L., Chen, L., Wu, T., Liu, L., Sun, S., et al. (2020). Mutagenesis of GmFT2a and GmFT5a mediated by CRISPR/Cas9 contributes to expanding the regional adaptability of soybean. *Plant Biotechnol. J.* 18, 298–309. doi: 10.1111/pbi.13199
- Camerlengo, F., Frittelli, A., Sparks, C., Doherty, A., Martignago, D., Larré, C., et al. (2020). CRISPR-Cas9 multiplex editing of the α -amylase/trypsin inhibitor genes to reduce allergen proteins in durum wheat. *Front. Sustain Food Syst.* 4, 104. doi: 10.3389/fsufs.2020.00104
- Carrière, Y., Degain, B., and Tabashnik, B. (2021). Effects of gene flow between bt and non-bt plants in a seed mixture of Cry1A.105 + Cry2Ab corn on performance of corn earworm in Arizona. *Pest Manag. Sci.* 77 (4), 2106–2113. doi: 10.1002/ps.6239
- Case, P. (2021). *Rothamsted gets green light to trial gene-edited wheat*. Available at: <https://www.fwi.co.uk/arable/wheat/rothamsted-gets-green-light-to-trial-gene-edited-wheat> (Accessed 3 June 2022).
- Cermák, T., Curtin, S., Gil-Humanes, J., Čegan, R., Kono, T., Konečná, E., et al. (2017). A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* 29, 1196–1217. doi: 10.1105/tpc.16.00922
- Chandrasekaran, J., Brumin, M., Wolf, D., Leibman, D., Klap, C., Pearlsman, M., et al. (2016). Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* 17, 1140–1153. doi: 10.1111/mpp.12375
- Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049
- Chen, S., Zhang, N., Zhou, G., Hussain, S., Ahmed, S., Tian, H., et al. (2021). Knockout of the entire family of ATR genes in arabidopsis leads to enhanced drought and salinity tolerance without fitness costs. *BMC Plant Biol.* 21, 137. doi: 10.1186/s12870-021-02907-9
- Cobb, M. (2017). 60 years ago, Francis crick changed the logic of biology. *PLoS Biol.* 15 (9), e2003243. doi: 10.1371/journal.pbio.2003243
- Cohen, S., Chang, A., Boyer, H., and Helling, R. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 70 (11), 3240–3244. doi: 10.1073/pnas.70.11.3240
- da Costa, G., da Conceição Nepomuceno, G., da Silva Pereira, A., and Simões, B. (2022). Worldwide dietary patterns and their association with socioeconomic data: An ecological exploratory study. *Global Health* 18 (1), 31. doi: 10.1186/s12992-022-00820-w
- Daunert, S., Deo, S., Morin, X., and Roda, A. (2008). The genetically modified foods debate: Demystifying the controversy through analytical chemistry. *Anal. Bioanal. Chem.* 392, 327–331. doi: 10.1007/s00216-008-2312-5
- Davison, J., and Ammann, K. (2017). New GMO regulations for old: Determining a new future for EU crop biotechnology. *GM Crops Food* 8 (1), 13–34. doi: 10.1080/21645698.2017.1289305
- Delgado-Zegarza, J., Alvarez-Risco, A., Cárdenas, C., Donoso, M., Moscoso, S., Román, B., et al. (2022). Labeling of genetically modified (GM) foods in Peru: Current dogma and insights of the regulatory and legal statutes. *Int. J. Food Sci.* 3489785. doi: 10.1155/2022/3489785
- Doebley, J. (2006). Unfallen grains: How ancient farmers turned weeds into crops. *Science* 312 (5778), 1318–1319. doi: 10.1126/science.1128836
- EFSA, Mullins, E., Bresson, J., Dalmay, T., Dewhurst, I., Epstein, M., et al. (2022). Scientific opinion on development needs for the allergenicity and protein safety assessment of food and feed products derived from biotechnology. *EFSA J.* 20 (1), e07044. doi: 10.2903/j.efsa.2022.7044
- European Commission (2010). *A decade of EU-funded GMO research 2001–2010* (Brussels: European Commission). Available at: <https://op.europa.eu/en/publication-detail/-/publication/d1be9ff9-f3fa-4f3c-86a5-beb0882e0e65>.
- Evanega, S., Conrow, J., Adams, J., and Lynas, M. (2022). The state of the 'GMO' debate - toward an increasingly favorable and less polarized media conversation on ag-biotech? *GM Crops Food* 13 (1), 38–49. doi: 10.1080/21645698.2022.2051243
- Fernandes, G., Silva, A., Maronhas, M., Dos Santos, A., and Lima, P. (2022). Transgene flow: Challenges to the on-farm conservation of maize landraces in the Brazilian semi-arid region. *Plants (Basel)* 11 (5), 603. doi: 10.3390/plants11050603
- Fister, A., Landherr, L., Maximova, S., and Guiltinan, M. (2018). Transient expression of CRISPR/Cas9 machinery targeting TcNPR3 enhances defense response in theobroma cacao. *Front. Plant Sci.* 9, 268. doi: 10.3389/fpls.2018.00268
- Fitzpatrick, S., and Reid, B. (2019). Does gene flow aggravate or alleviate maladaptation to environmental stress in small populations? *Evol. Appl.* 12 (7), 1402–1416. doi: 10.1111/eva.12768
- Forner, J., Kleinschmidt, D., Meyer, E., Fischer, A., Morbitzer, R., Lahaye, T., et al. (2022). Targeted introduction of heritable point mutations into the plant mitochondrial genome. *Nat. Plants* 8 (3), 245–256. doi: 10.1038/s41477-022-01108-y
- Fraleigh, R. (1983). Liposome-mediated delivery of tobacco mosaic virus RNA into petunia protoplast: improved conditions for liposome-protoplast incubations. *Plant Mol. Biol.* 2 (1), 5–14. doi: 10.1007/BF00187570
- Gao, Y., and Zhao, Y. (2013). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* 56, 343–349. doi: 10.1111/jipb.12152
- García-Alonso, M., Novillo, C., Kostolaniova, P., Parrilla, M., Alcalde, E., and Podevin, N. (2022). The EU's GM crop conundrum: Did the EU policy strategy to convert EFSA GMO guidance into legislation deliver on its promises? *Did EU Policy strategy to convert EFSA GMO guidance into legislation Deliver its promises? EMBO Rep.* 23 (5), e54529. doi: 10.15252/embr.202154529
- Giraldo, P., Shinozuka, H., Spangenberg, G., Cogan, N., and Smith, K. (2019). Safety assessment of genetically modified feed: Is there any difference from food? *Front. Plant Sci.* 10, 1592. doi: 10.3389/fpls.2019.01592
- Graham, N., Patil, G., Bubeck, D., Dobert, R., Glenn, K., Gutscheit, A., et al. (2020). Plant genome editing and the relevance of off-target changes. *Plant Physiol.* 183 (4), 1453–1471. doi: 10.1104/pp.19.01194
- Gumtow, R., Wu, D., Uchida, J., and Tian, M. (2018). A *Phytophthora palmivora* extracellular cystatin-like protease inhibitor targets papain to contribute to virulence on papaya. *Mol. Plant-Microbe Interact.* 31 (3), 363–373. doi: 10.1094/MPMI-06-17-0131-FI
- Hartline-Grafton, H., and Hassink, S. (2021). Food insecurity and health: Practices and policies to address food insecurity among children. *Acad. Pediatr.* 21 (2), 205–210. doi: 10.1016/j.acap.2020.07.006
- He, Y., and Zhao, Y. (2020). Technological breakthroughs in generating transgene-free and genetically stable CRISPR-edited plants. *ABIOTECH.* 1, 88–96. doi: 10.1007/s42994-019-00013-x
- Holme, I., Wendt, T., and Holm, P. (2013). Intragenesis and cisgenesis as alternatives to transgenic crop development. *Plant Biotechnol. J.* 11 (4), 395–407. doi: 10.1111/pbi.12055
- Huang, J., Gao, L., Luo, S., Liu, K., Qing, D., Pan, Y., et al. (2022). The genetic editing of GS3 via CRISPR/Cas9 accelerates the breeding of three-line hybrid rice with superior yield and grain quality. *Mol. Breeding* 42, 22. doi: 10.1007/s11032-022-01290-z
- Huang, L., Li, Q., Zhang, C., Chu, R., Gu, Z., Tan, H., et al. (2020). Creating novel wx alleles with fine-tuned amylose levels and improved grain quality in rice by promoter editing using CRISPR/Cas9 system. *Plant Biotechnol. J.* 18, 2164–2166. doi: 10.1111/pbi.13391
- Huang, J., Li, J., Zhou, J., Wang, L., Yang, S., Hurstet, L., et al. (2018). Identifying a large number of high-yield genes in rice by pedigree analysis, whole-genome sequencing, and CRISPR-Cas9 gene knockout. *PNAS U.S.A.* 115, e7559–e7567. doi: 10.1073/pnas.1806110115
- Hu, Y., and Li, W. (2022). Development and application of CRISPR-cas based tools. *Front. Cell Dev. Biol.* 10, 834646. doi: 10.3389/fcell.2022.834646
- Ibáñez, S., Carneros, E., Testillano, P. S., and Pérez-Pérez, J. M. (2020). Advances in plant regeneration: Shake, rattle and roll. *Plants (Basel)* 19 (7), 897. doi: 10.3390/plants9070897
- Iqbal, Z., Sattar, M., and Shafiq, M. (2016). CRISPR/Cas9: A tool to circumscribe cotton leaf curl disease. *Front. Plant Sci.* 7, 475. doi: 10.3389/fpls.2016.00475
- ISAAA 2019 *Biotech crop highlights*. Available at: <https://www.isaaa.org/resources/publications/pocketk/16/#:~:text=The%20most%20planted%20biotech%20crops,crops%20or%2091.9%20million%20hectares> (Accessed 5 June 2022).
- ISAAA 2017 Global status of commercialized biotech/GM crops in 2017: Biotech crop adoption surges as economic benefits accumulate in 22 years. In: *ISAAA briefs 53 Ithaca*. Available at: <https://www.isaaa.org/resources/publications/briefs/53/executivesummary/default.asp> (Accessed 19 July 2022).
- ISAAA (2020) *GM approval database*. Available at: <http://www.isaaa.org/gmapprovaldatabase/default.asp> (Accessed 4 June 2022).
- Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., and Venkataraman, G. (2018). CRISPR for crop improvement: An update review. *Front. Plant Sci.* 9, 985. doi: 10.3389/fpls.2018.00985
- James, C. (2011) Global status of commercialized biotech/GM crops. In: *ISAAA briefs 43* (Ithaca: International Service for the Acquisition of Agri-biotech Applications). Available at: <https://www.isaaa.org/resources/publications/briefs/43/download/isaaa-brief-43-2011.pdf> (Accessed 5 June 2022).
- Jiang, W., Henry, I., Lynagh, P., Comai, L., Cahoon, E., and Weeks, D. (2017). Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnol. J.* 15, 648–657. doi: 10.1111/pbi.12663
- John, D., and Babu, G. (2021). Lessons from the aftermaths of green revolution on food system and health. *Front. Sustain Food Syst.* 5, 644559. doi: 10.3389/fsufs.2021.644559

- Kamle, M., Kumar, P., Patra, J., and Bajpai, V. (2017). Current perspectives on genetically modified crops and detection methods. *3 Biotech.* 7 (3), 219. doi: 10.1007/s13205-017-0809-3
- Karalis, D., Karalis, T., Karalis, S., and Kleisiari, A. (2020). Genetically modified products, perspectives and challenges. *Cureus.* 12 (3), e7306. doi: 10.7759/cureus.7306
- Karunaratne, S., Han, Y., Zhang, X., and Li, C. (2022). CRISPR/Cas9 gene editing and natural variation analysis demonstrate the potential for *HvARE1* in improvement of nitrogen use efficiency in barley. *J. Integr. Plant Biol.* 64, 756–770. doi: 10.1111/jipb.13214
- Kaur, N., Sharma, S., Hasanuzzaman, M., and Pati, P. (2022). Genome editing: A promising approach for achieving abiotic stress tolerance in plants. *Int. J. Genomics* 5547231. doi: 10.1155/2022/5547231
- Keiper, F., and Atanassova, A. (2022). Enabling genome editing for enhanced agricultural sustainability. *Front. Genome Ed.* 4, 898950. doi: 10.3389/fgene.2022.898950
- Kettenburg, A., Hanspach, J., and Abson, D. (2018). From disagreements to dialogue: Unpacking the golden rice debate. *Sustain Sci.* 13 (5), 1469–1482. doi: 10.1007/s11625-018-0577-y
- Khanday, I., Skinner, D., Yang, B., Mercier, R., and Sundaresan, V. (2018). A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nature.* 565, 91–95. doi: 10.1038/s41586-018-0785-8
- Kieu, N., Lenman, M., Wang, E., Petersen, B., and Andreasson, E. (2021). Mutations introduced in susceptibility genes through CRISPR/Cas9 genome editing confer increased late blight resistance in potatoes. *Sci. Rep.* 11, 4487. doi: 10.1038/s41598-021-83972-w
- Kikulwe, E., Wesseler, J., and Falck-Zepeda, J. (2011). Attitudes, perceptions, and trust: insights from a consumer survey regarding genetically modified banana in Uganda. *Appetite.* 57 (2), 401–413. doi: 10.1016/j.appet.2011.06.001
- Klümper, W., and Qaim, M. (2014). A meta-analysis of the impacts of genetically modified crops. *PLoS One* 9, e111629. doi: 10.1371/journal.pone.0111629
- Knisley, S. (2021). *Gene editing innovations present many benefits to farmers and their customers.* Available at: <https://www.uswheat.org/wheatletter/gene-editing-innovations-present-many-benefits-to-farmers-and-their-customers/#:~:text=Gene%20editing%20also%20holds%20excellent,land%2C%20and%20improving%20disease%20resistance> (Accessed 23 July 2022).
- Koonin, E., Makarova, K., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-cas systems. *Curr. Opin. Microbiol.* 37, 67–78. doi: 10.1016/j.mib.2017.05.008
- Kumar, K., Gambhir, G., Dass, A., Tripathi, A., Singh, A., Jha, A., et al. (2020). Genetically modified crops: current status and future prospects. *Planta.* 251 (4), 91. doi: 10.1007/s00425-020-03372-8
- Kurup, V., and Thomas, J. (2020). Edible vaccines: Promises and challenges. *Mol. Biotechnol.* 62 (2), 79–90. doi: 10.1007/s12033-019-00222-1
- Lacroix, B., and Citovsky, V. (2020). Biolistic approach for transient gene expression studies in plants. *Methods Mol. Biol.* 2124, 125–139. doi: 10.1007/978-1-0716-0356-7_6
- Lencucha, R., Pal, N., Appau, A., Thow, A., and Drope, J. (2020). Government policy and agricultural production: A scoping review to inform research and policy on healthy agricultural commodities. *Global Health* 16 (1), 11. doi: 10.1186/s12992-020-0542-2
- Li, R., Fu, D., Zhu, B., Luo, Y., and Zhu, H. (2018a). CRISPR/Cas9-mediated mutagenesis of *IncRNA1459* alters tomato fruit ripening. *Plant J.* 94, 513–524. doi: 10.1111/tj.13872
- Li, S., Gao, F., Xie, K., Zeng, X., Cao, Y., Zeng, J., et al. (2016). The *OsmiR396c-OsGRF4-OsGIF1* regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* 14, 2134–2146. doi: 10.1111/pbi.12569
- Li, J., Jiao, G., Sun, Y., Chen, J., Zhong, Y., Yan, L., et al. (2021a). Modification of starch composition, structure and properties through editing of *TaSBEIIa* in both winter and spring wheat varieties by CRISPR/Cas9. *Plant Biotechnol. J.* 19, 937–951. doi: 10.1111/pbi.13519
- Li, R., Li, R., Li, X., Fu, D., Zhu, B., Tian, H., et al. (2018b). Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ -aminobutyric acid levels in *Solanum lycopersicum*. *Plant Biotechnol. J.* 16, 415–427. doi: 10.1111/pbi.12781
- Li, J., Li, Y., and Ma, L. (2021b). Recent advances in CRISPR/Cas9 and applications for wheat functional genomics and breeding. *ABIOTECH.* 2, 375–385. doi: 10.1007/s42994-021-00042-5
- Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., et al. (2016). Reassessment of the four yield-related genes *gn1a*, *depl1*, *gs3*, and *ip1* in rice using a CRISPR/Cas9 system. *Front. Plant Sci.* 7, 377. doi: 10.3389/fpls.2016.00377
- Liu, L., and Cao, C. (2014). Who owns the intellectual property rights to Chinese genetically modified rice? evidence from patent portfolio analysis. *Biotechnol. Law Rep.* 33 (5), 181–192. doi: 10.1089/blr.2014.9971
- Liu, C., Cao, Y., Hua, Y., Du, G., Liu, Q., Wei, X., et al. (2021). Concurrent disruption of genetic interference and increase of genetic recombination frequency in hybrid rice using CRISPR/Cas9. *Front. Plant Sci.* 12, 757152. doi: 10.3389/fpls.2021.757152
- Liu, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., et al. (2017). *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat. Plants.* 3, 17043. doi: 10.1038/nplants.2017.43
- Liu, F., Lockett, B., Sorichetti, R., Watmougha, S., and Eimersa, M. (2022). Agricultural intensification leads to higher nitrate levels in lake Ontario tributaries. *Sci. Total Environ.* 830, 154534. doi: 10.1016/j.scitotenv.2022.154534
- Li, X., Wang, Y., Chen, S., Tian, H., Fu, D., Zhu, B., et al. (2018). Lycopene is enriched in tomato fruit by CRISPR/Cas9-mediated multiplex genome editing. *Front. Plant Sci.* 9, 559. doi: 10.3389/fpls.2018.00559
- Li, Q., Zhang, D., Chen, M., Liang, W., Wei, J., Qi, Y., et al. (2016). Development of japonica photo-sensitive genic male sterile rice lines by editing carbon starved anther using CRISPR/Cas9. *J. Genet. Genom.* 43, 415–419. doi: 10.1016/j.jgg.2016.04.011
- Li, J., Zhang, H., Si, X., Tian, Y., Chen, K., Liu, J., et al. (2017a). Generation of thermosensitive male-sterile maize by targeted knockout of the *ZmTMS5* gene. *J. Genet. Genom.* 44, 465–468. doi: 10.1016/j.jgg.2017.02.002
- Li, X., Zhou, W., Ren, Y., Tian, X., Lv, T., Wang, Z., et al. (2017b). High-efficiency breeding of early-maturing rice cultivars via CRISPR/Cas9-mediated genome editing. *J. Genet. Genom.* 44, 175–178. doi: 10.1016/j.jgg.2017.02.001
- Lucht, J. (2015). Public acceptance of plant biotechnology and GM crops. *Viruses.* 7 (8), 4254–4281. doi: 10.3390/v7082819
- Lu, H., Luo, T., Fu, H., Wang, L., Tan, Y., Huang, J., et al. (2018a). Resistance of rice to insect pests mediated by suppression of serotonin biosynthesis. *Nat. Plants* 4, 338–344. doi: 10.1038/s41477-018-0152-7
- Lusser, M., Parisi, C., Plan, D., and Rodriguez-Cerezo, E. (2012). Deployment of new biotechnologies in plant breeding. *Nat. Biotechnol.* 30, 231–239. doi: 10.1038/nbt.2142
- Lu, K., Wu, B., Wang, J., Zhu, W., Nie, H., Qian, J., et al. (2018b). 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
- Macovei, A., Sevilla, N., Cantos, C., Jonson, G., Slamet-Loedin, I., Čermák, T., et al. (2018). Novel alleles of rice *elf4G* generated by CRISPR/Cas9-targeted mutagenesis confer resistance to rice tungro spherical virus. *Plant Biotechnol. J.* 16, 1918–1927. doi: 10.1111/pbi.12927
- Maghari, B., and Ardekani, A. (2011). Genetically modified foods and social concerns. *Avicenna J. Med. Biotechnol.* 3 (3), 109–117.
- Maqbool, M., Aslam, M., Beshir, A., and Khan, M. (2018). Breeding for provitamin A biofortification of maize (*Zea mays* L.). *Plant Breed* 137 (4), 451–469. doi: 10.1111/pbr.12618
- Martin, A., Alabaddullah, A., and Moore, G. (2021). A separation-of-function *ZIP4* wheat mutant allows crossover between related chromosomes and is meiotically stable. *Sci. Rep.* 11, 1–13. doi: 10.1038/s41598-021-01379-z
- Mathers, J. (2007). Overview of genes, diet and cancer. *Genes Nutr.* 2 (1), 67–70. doi: 10.1007/s12263-007-0015-8
- Mathur, C., Kathuria, P., Dahiya, P., and Singh, A. (2015). Lack of detectable allergenicity in genetically modified maize containing “Cry” proteins as compared to native maize based on in silico & in vitro analysis. *PLoS One* 10 (2), e, 0117340. doi: 10.1371/journal.pone.0117340
- Matouskova, K., and Vandenberg, L. (2022). Towards a paradigm shift in environmental health decision-making: A case study of oxybenzone. *Environ. Health* 21 (1), 6. doi: 10.1186/s12940-021-00806-y
- Metje-Sprink, J., Menz, J., Modrzejewski, D., and Sprink, T. (2019). DNA-Free genome editing: Past, present and future. *Front. Plant Sci.* 9, 1957. doi: 10.3389/fpls.2018.01957
- Miao, C., Xiao, L., Hua, K., Zou, C., Zhao, Y., Bressan, R., et al. (2018). Mutations in a subfamily of abscisic acid receptor genes promote rice growth and productivity. *PNAS U.S.A.* 115, 6058–6063. doi: 10.1073/pnas.1804774115
- Midtvedt, T. (2014). Antibiotic resistance and genetically modified plants. *Microb. Ecol. Health Dis.* 25, 25918. doi: 10.3402/mehd.v25.25918
- Mieulet, D., Jolivet, S., Rivard, M., Cromer, L., Vernet, A., Mayonove, P., et al. (2016). Turning rice meiosis into mitosis. *Cell Res.* 26 (11), 1242–1254. doi: 10.1038/cr.2016.117
- Mrówczyńska-Kamińska, A., Bajan, B., Pawłowski, K. P., Genstwa, N., and Zmyślona, J. (2021). Greenhouse gas emissions intensity of food production systems and its determinants. *PLoS One* 16 (4), e0250995. doi: 10.1371/journal.pone.0250995
- Muehlfeld, K., and Wang, M. (2022). Intellectual property rights in China—a literature review on the public’s perspective. *Front. Sociol.* 7, 793165. doi: 10.3389/fsoc.2022.793165

- Mustapa, M., Batcha, M., Amin, L., Arham, A., Mahadi, Z., Yusoff, N., et al. (2021). Farmers' attitudes towards GM crops and their predictors. *J. Sci. Food Agric.* 101 (13), 5457–5468. doi: 10.1002/jsfa.11194
- Nakayasu, M., Akiyama, R., Lee, H., Osakabe, K., Osakabe, Y., Watanabe, B., et al. (2018). Generation of α -solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiol. Biochem.* 131, 70–77. doi: 10.1016/j.plaphy.2018.04.026
- Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J., and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 691–693. doi: 10.1038/nbt.2655
- Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., and Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* 7, 482. doi: 10.1038/s41598-017-00578-x
- Nieves-Cordones, M., Mohamed, S., Tanoi, K., Kobayashi, N., Takagi, K., Vernet, A., et al. (2017). Production of low-cs+ rice plants by inactivation of the k+ transporter OsHAK1 with the CRISPR-cas system. *Plant J.* 92, 43–56. doi: 10.1111/tjpi.13632
- Niiler, E. (1999). Terminator technology temporarily terminated. *Nat. Biotechnol.* 17 (11), 1054. doi: 10.1038/15034
- Nishizawa, T., Tamaoki, M., Aono, M., Kubo, A., Saji, H., and Nakajima, N. (2010). Rapeseed species and environmental concerns related to loss of seeds of genetically modified oilseed rape in Japan. *GM Crops*. 1 (3), 143–156. doi: 10.4161/gmcr.1.3.12761
- Nonaka, S., Arai, C., Takayama, M., Matsukura, C., and Ezura, H. (2017). Efficient increase of γ -aminobutyric acid (GABA) content in tomato fruits by targeted mutagenesis. *Sci. Rep.* 7, 7057. doi: 10.1038/s41598-017-06400-y
- Nordlee, J., Taylor, S., Townsend, J., Thomas, L., and Bush, R. (1996). Identification of a Brazil-nut allergen in transgenic soybeans. *N. Engl. J. Med.* 334 (11), 688–692. doi: 10.1056/NEJM199603143341103
- Ogata, T., Ishizaki, T., Fujita, M., and Fujita, Y. (2020). CRISPR/Cas9-targeted mutagenesis of OsERA1 confers enhanced responses to abscisic acid and drought stress and increased primary root growth under nonstressed conditions in rice. *PLoS One* 15 (12), e0243376. doi: 10.1371/journal.pone.0243376
- Okuzaki, A., Ogawa, T., Koizuka, C., Kaneko, K., Inaba, M., Imamura, J., et al. (2018). CRISPR/Cas9-mediated genome editing of the fatty acid desaturase 2 gene in brassica napus. *Plant Physiol. Biochem.* 131, 63–69. doi: 10.1016/j.plaphy.2018.04.025
- Oliver, M. (2014). Why we need GMO crops in agriculture. *Mo Med.* 111 (6), 492–507.
- Omobowale, E., Singer, P., and Daar, A. (2009). The three main monotheistic religions and gm food technology: An overview of perspectives. *BMC Int. Health Hum. Rights*. 9, 18. doi: 10.1186/1472-698X-9-18
- Park, J., Kim, E., Jang, Y., Jan, R., Farooq, M., Ubaidillah, M., et al. (2022). Applications of CRISPR/Cas9 as new strategies for short breeding to drought gene in rice. *Front. Plant Sci.* 13, 850441. doi: 10.3389/fpls.2022.850441
- Paschon, D., Lussier, S., Wangzor, T., Xia, D., Li, P., Hinkley, S., et al. (2019). Diversifying the structure of zinc finger nucleases for high-precision genome editing. *Nat. Commun.* 10 (1), 1133. doi: 10.1038/s41467-019-08867-x
- Paula, D., Brail, Q., Dahlbeck, D., and Staskawicz, B. (2016). CRISPR-Cas9 mediated mutagenesis of a DMR6 ortholog in tomato confers broad-spectrum disease resistance. *bioRxiv* 064824. doi: 10.1101/064824
- Pellegrino, E., Bedini, S., Nuti, M., and Ercoli, L. (2018). Impact of genetically engineered maize on agronomic, environmental and toxicological traits: A meta-analysis of 21 years of field data. *Sci. Rep.* 8, e3113. doi: 10.1038/s41598-018-21284-2
- Peng, R., Jones, D., Liu, F., and Zhang, B. (2021). From sequencing to genome editing for cotton improvement. *Trends Biotechnol.* 39 (3), 221–224. doi: 10.1016/j.tibtech.2020.09.001
- Podevin, N., Davies, V., Hartung, F., Nogué, F., and Casacuberta, J. (2013). Site-directed nucleases: a paradigm shift in predictable, knowledge-based plant breeding. *Trends Biotechnol.* 31, 375–383. doi: 10.1016/j.tibtech.2013.03.004
- Raffan, S., Sparks, C., Huttly, A., Hyde, L., Martignago, D., Mead, A., et al. (2021). Wheat with greatly reduced accumulation of free asparagine in the grain, produced by CRISPR/Cas9 editing of asparagine synthetase gene TaASN2. *Plant Biotechnol. J.* 19, 1602–1613. doi: 10.1111/pbi.13573
- Raman, R. (2017). The impact of genetically modified (GM) crops in modern agriculture: A review. *GM Crops Food*. 8 (4), 195–208. doi: 10.1080/21645698.2017.1413522
- Raza, S., Hassanin, A., Pant, S., Bing, S., Sitohy, M., Abdelnour, S., et al. (2022). Potentials, prospects and applications of genome editing technologies in livestock production. *Saudi J. Biol. Sci.* 29 (4), 1928–1935. doi: 10.1016/j.sjbs.2021.11.037
- Razzaq, M., Akhter, M., Ahmad, R., Cheema, K., Hina, A., Karikari, B., et al. (2022). CRISPR-Cas9 based stress tolerance: New hope for abiotic stress tolerance in chickpea (*Cicer arietinum*). *Mol. Biol. Rep.* 49 (4), 8977–8985. doi: 10.1007/s11033-022-07391-4
- Redden, R. (2021). Genetic modification for agriculture-proposed revision of GMO regulation in Australia. *Plants (Basel)*. 10 (4), 747. doi: 10.3390/plants10040747
- Roberts, A., Boeckman, C., Mühl, M., Romeis, J., Teem, J., Valicente, F., et al. (2020). Sublethal endpoints in non-target organism testing for insect-active GE crops. *Front. Bioeng Biotechnol.* 8, 556. doi: 10.3389/fbioe.2020.00556
- Rodrigues, R., Lage, C., and Vasconcellos, A. (2011). Intellectual property rights related to the genetically modified glyphosate tolerant soybeans in Brazil. *Acad. Bras. Cienc.* 83 (2), 719–730. doi: 10.1590/S0001-37652011000200029
- Salsman, J., and Delliare, G. (2016). Precision genome editing in the CRISPR era. *Biochem. Cell Biol.* 95, 187–201. doi: 10.1139/bcb-2016-0137
- Sanchez-Leon, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M., Sousa, C., Voytas, D., et al. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol. J.* 16, 902–910. doi: 10.1111/pbi.12837
- Sashidhar, N., Harloff, H., Potgieter, L., and Jung, C. (2020). Gene editing of three BnITPK genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnol. J.* 18, 2241–2250. doi: 10.1111/pbi.13380
- Schouten, H., Krens, F., and Jacobsen, E. (2006). Cisgenic plants are similar to traditionally bred plants: International regulations for genetically modified organisms should be altered to exempt cisgenesis. *EMBO Rep.* 7, 750–753. doi: 10.1038/sj.embor.7400769
- Sendhil, R., Nyika, J., Yadav, S., Mackolil, J., Prashat, R., Workie, E., et al. (2022). Genetically modified foods: Bibliometric analysis on consumer perception and preference. *GM Crops Food*. 13 (1), 65–85. doi: 10.1080/21645698.2022.2038525
- Shan, Q., Zhang, Y., Chen, K., and Gao, C. (2015). Creation of fragrant rice by targeted knockout of the OsBADH2 gene using TALEN technology. *Plant Biotechnol. J.* 13, 791–800. doi: 10.1111/pbi.12312
- Sharma, P., Singh, S., Iqbal, H., Parra-Saldivar, R., Varjani, S., and Tong, Y. (2022). Genetic modifications associated with sustainability aspects for sustainable developments. *Bioengineered*. 13 (4), 9508–9520. doi: 10.1080/21655979.2022.2061146
- Shen, R., Wang, L., Liu, X., Wu, J., Jin, W., Zhao, X., et al. (2017). Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. *Nat. Commun.* 8, 1310. doi: 10.1038/s41467-017-01400-y
- Shrawat, A., and Armstrong, C. (2018). Development and application of genetic engineering for wheat improvement. *CRC Crit. Rev. Plant Sci.* s37, 35–421. doi: 10.1080/07352689.2018.1514718
- Singh, M., Kumar, M., Albertsen, M., Young, J., and Cigan, A. (2018). Concurrent modifications in the three homeologs of Ms45 gene with CRISPR-Cas9 lead to rapid generation of male sterile bread wheat (*Triticum aestivum* L.). *Plant Mol. Biol.* 97, 371–383. doi: 10.1007/s11103-018-0749-2
- Smyth, S. (2020a). Regulatory barriers to improving global food security. *Glob Food Sec* 26, 100440. doi: 10.1016/j.gfs.2020.100440
- Smyth, S. (2020b). The human health benefits from GM crops. *Plant Biotechnol. J.* 18 (4), 887–888. doi: 10.1111/pbi.13261
- Smyth, S. (2022). Contributions of genome editing technologies towards improved nutrition, environmental sustainability and poverty reduction. *Front. Genome Ed.* 4, 863193. doi: 10.3389/fgeed.2022.863193
- Snow, A., and Palma, P. (1997). Commercialization of transgenic plants: Potential ecological risks. *Bioscience*. 47, 86–96. doi: 10.2307/1313019
- Southy, F. (2022) *Nutritionally charged lettuce developed with CRISPR/Cas gene editing tech*. Available at: <https://www.foodnavigator.com/Article/2022/01/04/Nutritionally-charged-lettuce-developed-with-CRISPR-Cas-gene-editing-tech#:~:text=Nutritionally%20charged%20lettuce%20developed%20with%20CRISPR%2FCas%20gene%20editing%20tech,-AddThis%20Sharing%20Buttons&text=A%20scientist%20in%20Israel%20is,of%20genetic%20modification%20is%20polarising> (Accessed 5 June).
- Subramanian, A., and Qaim, M. (2010). The impact of bt cotton on poor households in rural India. *J. Dev. Stud.* 46, 2. doi: 10.1080/00220380903002954
- Sun, Y., Jiao, G., Liu, Z., Zhang, X., Li, J., Guo, X., et al. (2017). Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* 8, 1298. doi: 10.3389/fpls.2017.00298
- Swamy, B., Samia, M., Boncodin, R., Marundan, S., Rebong, D., Ordonio, R., et al. (2019). Compositional analysis of genetically engineered GR2E “Golden rice” in comparison to that of conventional rice. *J. Agric. Food Chem.* 67 (28), 7986–7994. doi: 10.1021/acs.jafc.9b01524
- Szymczyk, B., Szczurek, W., Świątkiewicz, S., Kwiatek, K., Sieradzki, Z., Mazur, M., et al. (2018). Results of a 16-week safety assurance study with rats fed genetically modified bt maize: Effect on growth and health parameters. *J. Vet. Res.* 62 (4), 555–561. doi: 10.2478/jvetres-2018-0060
- Talakayala, A., Katta, S., and Garladinne, M. (2020). Genetic engineering of crops for insect resistance: An overview. *J. Biosci.* 45, 114. doi: 10.1007/s12038-020-00081-y

- Tang, L., Mao, B., Li, Y., Lv, Q., Zhang, L., Chen, C., et al. (2017). Knockout of OsNramp5 using the CRISPR/Cas9 system produces low Cd-accumulating indica rice without compromising yield. *Sci. Rep.* 7, 14438. doi: 10.1038/s41598-017-14832-9
- Teffer, T. (2021). Should we still worry about the safety of GMO foods? why and why not? a review. *Food Sci. Nutr.* 9 (9), 5324–5331. doi: 10.1002/fsn3.2499
- Touyz, L. (2013). Genetically modified foods, cancer, and diet: Myths and reality. *Curr. Oncol.* 20 (2), e59–e61. doi: 10.3747/co.20.1283
- Tran, M., Doan, D., Kim, J., Song, Y., Sung, Y., Das, S., et al. (2021). CRISPR/Cas9-based precise excision of SlHyPRP1 domain(s) to obtain salt stress-tolerant tomato. *Plant Cell Rep.* 40 (6), 999–1011. doi: 10.1007/s00299-020-02622-z
- Tripathi, L., Dhugga, K., Ntui, V., Runo, S., Syombua, E., Muiruri, S., et al. (2022). Genome editing for sustainable agriculture in Africa. *Front. Genome Ed.* 4, 876697. doi: 10.3389/fgeed.2022.876697
- Tripathi, J., Ntui, V., Shah, T., and Tripathi, L. (2021). CRISPR/Cas9-mediated editing of DMR6 orthologue in banana (*Musa spp.*) confers enhanced resistance to bacterial disease. *Plant Biotechnol. J.* 19 (7), 1291–1293. doi: 10.1111/pbi.13614
- Tseng, M., Roel, A., Deambrosi, E., Zorrilla, G., Ricetto, S., and Pittelkow, C. (2020). Towards actionable research frameworks for sustainable intensification in high-yielding rice systems. *Sci. Rep.* 10 (1), 9975. doi: 10.1038/s41598-020-63251-w
- Tsuboi, Y., Sakuma, T., Yamamoto, T., Horiuchi, H., Takahashi, F., Igarashi, K., et al. (2022). Gene manipulation in the mucorales fungus *rhizopus oryzae* using TALENs with exonuclease overexpression. *FEMS Microbiol. Lett.* 369 (1), fnac010. doi: 10.1093/femsle/fnac010
- Tuncel, A., Corbin, K., Ahn-Jarvis, J., Harris, H., Hawkins, E., Smedley, M., et al. (2019). Cas9-mediated mutagenesis of potato starch-branching enzymes generates a range of tuber starch phenotypes. *Plant Biotechnol. J.* 17, 2259–2271. doi: 10.1111/pbi.13137
- Turnbull, C., Lillemo, M., and Hvostlef-Eide, T. (2021). Global regulation of genetically modified crops amid the gene edited crop boom - a review. *Front. Plant Sci.* 12, 630396. doi: 10.3389/fpls.2021.630396
- Vega Rodríguez, A., Rodríguez-Oramas, C., Sanjuán Velázquez, E., Torre, A., Armendáriz, C., and Iruzueta, C. (2022). Myths and realities about genetically modified food: A risk-benefit analysis. *Appl. Sci.* 12 (6), 2861. doi: 10.3390/app12062861
- Viana, C., Freire, D., Abrantes, P., Rocha, J., and Pereira, P. (2022). Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. *Sci. Total Environ.* 806 (Pt 3), 150718. doi: 10.1016/j.scitotenv.2021.150718
- VIB (2022) Applications submitted for new field trials with genome-edited maize. Available at: <https://vib.be/news/applications-submitted-new-field-trials-genome-edited-maize> (Accessed 5 June).
- Waltz, E. (2016a). CRISPR-edited crops free to enter market, skip regulation. *Nat. Biotechnol.* 34, 582. doi: 10.1038/nbt0616-582
- Waltz, E. (2016b). Gene-edited CRISPR mushroom escapes US regulation. *Nature* 532 (7599), 293. doi: 10.1038/nature.2016.19754
- Waltz, E. (2018). With a free pass, CRISPR-edited plants reach market in record time. *Nat. Biotechnol.* 36, 6–7. doi: 10.1038/nbt0118-6b
- Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969
- Wang, F., Chen, M., Yu, L., Xie, L., Yuan, L., Qi, H., et al. (2017a). OsARM1, an R2R3 MYB transcription factor, is involved in regulation of the response to arsenic stress in rice. *Front. Plant Sci.* 8, 1868. doi: 10.3389/fpls.2017.01868
- Wang, L., Chen, L., Li, R., Zhao, R., Yang, M., Sheng, J., et al. (2017b). Reduced drought tolerance by CRISPR/Cas9-mediated SLMAPK3 mutagenesis in tomato plants. *J. Agric. Food Chem.* 65, 8674–8682. doi: 10.1021/acs.jafc.7b02745
- Wang, Z., Hong, Y., Zhu, G., Li, Y., Niu, Q., et al. (2020). Loss of salt tolerance during tomato domestication conferred by variation in a Na(+)/K(+) transporter. *EMBO J.* 39, e103256. doi: 10.15252/embj.2019103256
- Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., et al. (2019a). Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. *Nat. Biotechnol.* 37, 283–286. doi: 10.1038/s41587-018-0003-0
- Wang, M., Mao, Y., Lu, Y., Tao, X., and Zhu, J. (2017). Multiplex gene editing in rice using the CRISPR-Cpf1 system. *Mol. Plant* 10, 1011–1013. doi: 10.1016/j.molp.2017.03.001
- Wang, W., Pan, Q., Tian, B., He, F., Chen, Y., Bai, G., et al. (2019b). Gene editing of the wheat homologs of TONNEAU1-recruiting motif encoding gene affects grain shape and weight in wheat. *Plant J.* 100, 251–264. doi: 10.1111/tpj.14440
- Wang, F., Wang, C., Liu, P., Lei, C., Hao, W., Gao, Y., et al. (2016). Enhanced rice blast resistance by CRISPR/Cas9-targeted mutagenesis of the ERF transcription factor gene OsERF922. *PLoS One* 11, e0154027. doi: 10.1371/journal.pone.0154027
- Wang, T., Xun, H., Wang, W., Ding, X., Tian, H., Hussain, S., et al. (2021). Mutation of GmAITR genes by CRISPR/Cas9 genome editing results in enhanced salinity stress tolerance in soybean. *Front. Plant Sci.* 12, 779598. doi: 10.3389/fpls.2021.779598
- Wen, A., Havens, K., Bloch, S., Shah, N., Higgins, D., Davis-Richardson, A., et al. (2021). Enabling biological nitrogen fixation for cereal crops in fertilized fields. *ACS Synth. Biol.* 10 (12), 3264–3277. doi: 10.1021/acssynbio.1c00049
- Wenjing, W., Chen, Q., Singh, P., Huang, Y., and Pei, D. (2020). CRISPR/Cas9 edited HSA6a and HSA6b of *Arabidopsis thaliana* offers ABA and osmotic stress insensitivity by modulation of ROS homeostasis. *Plant Signal Behav.* 15 (12), 1816321. doi: 10.1080/15592324.2020.1816321
- Wu, F., and Butz, W. (2004). “The gene revolution: Genetically modified crops,” in *The future of genetically modified crops: Lessons from the green revolution*, 1st ed (Santa Monica, CA, USA: RAND Corporation), 39–64.
- Wunderlich, S., and Gatto, K. (2015). Consumer perception of genetically modified organisms and sources of information. *Adv. Nutr.* 6 (6), 842–851. doi: 10.3945/an.115.008870
- Xiao, Z., and Kerr, W. (2022). Biotechnology in China - regulation, investment, and delayed commercialization. *GM Crops Food* 13 (1), 86–96. doi: 10.1080/21645698.2022.2068336
- Xiao, Y., Yu, Y., Li, G., Xie, L., Guo, X., Li, J., et al. (2020). Genome-wide association study of vitamin E in sweet corn kernels. *Crop J.* 8 (2), 341–350. doi: 10.1016/j.cj.2019.08.002
- Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *PNAS* 112, 3570–3575. doi: 10.1073/pnas.1420294112
- Xie, Y., Niu, B., Long, Y., Li, G., Tang, J., Zhang, Y., et al. (2017a). Suppression or knockout of SaF/SaM overcomes the sa-mediated hybrid male sterility in rice. *J. Integr. Plant Biol.* 59, 669–679. doi: 10.1111/jipb.12564
- Xie, Y., Xu, P., Huang, J., Ma, S., Xie, X., Tao, D., et al. (2017b). Interspecific hybrid sterility in rice is mediated by OgTPR1 at the S1 locus encoding a peptidase-like protein. *Mol. Plant* 10, 1137–1140. doi: 10.1016/j.molp.2017.05.005
- Xu, R., Yang, Y., Qin, R., Li, H., Qiu, C., Li, L., et al. (2016). Rapid improvement of grain weight via highly efficient CRISPR/Cas9-mediated multiplex genome editing in rice. *J. Genet. Genom.* 43, 529–532. doi: 10.1016/j.jgg.2016.07.003
- Yang, Y., Xu, C., Shen, Z., and Yan, C. (2022). Crop quality improvement through genome editing strategy. *Front. Genome Ed.* 3, 819687. doi: 10.3389/fgeed.2021.819687
- Yang, Q., Zhong, X., Li, Q., Lan, J., Tang, H., Qi, P., et al. (2020). Mutation of the d-hordein gene by RNA-guided Cas9 targeted editing reducing the grain size and changing grain compositions in barley. *Food Chem.* 311, 125892. doi: 10.1016/j.foodchem.2019.125892
- Yao, L., Zhang, Y., Liu, C., Liu, Y., Wang, Y., Liang, D., et al. (2018). OsMATL mutation induces haploid seed formation in indica rice. *Nat. Plants* 4, 530–533. doi: 10.1038/s41477-018-0193-y
- Yeh, D., Gómez, M., and Kaiser, H. (2019). Signaling impacts of GMO labeling on fruit and vegetable demand. *PLoS One* 14 (10), e0223910. doi: 10.1371/journal.pone.0223910
- Ye, M., Peng, Z., Tang, D., Yang, Z., Li, D., Xu, Y., et al. (2018). Generation of self-compatible diploid potato by knockout of s-RNase. *Nat. Plants* 4, 651–654. doi: 10.1038/s41477-018-0218-6
- Yin, Y., Qin, K., Song, X., Zhang, Q., Zhou, Y., Xia, X., et al. (2018). BZR1 transcription factor regulates heat stress tolerance through FERONIA receptor like kinase-mediated reactive oxygen species signaling in tomato. *Plant Cell Physiol.* 59, 2239–2254. doi: 10.1093/pcp/pcy146
- Yu, W., Wang, L., Zhao, R., Sheng, J., Zhang, S., Li, R., et al. (2019). Knockout of SLMAPK3 enhances tolerance to heat stress involving ROS homeostasis in tomato plants. *BMC Plant Biol.* 19, 1–13. doi: 10.1186/s12870-019-1939-z
- Yu, X., Zhao, Z., Zheng, X., Zhou, J., Kong, W., Wang, P., et al. (2018). A selfish genetic element confers non-mendelian inheritance in rice. *Science* 360, 1130–1132. doi: 10.1126/science.aar4279
- Zafar, K., Khan, M., Amin, I., Mukhtar, Z., Yasmin, S., Arif, M., et al. (2020). Precise CRISPR-Cas9 mediated genome editing in super basmati rice for resistance against bacterial blight by targeting the major susceptibility gene. *Front. Plant Sci.* 11, 575. doi: 10.3389/fpls.2020.00575
- Zaraska, M. (2022) *The tomatoes at the forefront of a food revolution*. Available at: <https://www.bbc.com/future/article/20211207-the-tomatoes-at-the-forefront-of-a-food-revolution> (Accessed 3 June).
- Zhang, P., Du, H., Wang, J., Pu, Y., Yang, C., Yan, R., et al. (2020). Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. *Plant Biotechnol. J.* 18, 1384–1395. doi: 10.1111/pbi.13302

- Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* 7, 12617. doi: 10.1038/ncomms12617
- Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., et al. (2018a). Analysis of the functions of TaGW2 homoeologs in wheat grain weight and protein content traits. *Plant J.* 94, 857–866. doi: 10.1111/tpj.13903
- Zhang, J., Zhang, H., Botella, J., and Zhu, J. (2018b). Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the waxy gene in elite rice varieties. *J. Integr. Plant Biol.* 60, 369–375. doi: 10.1111/jipb.12620
- Zheng, Z., Appiano, M., Pavan, S., Bracuto, V., Ricciardi, L., Visser, R., et al. (2016). Genome-wide study of the tomato SLMLO gene family and its functional characterization in response to the powdery mildew fungus *oidium neolycopersici*. *Front. Plant Sci.* 7, 380. doi: 10.3389/fpls.2016.00380
- Zheng, M., Zhang, L., Tang, M., Liu, J., Liu, H., Yang, H., et al. (2020). Knockout of two BnaMAX1 homologs by CRISPR/Cas9-targeted mutagenesis improves plant architecture and increases yield in rapeseed (*Brassica napus* L.). *Plant Biotechnol. J.* 18 (3), 644–654. doi: 10.1111/pbi.13228
- Zhou, H., He, M., Li, J., Chen, L., Huang, Z., Zheng, S., et al. (2016). Development of commercial thermo-sensitive genic male sterile rice accelerates hybrid rice breeding using the CRISPR/Cas9-mediated TMS5 editing system. *Sci. Rep.* 6, 37395. doi: 10.1038/srep37395
- Zhou, J., Peng, Z., Long, J., Sosso, D., Liu, B., Eom, J., et al. (2015). Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *Plant J.* 82, 632–643. doi: 10.1111/tpj.12838
- Zupan, J., and Zambryski, P. (1995). Transfer of T-DNA from *agrobacterium* to the plant cell. *Plant Physiol.* 107, 1041–1047. doi: 10.1104/pp.107.4.1041



OPEN ACCESS

EDITED BY

Md. Anowar Hossain,
University of Rajshahi, Bangladesh

REVIEWED BY

Jlabao Ye,
Yangtze University, China
Haidong Yan,
University of Georgia, United States

*CORRESPONDENCE

Rajiv Ranjan
rajivranjanbt@gmail.com

SPECIALTY SECTION

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

RECEIVED 29 August 2022

ACCEPTED 27 October 2022

PUBLISHED 15 December 2022

CITATION

Tyagi P, Singh D, Mathur S, Singh A
and Ranjan R (2022) Upcoming
progress of transcriptomics studies on
plants: An overview.
Front. Plant Sci. 13:1030890.
doi: 10.3389/fpls.2022.1030890

COPYRIGHT

© 2022 Tyagi, Singh, Mathur, Singh and
Ranjan. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](#)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Upcoming progress of transcriptomics studies on plants: An overview

Parul Tyagi, Deeksha Singh, Shivangi Mathur, Ayushi Singh
and Rajiv Ranjan*

Plant Molecular Biology Lab, Department of Botany, Faculty of Science, Dayalbagh Educational
Institute, Agra, India

Transcriptome sequencing or RNA-Sequencing is a high-resolution, sensitive and high-throughput next-generation sequencing (NGS) approach used to study non-model plants and other organisms. In other words, it is an assembly of RNA transcripts from individual or whole samples of functional and developmental stages. RNA-Seq is a significant technique for identifying gene predictions and mining functional analysis that improves gene ontology understanding mechanisms of biological processes, molecular functions, and cellular components, but there is limited information available on this topic. Transcriptomics research on different types of plants can assist researchers to understand functional genes in better ways and regulatory processes to improve breeding selection and cultivation practices. In recent years, several advancements in RNA-Seq technology have been made for the characterization of the transcriptomes of distinct cell types in biological tissues in an efficient manner. RNA-Seq technologies are briefly introduced and examined in terms of their scientific applications. In a nutshell, it introduces all transcriptome sequencing and analysis techniques, as well as their applications in plant biology research. This review will focus on numerous existing and forthcoming strategies for improving transcriptome sequencing technologies for functional gene mining in various plants using RNA-Seq technology, based on the principles, development, and applications.

KEYWORDS

NGS, transcriptome analysis, gene function, molecular markers, secondary metabolites

Introduction

Since the advent of the post-genomic era, numerous omics approaches have been developed, such as genomics, transcriptomics, metabolomics, and proteomics. As one of these technologies, transcriptomics is the second earliest and most commonly used (Lockhart and Winzler, 2000; Levy and Myers, 2016). Transcriptomics studies focus on

the transcriptome. Because of its high throughput, improved precision, and cost effectiveness during the past two decades, genomic sequence databases have increased massively (Lathe et al., 2008; Jain, 2012; Karsch-Mizrachi et al., 2018). However, the complex mapping of a genome to various phenotypes, tissues, developmental stages, and environmental factors continues to pose a major challenge in molecular biology. A more in-depth understanding of gene regulation transcripts and expression is not only difficult but is also at the core of the problem. Transcriptomics has been extensively studied in a variety of organisms and provides critical insights into gene structure, expression, and regulation (Lowe et al., 2017). In recent years, transcriptomics research has grown tremendously due to rapid progress in sequencing technologies (Wang et al., 2016; Abdel-Ghany et al., 2016).

In plant research, there are some significant differences between transcriptomics and genomics. First, genome assembly is more complex and costly in plant research compared to transcriptomics (RNA-Seq); in the absence of a reference genome, the transcriptome can be used to assess an organism or plant's overall transcriptional activity. Second, the transcriptome changes over time and space because it includes information on secondary metabolic pathways as well as variations in gene expression at different times and spatial regions. According to research, the accumulation of biologically active compounds in different plant tissues depends on the gene expression levels of that tissue as well as the time they are produced. This is because plants have various growth conditions and periods, even within the same species. As a result, the transcriptome outperforms the genome in terms of identifying genes related to therapeutic plant components (Wang et al., 2009). These differences are important for studies on plant functional genome mining, gene regulatory domains development, genetic diversity (dominant and recessive genes), and bioactive compounds (Tyagi and Ranjan, 2021; Tyagi et al., 2022).

Transcriptome study approaches have progressed from basic DNA microarrays platform to RNA-Seq technology in recent times due to the continual developments in sequencing technology (Mironova et al., 2015). It has several perks including high sensitivity, high throughput, and efficacy, which may be used to analyses a complete transcriptome without a genomic reference sequence. RNA-Seq technology is a popular sequencing approach in molecular biology, biotechnology, and bioinformatics (Sun and Wei, 2018). This method has been widely exploited in model plants, including *Arabidopsis thaliana* (Zhang et al., 2019), *Oryza sativa* (Li et al., 2012; Pradhan et al., 2019; Yang et al., 2021), *Zea mays* (Xu B et al., 2014; Liu et al., 2019; Xu et al., 2021), *Rehmannia glutinosa* (Ma et al., 2021), *Polygonum cuspidatum* (Wang et al., 2021), *Asarum sieboldii* (Chen et al., 2021), *Calotropis gigantea* (Hoopes et al., 2018). Transcriptome can be defined as the total number of RNA molecules transcribed from a particular tissue or cell at a given

functional or developmental stage, which includes mRNA (messenger RNA) and nc-RNA (non-coding RNA). They are referred to as “bridges” because they accurately manage the transport of genetic information from DNA to protein (Costa et al., 2010), whereas non-coding RNA influences gene expression, protein synthesis, and various cellular processes at several levels (Kumar et al., 2016). Understanding transcriptomics thus enhances research with reference to the functions of cells, tissues, and organisms. RNA-Seq is a fairly new technology that measures all the biological amounts of the transcriptome. This makes it easier to study the transcriptome (Wang et al., 2009).

Computing and analysing RNA-seq still remains a challenging task. Accurate alignment of sequencing reads, and accurate expression level inference are some of these challenges. In RNA-seq, the central computational challenge is accurately and efficiently assigning short sequencing reads to their transcripts and determining gene expression based on that information. There have been some benchmark studies conducted, but they have generally been conducted using simulated RNA-seq datasets or RNA-seq datasets focusing only on long RNAs, such as messenger RNAs and long noncoding RNAs. Therefore, they did not assess the suitability of these tools for measuring total RNA in datasets that included small RNAs like transfer RNAs and small nucleolar RNAs. This review paper focuses on various methodologies, including transcriptome sequencing. Additionally, it provides an overview of the advancements made in using the RNA-Seq approach in plant research, including the molecular markers identification, functional gene expression analysis, biosynthetic pathways of secondary metabolite, and mechanisms for plant development.

Techniques of next generation sequencing (NGS) and their potential applications

RNA-seq: A history and evolution

Sequencing of events is a crucial comprehension method for a variety of reasons. Researchers of various abilities can efficiently organize material and ideas using sequencing structures. It is crucial to problem-solving in all subject areas, including science and social studies. So, differences of sequencing are coming according to generation, in first-generation sequencing platform, enabled sequencing of clonal DNA populations, in second generation sequencing, massively increased throughput by parallelizing many reactions, and in third generation sequencing, allow direct sequencing of single DNA molecules. In the 1970s, Walter Fiers and colleagues marked the start of RNA-Seq studies by sequencing the MS2

bacteriophage (3,569 nucleotides) whole transcriptome (Crowgey and Mahajan, 2019). RNA is highly unstable and susceptible to breakdown by cell membrane-associated RNases due to its single-stranded structure. Therefore, extensive efforts are required to sequence the whole transcriptome accurately. The discovery of reverse transcriptase (Zhang, 2019) enables the transformation of messenger RNA and non-coding RNA into stabilized DNA. The reverse-synthesized DNA is referred to as complementary DNA (cDNA). Sanger, a British scientist, invented DNA sequencing technology in 1975 (Zhang, 2019; Sanger and Coulson, 1975). Iscoe (Hwang et al., 2018) used the technique of RNA-Sequencing PCR, which exponentially amplifies cDNA. As a result of these pioneering efforts, microarray (chip) technology was developed (Bolón-Canedo et al., 2019). Through Arrayed technology, hundreds of known partial DNA sequences can be mounted on nylon membranes or solid support slides using molecular hybridization technology. Numerous genes can be quantitatively detected through hybridization. The reverse synthesis of cDNA can rapidly be used to sequence the transcriptomes of different biological samples, making microarrays an excellent way of exploring gene function (Clark et al., 2002). As of now, the technology is primarily used to study known genes, rather than unknown genes. Additionally, the microchip has difficulty detecting numerous transcripts generated by alternative splicing.

Sanger sequencing is an expensive, laborious, and time-consuming technique. NGS sequencing, however, has met a number of critical needs for molecular biologists since 2006, when it was introduced. In comparison to first-generation sequencing technology, NGS is fast, high-throughput, and economical. It is possible to sequence millions or billions of nucleic acids simultaneously, enabling the analysis of transcriptomes and genomes of all species using NGS. As of today, NGS is used for constructing different organism genomes for the purpose of collecting the complete gene sequences of various species of plants, including SARS-CoV-2 (Li et al., 2020) and *Chosenia arbutifolia* (Khoar), a Korean plant (Chen et al., 2014; Mei et al., 2016; Feng et al., 2019). Next generation sequencing is predominantly employed to sequence and analyze messenger RNA (mRNA) and small RNA (small RNA) from the transcriptome. As indicated in Table 1 (Zhang et al., 2016), the Roche 454, Illumina Solexa, and ABI SOLiD are all prominent next-generation sequencing platforms.

In 2005, Roche introduced the high-throughput Roche 454 sequencing technology, based on pyro-sequencing technique (Margulies et al., 2005). A longer read length and short run time make 454 sequencing stand out from other NGS technologies. This approach is non-fluorescent and does not require nucleic acid probes. However, it has the potential to introduce errors such as deletions or insertions during the

TABLE 1 Comprehensive overview of various next-generation sequencing (NGS) technologies.

| Name of Sequencer | Sanger Sequencing | Roche 454 sequencing | Solexa-Illumina | ABI SOLiD | SMRT | Nanopore |
|------------------------------|--|---|---|--|--|--|
| Based on Principle | "Dideoxy chain termination" | "Sequencing by synthesis" | "Sequencing by synthesis" | "Sequencing by ligation" | "Sequencing by synthesis" | "Electrical signal sequencing" |
| Advantage | Higher sensitivity to detect low-frequency variants | Long read length, ferments are generated in large number | Enables a wide variety of applications, allowing researchers to ask virtually any question related to the genome, transcriptome, or epigenome of any organism | High accuracy, Coverage is more than 30x, Detecting targeted re-sequencing, and transcriptome sequencing | Short time-consuming and no need for PCR amplification | Ability to produce ultra-long reads |
| Disadvantage | Sequencing could only small DNA sequence, higher cost, and minimum high throughput | More than 6 error rates with polybases, high cost, and low throughput | Short reads length, low multiplexing capability, only highly trained person can operate | Short read length, prone to chain decoding error | High cost and minimum high throughput | High single-base error rate and long sequencing time |
| Length of base pair | 15-40bp | 300-800 bp | 2×150bp | ~75 bp | 8-15 to 40-70kb | 500bp-2.3Mb |
| Developed by and year | Frederick Sanger and colleagues in 1977 | Shankar Balasubramania and David Klenerman in 2005 | Shankar Balasubramanian and David Klenerman in 2006 | Robert C. Martin (also known as Uncle Bob) in 2006 | Christian Henry, and John F. Milligan in 2004 | Hagan Bayley Clive G. Brown et al., in 2005 |
| Run time | 4-5hrs | 24 hrs | 6-7 days | 6-7 days | ~6 hrs | 10 minutes |
| Output data | 1.9~84 Kb | ~7 Gb | 600 Gb | 120 Gb | 5~10Gb | More than 50 Gb |
| References | Sanger et al., 1977 | Zhang et al., 2016 | Wang et al., 2009 | Xu, 2018 | Li et al., 2018 | Ma et al., 2019 |

process of sequencing (Liang et al., 2017). The method utilized by Solexa (Illumina) primarily rely on the principle of “sequencing by synthesis”, in which DNA fragments are randomly linked on a flow cell. Following extension and amplification, the surface of the glass generates millions of clusters containing thousands of identical DNA fragments. Sequencing of labeled dNTPs (fluorescence) is performed on the stretched DNA strand. This technology permits the use of synthetic probes and reference sequences in genome-wide expression analysis (Liu et al., 2012).

There are a few drawbacks to the platform, including its shorter read length as well as its complexity in assembling reads from scratch. With solid sequencing (2007), magnetic beads are used to sequence data in a highly parallel manner. This approach allows massive DNA amplification without incurring high costs because of constant ligation and development of fluorescently labeled oligonucleotides (Li and Xu, 2019). While high precision is the significant benefit of this technology, the primary disadvantage is that it is prone to chain decoding errors if an error occurs, as mentioned in Table 1.

Third-generation sequencing has been enabled ‘by improvements in sequencing technologies’. In addition to amplification requirements for NGS platforms, template relocation and mismatches in nucleotides and GC percentages are common challenges related to these approaches. These challenges reduce the accuracy and completeness of our sequencing data. Because third-generation sequencing offers such long reads, it has been widely used in structural variation, methylation, transcriptome and genomic analysis, among others. SMRT (Single-molecule real-time) and Nanopore sequencing have recently become key technologies for third-generation sequencing (Ma et al., 2019). As a sequencing method, the SMRT employs the “synthesis-based principle.” SMRT technology has two unique properties compared to classical methods. This was followed by adding and linking the fluorescent group to the phosphoric acid (H_3PO_4) group to eliminate the background noise. As a second benefit, there is no need for amplification and the SMRT allows for more precise measurements through self-correction (Li et al., 2018). On the other hand, SMRT sequencing technology has several disadvantages, one of which being the integration of random errors. By using circular consensus sequencing (CCS) and improving sequence coverage, it is now possible to increase the accuracy of a single read to 99.8% in comparison to second generation sequencing (Wenger et al., 2019).

Nanopore is a next-generation SMRT sequencing platform that determines the base composition mostly through changes in electrical impulses. Nanopore sequencing is more cost-effective and offers longer reads than other platforms. However, the major drawback of this technology is its high error rate which ranges from 5–20% (Huang et al., 2021). Despite its advantages, it has

some disadvantages, including a high probability of random errors and single-base errors as mentioned in Table 1. Using the above approaches without a reference sequence can be cost-prohibitive and time-consuming. RNA-Seq is a crucial part of NGS technology and an important tool for transcriptome analysis. It overcomes the limitations of microarray analysis and provides a greater understanding of transcriptome research.

Sample preparation and assembly for RNA-seq

The total RNA from biological samples is isolated using kits or manually (TRIzol) protocols (Drygin et al., 2021). m-RNA was extracted from total RNA “Experiments should be employed in triplicate form.” The enriched mRNA was fragmented and converted into first-strand cDNA, which was then followed by second-strand generation, A-tailing, adaptor ligation, and a limited number of PCR amplifications of the adaptor-ligated libraries. The amplified libraries were analyzed on the Bio-analyzer according to the instructions of the manufacturer. Quantification and qualification of the library were performed using any sensitivity kit by the Nanodrop spectrophotometer. NGS for cDNA of all plants was performed by library on any sequencing platform according to samples. The raw reads were first filtered to exclude the reads containing adaptors or with ambiguous nucleotides (‘N’). Below than 20% Q < 20 nucleotide bases were also trimmed. The obtained high-quality (HQ) clean reads were used to construct a sequence assembly using Trinity (Tyagi and Ranjan, 2021) or any other software. After removing duplicate Trinity-generated sequences with the TGICL, clusters and unigenes were recovered, as illustrated in Figure 1. Cogent creates a k-mer profile of non-redundant (NR) transcripts, calculates pairwise distances, and then categorizes transcripts into families based on k-mer similarity. Using any of the graph techniques, each transcript family was further reconstructed into one or more distinct transcript model (s).

Transcriptome assembly is required for a variety of downstream analyses. Raw transcriptome data is error-prone due to the vast amount of transcriptome data (Nagalakshmi et al., 2008). Thus, selecting assemblies based on distinct transcriptome data and research ideas is important. The primary thing to consider when assembling is whether there is a reference sequence available; assembly can be classified into two types, reference-based assembly and *de novo* assembly as shown in Figure 1. It is impossible to carry out the reference assembly for most plants because their genomic sequence data is frequently lacking; therefore, many non-model species are not provided with genomic sequence information. The only assembly approach appropriate for non-model plants is *de novo* assembly, which is depicted in Figure 1.

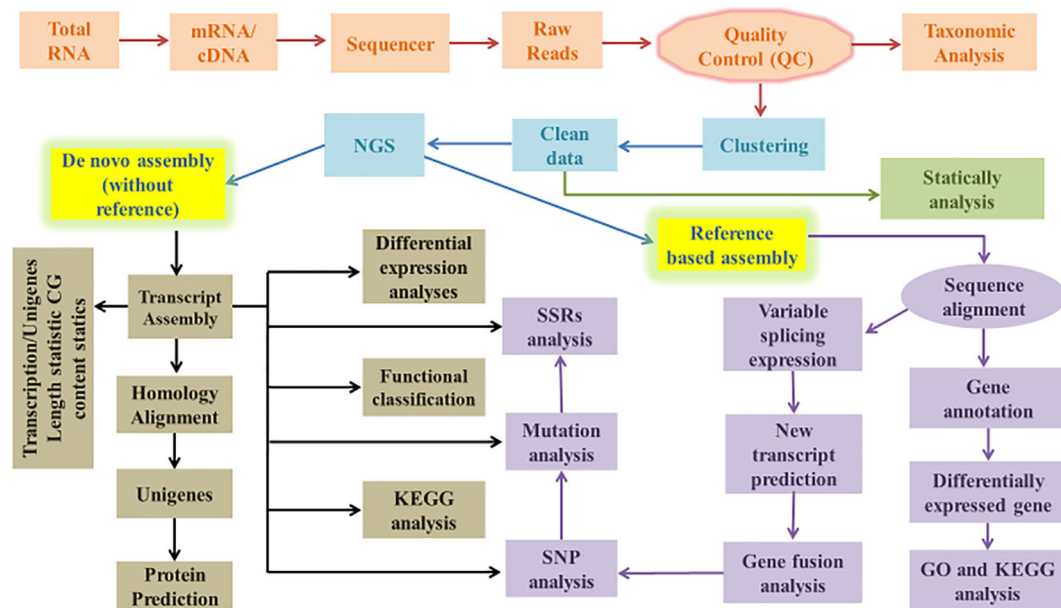


FIGURE 1
Flow chart of transcriptome (*De novo* and reference based) data analysis.

The assembly and function annotation process

Software such as Cufflinks and Scripture (Guo et al., 2021) is used to assemble genomes, whereas Oases, Trinity, Rnnotator (Martin et al., 2010), and SOAPdenovo-Trans (Madritsch et al., 2021), are used to assemble *de novo* genomes (Martin et al., 2010; Schulz et al., 2012; Grabherr et al., 2013; Trapnell et al., 2013; Xie et al., 2014). Assembling a single K-mer method with SOAPdenovo-Trans and Trinity is one stage of the *de novo* (without reference) assembly, while multiple K-mer methods are assembled with Oases and Rnnotator. Numerous k-mer methods can be utilized for obtaining huge transcripts of data, but when the findings of numerous k-mer methods are merged, a huge amount of duplicated data is generated, thus increasing the error rate and data complexity. As a result, at present time, a single K-mer assembly method is the most significant for improved precision that can be adopted, as mentioned in Table 2 (Strickler et al., 2012). SOAPdenovo-Trans connect to the contigs using an unknown nucleotide base (N). This approach is less accurate than the other three software programs in terms of assembly correctness, does not produce longer transcripts, and has a lower average accuracy rate (Lu, 2013). In recent years, Trinity has become an increasingly popular *de novo* assembly program due to its ability to produce accurate and efficient assembly outputs. In that respect, Scripture and Cufflinks use sensitive and conservative assembly methods, respectively. Despite the superior quality of

Cufflinks' assembly, Scripture's transcript quantity greatly exceeds that of Cufflinks, as shown in Table 2. Cufflinks can provide higher accuracy and quality assembly results by using high-quality reference sequences.

Gene function annotation is the process of identifying the functions related with individual gene based on existing data by comparing unidentified gene sequences to those published in public databases. GenOntology (GO) (Yu, 2020) and KEGG (Kyoto Encyclopedia of Genes and Genomes) (Yousef et al., 2021) are the two widely used approaches for functionally categorizing gene functions. A GO is divided into three categories, namely biological processes (BP), cellular compounds (CC) and molecular functions (MF). Unigene sequences are commonly annotated by databases such as Non-Redundant (NR), Protein Sequence Database (Zhao et al., 2021), GO, KEGG, Clusters of Orthologous Groups (COG) (Lee and Lee, 2021), Pfam (Mistry et al., 2021), NCBI Nucleotide Sequence Database (Schoch et al., 2020), and Swiss-Prot database (Liu FX et al., 2018).

Application of transcriptome sequencing

The RNA-Seq method provides insight into gene expression under a variety of environments and enables the discovery of new genes (Clamp et al., 2007; Mortazavi et al., 2008; Shendure, 2008; Pickrell et al., 2010; Touch et al., 2010; Filichkin et al.,

TABLE 2 List of softwares used in *De novo* assembly pipeline with their functions.

| Name of Software | Raw data | Data filtration | Output |
|---|------------------------------------|--|---|
| Through Sequencer (Leng et al., 2012) | RNA-Seq Raw Reads | Short Reads- Obtained from the common NGS platforms, including Illumina, SOLiD and 454, are often very short bases (35-500 bp). Long Reads- Oxford Nanopore/PacBio Sequencers can sequence up to long 5 to 100Kb reads. | Massively parallel millions to billions sequence that offers high- throughput, scalability, and takes lesser time. |
| FastQc (Andrews, 2010) | Quality Check (QC) | Quality assessment- Evaluate the raw read quality, identify the adaptor contaminations, and identify low quality samples | Good/Bed- According to Phred Score quality (Q-value) |
| Trimgalore | Read clean up (If contaminated) | Trimming- Removes the bad bases (adaptor sequences and low-quality bases) at start and end of the reads Filtering- Removes contaminants, low complexity reads (repeats), short reads less than 20 bases | K-mer- Shorter nucleotides than the read length De Bruijn graph- Several transcriptome assembly programs. Every path in the graph denotes a potential transcript for transcriptome assembly. |
| De novo assembly (Zhao et al., 2011) | Trinity (Henschel et al., 2012) | Quality/Phred Score quality (Q-score)- prediction of the probability (P) of an error in base calling. $Q(\text{phred}) = -10 \log_{10} P$ Or $P = 10^{-Q/10}$ | De novo Assembler- A novel method for the efficient and robust <i>de novo</i> reconstruction of transcriptomes. Software modules- Inchworm, Chrysalis, and Butterfly |
| RSEM (Li and Dewey, 2011) | Transcript Abundance Estimation | Assembly Statistics | N50 length is defined as the shortest sequence length at 50% of the transcriptome |
| Cd-hit (Suzek et al., 2007) | Transcript clustering | Generate unigenes | Group of transcript sequences |
| edgeR (Robinson et al., 2010) | Differential Expression Analysis | <i>edgeR</i> can be applied to differential expression at the gene, exon, transcript, or tag level. In fact, any genetic feature can be utilized to calculate read counts. There are two testing methods: likelihood ratio tests and quasi-likelihood F-tests. | The key abilities of package, and then gives several fully worked case studies, from counts to list of genes |
| TransDecoder (Wang et al., 2009) | Coding DNA Prediction | CDS prediction from unigenes | Segments of a gene's (mRNA) that code for protein. |
| Blast2GO (Martin et al., 2004) | Gene Ontology (GO) | Mapping and annotation | In detail, describe a gene/gene product, including three main characteristics: molecular function (MF), Biological process (BP), cellular compound (CC) |
| Trinotate (Wang et al., 2009) | Functional Annotation | (COG) Clusters of Orthologous Groups (for prediction of individual proteins function), Ven diagram (to identify common genes of all software), Pfam domain (to identification of protein family), Volcano plot (gene expression), Scattered plot (for normalization of obtained values), Heatmap (for highly significant differential expressed genes) | Portion identification and Gene prediction (process of collecting information about and describing a gene's) |
| KAAS (Moriya et al., 2007) | Pathway Prediction | Pathway analysis against KEGG databases | Identification of biological functions |
| DESeq2 (Love et al., 2017) | Differential Expression Analysis | Normalization, differential analysis and visualization of high- dimensional count data | Count matrices can be collapsed using collapse Replicates, which helps to combine counts from technical replications into single columns. |

2010), which in turn help us understand how cells function and how they function metabolically. The primary benefit of RNA-Seq is that it allows the comparison of gene expression patterns across samples. Before the advent of deep sequencing technology, the main means of measuring the expression levels of different genes was the microarray (Barrett et al., 2013; Anita and Caterina, 2018). However, hybridization techniques have low sensitivity, making it difficult to identify low-abundance targets and are incapable of finding tiny changes in the expression level of the target gene (Birney et al., 2007). Therefore, RNA-Seq is more accurate than microarray. RNA-Seq, in general, may measure the absolute number of each molecule in a cell

population and directly compare the results between tests. Second, RNA-Seq promotes new gene discovery. The annotations to transcripts in existing databases may not be comprehensive. The RNA-Seq results might self-assemble without the need for known genome annotations, facilitating the discovery of new genes (Zhao et al., 2014).

Third, RNA-Seq has shown satisfactory results in detecting sequence differences, including the identification of fusion candidate genes and the investigation of coding sequence polymorphisms. Because of alternative splicing, a single gene can produce many mRNA transcripts, each of which can be translated into a distinct protein with a different function.

Alternative splicing is universal, in eukaryotic species. The introduction or removal of various introns and exon regions during the splicing process forms different mRNA precursor obtained during gene transcription. Including the sequences spanning the splice junction region, sequences of all transcripts could be found with appropriate sequencing depth in RNA-Seq. The depth is defined as the ratio of the total number of bases (bp) retrieved by sequencing to the genome size, and it is one of the indicators used to evaluate the quantity of sequencing. Lastly, the main aspect of transcriptomics study is the finding and analysis of non-coding RNA (nc-RNA), as shown in Figure 1.

Single cell RNA sequencing technology (scRNA-seq)

In recent years, single-cell RNA sequencing (scRNA-seq) technologies have revolutionized the way we think about biological systems, as they are able to provide a high degree of spatial and temporal resolution of analyses (Zhang et al., 2019). Through scRNA-seq of plants, it is possible to identify new cell types and reveal how the different cell types interact spatially and developmentally revealing both common and rare cell types and cell states. By using single-cell RNA sequencing technology (scRNA-seq), stem cells can be systematically studied at a cellular and molecular level to gain insight into their differentiation trajectory (Islam et al., 2014; Macosko et al., 2015). An early developmental stage transcriptome was described in 2009 based on a next-generation sequencing platform as the first example of single-cell transcriptome analysis (Tang et al., 2009). High-resolution global views of single-cell heterogeneity have become increasingly popular since this study. A critical aspect of this study is analyzing the gene expression differences (Hwang et al., 2018). When analysis is conducted on individual cells, it may be possible to detect rare populations that may not be detectable in pooled analysis. Furthermore, recent advances in immunological research techniques and bioinformatics pipelines have made it possible for researchers to deconvolute highly diverse immune cell populations (Shalek et al., 2013). Additionally, scRNA-seq is increasingly used to analyze myoblast differentiation (Trapnell et al., 2014), lymphocyte fate (Stubbington et al., 2016), and early development (Petropoulos et al., 2016).

Advances in scRNA-seq technologies have opened new possibilities for uncovering the cellular and molecular differentiation trajectory of plant stem cells. The first application of single-cell technologies is to identify cell subtypes in heterogeneous populations of cells. To date, the Arabidopsis primary root tip is the most studied tissue using single-cell RNA sequencing. By analysing briTRIPLE mutants at single-cell resolution, Graeff et al., 2021 studied the impact of brassinosteroids (BR) signalling in Arabidopsis root tissue. In

this study, the researchers found that BR signalling does not affect cell proliferation or cellular development, rather it promotes cellular anisotropy and cell division plane orientation. It is possible to identify phenotypic variations among cell types through single-cell sequencing profiling. ScRNA-seq was used to examine the phenotypes of epidermal cells in the mutant, including root hair deficient cells (rhd6) and glabrous2 cells (gl2). Cell identity phenotypes were identified based on generated data. A study of rice radicals using single-cell sequencing and chromatin accessibility was conducted by Zhang et al., 2021. The development trajectories of epidermal cells and ground tissues were reconstructed using root tip cell profiling, which provided insight into the mechanism that controls cell fate determination in these lineages. Transcriptome profiles and marker genes for these cell types were uncovered through further analysis.

Furthermore, scRNA-seq has been applied to crop improvement in maize by highlighting transcriptional differentiation at high resolution in maize cells. An analysis of 12,525 maize ear cells using scRNA-seq technology was recently published by Xu et al. (2021). A scRNA-seq map of an inflorescence was generated as a result of this profiling. By identifying genetic redundancy in maize, establishing gene regulatory networks at the cell level, and identifying key loci with high ear-yielding characteristics, the generated data can help promote maize genetics. A similar study was conducted by Bezruczyk and colleagues in the same year, using scRNA-seq to study bundle sheath (BS) differentiation in maize. In maize, single-cell sequencing profiling helped identify cells with unique characteristics on the adaxial side, enhancing the possibility of bioengineering the plant (Bezruczyk et al., 2021). As single-cell transcriptomics is increasingly applied to several plant species, future research will indeed introduce it to crop species. This will pave the way for its incorporation into applied plant research, which could benefit our agricultural systems in the future.

Functional gene extraction from plants

To withstand external pressures and adapt to their environments, medicinal plants have evolved a plethora of regulatory mechanisms. A functional gene mining technique identifies key enzymes, pathways, and regulatory mechanisms in plants, thus helping us better understand their molecular biology. Xu WR et al., (2014) analyzed *Vitis amurensis* (Amur grape) transcriptome using the Illumina GA-II sequencing platform and found that in cold regulation a total of 6,850 transcripts involved. There were 3,676 as well as 3,147 copies of transcripts that were upregulated and downregulated, respectively, and 38 key TF families that were implicated in cold regulation. The results of this study provide a foundation for further research into the mechanisms involved in *Vitis* species' cold stress tolerance. Pragati et al., 2018 reported that

a total of 221,792 and 161,733 transcripts in which 141,310 and 113,062 unigenes were obtained from leaf and root tissues of *Aloe vera* (Gwarpatha), respectively, were used on the Illumina platform. It has been determined that 16 genes are involved in the production of lignin, carotenoids saponins and anthraquinone. The results of this study will be useful in future research into genes involved in secondary metabolite biosynthesis and metabolic regulation in *A. vera* and other *Aloe* species.

The transcriptome of *Paeonia suffruticosa* was sequenced and examined in 2012 by Mutasa-Gottgens and colleagues, and 81,725 copies of unigenes associated with drought resistance were found. It has been predicted that genes associated with hormone signaling pathways are important for drought adaptation and setting framework to study *P. suffruticosa*'s drought stress response mechanism in the future. Singh et al. (2017) sequenced the *Trillium govanianum* transcriptome using the Illumina sequencing platform, collecting 69,174 transcripts, and discovering many genes involved in steroidal saponin production and biosynthetic pathways of various secondary metabolites. Researchers identified tissues (leaf and fruit) as the primary sites for producing steroidal saponins in the biosynthesis of terpenoids, brassinosteroids, carotene, flavonoids, steroids and phenylpropane. Genetic manipulation is valuable for the identification of biologically active metabolites, as well as for the development of molecular markers that are functionally related to the identified metabolites. A transcriptome sequence of *Polygonum minus* was published by Loke et al. (2017), which revealed 188,735 transcripts. They also reported 163,200 (86.5%) *P. minus* transcript similarity matches, the vast majority of which were with *Arabidopsis* transcripts (58.9%). Root and leaf tissues have improved metabolite pathways. The findings will contribute to the development of this species' genetic resources.

Callerya speciosa genome sequencing was conducted by Li et al. (2016) using Illumina's platform through which 161,926 unigenes and 4,538 differentially expressed genes were obtained. Store roots may be implicated in starch synthesis, cell wall loosening and light signaling. Additionally, they may play a role in the development of store roots. Using these findings, subsequent research was conducted on growing *C. speciosa* roots, producing therapeutic substances, and breeding the plant. In a study by Hou et al. (2018), 56,392 unigenes and 4,585 significant DEGs were found in *Cornus officinalis* leaf and fruit tissues using next-generation sequencing (NGS). A total of 1,392 genes were up-regulated in fruit tissues, while 3,193 genes were down-regulated. Most DEGs are involved in the regulation of secondary metabolism and the production of terpenoids. This knowledge contributes to the understanding of plant metabolism and gene expression. The phenolic compound rosmarinic acid has antimicrobial and antioxidant properties and is physiologically active. It was reported that *Dracocephalum tanguticum* revealed 151,463 unigenes in its transcriptome by

Li et al. (2017). A total of 22 genes are predicted to be involved in the biosynthesis of rosmarinic acid, providing references for future research on rosmarinic acid biosynthesis genes.

Development of plant molecular markers

SSR markers are a frequently deployed form of microsatellite DNA marker. A tandem repetition sequence consists of 1-6 nucleotide base pairs, with the most common sequence being di-nucleotide repeats. These markers were polymorphic and different numbers of tandem repetitions were associated with them. Historically, SSRs have been used in creating genetic maps, defining genetic diversity, mapping genes, and identifying parental ties due to their high polymorphism, simplicity, codominance, and easy detection. To understand the genetic diversity within asparagus species, Kapoor et al. (2020) used SSR markers to study asparagus varieties grown in different regions of northwest India. There were more than 120 alleles amplified, ranging between 3 and 8, with an average number of five alleles per marker. The lengths of the alleles ranged from 90 to 680 bp. Based on genetic diversity analysis, most *Asparagus* varieties have a conservative genetic base, except for *A. adscendens*, which indicates that this species has an extensive genetic base.

Future hybridization and conservation of *Asparagus* species are likely to be affected by these findings. Bhandari et al. (2020) used Illumina paired-end sequencing technology to create new SSR markers for *Salvadora oleoides* (Bada Peelu). From 21,055 microsatellite repeats, they developed 14,552 SSR markers, and randomly selected and confirmed 7,101 SSRs; 94 primers exhibited polymorphisms, and 34 primers failed to amplify. This study provides a foundation for future research on *S. oleoides*.

A HiSeq 4000 sequencing platform was used to analyze transcripts from *Populus alba* (root, leaf, and stem) (Dinh et al., 2020). A total of 11,343 EST-SSRs were identified, of which 101 primer pairs (forwards and reverses) were selected for polymorphism validation. Polymorphisms in populations were discovered by amplifying DNA fragments with 20 primers. Conservation, restoration, and management strategies can benefit greatly from this conclusion. Wang et al. (2020) sequenced and analyzed the transcriptome of *Gastrodia elata* and identified 34,322 unigenes. In 2,007 unigenes (5.85%), at least one SSR was present.

There were 498 detections (21.67%) of a repeating pattern of AG/CT among these SSRs. As a result of this research, Lade et al. (2020) have gained deeper insight into the molecular mechanisms that regulate the growth, development and metabolism of *G. elata* (Tianma). Total 96 sample of *Tinospora cordifolia* were gathered from 10 different geographically diverse locales of the India. In *T. cordifolia*, a total of 268,149 transcripts were assembled. Amongst them,

7,611 SSRs were identified. Tc16, Tc17, Tc31, Tc38, Tc59, Tc60, Tc129, Tc106, Tc130, and Tc131 were shown to contain genetic diversity potential. The potential markers SSR-18, TcSSR-37, TcSSR-59, TcSSR-92, TcSSR-123, and TcSSR-126 have been identified. Genetic enhancement of *T. cordifolia* will be assisted by these components and the newly discovered SSR markers. A comprehensive genetic analysis of two *Menispermum* species was conducted by Hina et al. (2020) using Illumina's transcriptomic platform and *de novo* assembly. Sum of 521 polymorphic EST-SSRs were found out of a total of 53,712 and 78,921 unigenes. The newly designed EST-SSR marker was also shown to be highly transferrable throughout the *Menispermum* species studied. In order to genetically map *Menispermum* populations, these unique microsatellites will be used. A total of 86,195 unigenes were identified in *P. lactiflora* by using microsatellite software, while 21,998 SSR sites were identified dispersed over 17,567 unigenes. Among the 100 primer pairs, 45 were selected at random and amplified bands of polymorphism. For the cluster analysis of sixteen *P. lactiflora* variations, these 45 primer pairs were used. Molecular markers-assisted breeding with *P. lactiflora* will be facilitated by the novel SSR marker.

The metabolic pathways of secondary metabolites

Secondary metabolites are often the most significant components of plants. They play an essential part in the process by which plants adapt to their surroundings and build up a defense mechanism against the impact of various stress. There are many factors that affect the accumulation of secondary metabolites, including the growing environment as well as the developmental stage of the part from which it produced. In various stages of development, transcriptomics is used to investigate pathways of biosynthesis, which lead to secondary metabolites, and to mine genes involved in biogenesis. A scientific foundation has been laid for determining how plants accumulate and utilize active components. *Entada phaseoloides* has been used for medical purposes for centuries. Traditional medicine makes extensive use of the stems due to the wind-dampness-eliminating and anti-inflammatory properties that these stems possess. The triterpenoid saponins found in *E. phaseoloides* are the compounds with the highest level of physiological activity. *E. phaseoloides* root, stem, and leaf tissues to uncover 26 candidate genes for cytochrome P450 and 17 uridine diphosphate glycosyltransferases that are involved in the production of triterpene saponin (Liao et al., 2020). As can be seen in Supplementary Table 1, the findings were beneficial to both the production of triterpenoid saponin and research into functional genomics.

There are several physiologically active compounds in *Lantana camara* (Lantana), including steroidal saponins,

flavonoids, and glycosides. Using tools for sequencing the transcriptome, Shah et al. (2020) put together *L. camara* leaves and roots from scratch. It was found that 72,877 and 513,985 unigenes were present in leaves and roots, respectively. Of these, 229 and 943 genes were responsible for the production of phenyl-propanoic acid. As a broad-spectrum antibiotic, *Tetragium hemsleyanum* extract is used to treat fever and sore throats. An in-depth analysis of the metabolome and transcriptome of purple and green *T. hemsleyanum* leaves was performed by Yan et al. (2020). 4211 transcripts have been identified in the purple and green leaves, 209 metabolites have been found to be differentially expressed, and 16 chemicals have been associated with 14 transcripts implicated in the pathway for anthocyanin synthesis. The sesquiterpene lactones produced by *Saussurea lappa* have a high medicinal value. In a study conducted by Bains et al. (2019), the leaf transcriptome of *S. lappa* was sequenced to identify flavonoid and sesquiterpene-producing transcripts. Transcripts from genes implicated in alkaloid metabolism have been identified in a small number of cases. As a result of these insights, scientists will be able to learn more about how plants' functional genomes work. Transcriptome analysis of *Arisaema heterophyllum* Blume and its leaf, tuber, and root tissues identified 47783, 43363, and 35686 unigenes, respectively, implicating genes involved in isoflavone biosynthesis. Experimental confirmation of 87 candidate genes encoding isoflavone-producing enzymes was accomplished (Wang et al., 2018). The findings of this study pave the way for further research on the pharmacological action of *Arisaema*. The antioxidative and anti-inflammatory properties of flavonoids, along with their application in the treatment of diseases, are illustrated in Supplementary Table 1.

A significant percentage of flavonoids may be found in the leaves of *Ginkgo biloba*. Wu et al. (2018) identified 37,625 unigenes from transcriptome sequencing of *G. biloba* with various flavonoid concentrations. According to the research, several potential genes are involved in the manufacture, transport, and regulation of flavonoids, according to the research. It was found that MYB transcription factor and dihydroflavonol-4-reductase, two of the fourteen flavonoid transport genes, participate in flavonoid transport. It is anticipated that the discoveries will help expand the current *G. biloba* gene database, broaden Ginkgo species research, and provide crucial information for the development of Ginkgo-related pharmaceuticals. *De novo* transcriptome sequencing of *Abrus mollis* leaves enabled analysis of flavonoid synthesis routes and associated precursors (Yuan et al. (2018)). Liu MM et al. (2018) found 99,807 unigenes in *Artemisia argyi* leaf, root, and stem tissues, including many genes that encode terpene-synthesis enzymes and transcription factors. It is anticipated that the findings will be used to investigate the molecular pathways of *A. argyi*. An analysis of *Panax ginseng* root tissues using 454-sequencing technology was conducted by Jayakodi et al. (2014). There were 17 percent difference in transcript levels

between adventitious and common roots, as well as a 21 percent difference in ginsenoside-producing genes of *P. notoginseng* (Luo et al., 2011; Liu MH et al., 2015). The transcripts of *P. notoginseng* differed by 17% between adventitious and common roots, as did 21 ginsenoside-producing genes (Luo et al., 2011; Liu MH et al., 2015). Jayakodi et al., 2015 studied the transcriptomic profile of *P. ginseng* leaves, roots, and flowers and identified 107,340 unigenes, including 9,908 metabolic pathways and 270 triterpene saponin-producing genes. Among 32 genes expressed specifically in annual ginseng roots, seven genes were expressed specifically in 6-year-old ginseng roots, and 38 genes were implicated in triterpene saponin synthesis, as shown in Supplementary Table 1.

The transcriptomic process

Transcriptomic approaches have been widely used to identify genes involved in plant growth and development, to identify genes that are expressed differently under abiotic stress, and to study their resistance to it. Identifying key influencing components can be helpful in plant cultivation and breeding, as well as simplifying the selection of improved varieties (Rastogi et al., 2019). Using transcriptome sequencing, Liu MM et al. (2018) studied the response of *A. argyi* leaves to cold, drought, waterlogging, and salt stress. Cold stress was the most damaging condition to the plants. Treatments with abiotic stress also reduced eugenol synthesis. The discovery of several stress-tolerance genes in *A. argyi* has enabled transgenic or polymerized plants to become stress-tolerant. According to Li et al. (2020), transcriptome sequencing was used to investigate the molecular mechanisms behind *Salvia miltiorrhiza* tissues' responses to mild abiotic stress (drought). In total, 58,085 unigenes were discovered, of which 28,846 could be identified as such. Significant enrichment in metabolic processes and catalytic activities were found among differential transcripts in roots and leaves based on GO enrichment studies. The expression of genes that encode enzymes involved in the synthesis of phenylpropanoids and terpenoids increased in response to moderate drought stress. These findings have provided a solid foundation for further research into the process of manufacturing therapeutic components in *S. miltiorrhiza* as well as irrigation techniques for successful cultivation, as shown in Supplementary Table 1.

Using the third sequencing technique, Feng et al. (2019) sequenced the full-length transcriptome of *Angelica sinensis*, and differential expression sequencing was done on the wild-type transcriptome. An analysis of NGS data identified 25,463 transcripts with differential expression. There was a significant difference between transcripts in the pathway for plant-pathogen interaction and signal transduction of plant hormones. The purpose of this study is to lay out a platform for screening and developing *A. sinensis*. The transcriptome sequencing of callus

tissue of *S. laniceps* was used to identify genes associated with frost resistance (Xu, 2018). In the GO enrichment investigation, 155 substances related to low temperatures, oxidative stress, and plant hormone responses were identified. Based on the KEGG enrichment analysis, several pathways were significantly enriched during low-temperature responses, including ribosomes, fatty acid metabolism, and unsaturated fatty acid biosynthesis (Supplementary Table 1). The findings of this study provide a framework for future research on genes associated with frost resistance in *Saussurea laniceps*.

Discussion

In the biological sciences, transcriptome sequencing is an important sequencing technique that can be used without genomic reference sequences. Transcriptomic analysis can be used to analyze a wide range of plants and has many applications. The application of transcriptomics for obtaining genetic information about plants is growing rapidly due to its fast, high coverage, efficiency, and high throughput characteristics. This technology has been applied to mining new functional genes, analyzing secondary metabolite pathways, identifying plant developmental pathways, and obtaining helpful information for plant breeding (Li et al., 2018). Understanding secondary metabolite synthesis pathways and associated genes will benefit secondary metabolism regulatory network analysis and secondary metabolism studies in plants. Numerous plants have been analyzed by transcriptomic analysis, and the technique has been applied in many different fields. The use of transcriptomics to gather genetic information on plants is expanding as a fast, high-coverage, high-throughput, and high-efficiency analytical approach. Using the technology, researchers have been able to identify plant developmental pathways, mine novel functional genes, and analyze secondary metabolite synthesis pathways (Li et al., 2018).

Some plants produce multiple secondary metabolites to survive with biological and abiotic stresses, some of which can be used to treat a variety of human ailments. These plants are commonly referred to as plants because of their medicinal potential. Approximately 270,000 plants species have been identified globally, with less than 40,000 species having putative medical significance (Mamedov and Nazim, 2012; Tan et al., 2015). Except for a few model plants that are important research tools and sources of genomic data, most plants are largely underexplored in terms of biological information. There is a significant gap in plant genomic data due to a lack of research knowledge about the characterization of plant transcriptomes. This hinders research on important topics like the identification of significant differentially expressed genes (DEGs) and pathways related to secondary metabolite biosynthesis. In order to gather data for future plant research,

plant transcriptome research should be encouraged (Navin and Hicks, 2011). In the future, transcriptome data analysis and study will assist in identifying functional genes associated with secondary metabolism pathways.

NGS techniques help us understand the transcriptome's complexity. In several transcriptome researches on plants, NGS reads are limited by the requirement for assembly or reference genomes. The short read length of NGS technology makes it difficult to study full-length transcripts in plants. It is usually possible to investigate only the local structure of the gene and alternative splicing mechanisms for full-length transcripts. Third-generation (3G) sequencing technology is becoming more common with the advancement of sequencing techniques. Due to its relatively high sequencing cost and low throughput, third-generation sequencing is limited in its utility for transcription at this time. RNA-Seq research is mostly conducted using NGS technology, with the third generation serving as a backup. Genomic, transcriptomic, metabolomic, and proteomic technologies are also available for collaborative research. Biological functions are currently hot research topics for high-throughput sequencing, and the data will be useful for mining genes and algorithms, for fast analysis, and for showing that these topics are very useful for mining genes and algorithms.

A transcriptome profiling approach is required to identify isolated cells in a population of plants or animals, and the observations are often reflective of the number of cells predominating (Ranzoni et al., 2019). Due to cell heterogeneity, phenotypic characteristics may appear to be the same, but genetic information will vary dramatically. Consequently, transcriptome sequencing typically results in the loss of a great deal of low-abundance information (Navin and Hicks, 2011). It is predicted that single-cell transcriptome research will move into a new phase with the development of sequencing technology and the sharp decline in sequencing costs. It is possible to systematically track the dynamic changes of individual cells using single-cell transcriptome sequencing, which can effectively support the heterogeneity of single-cell gene expression that conventional sequencing ignores. We can thus gain a better understanding of cell state, genetic makeup, gene expression, and regulation, as well as develop herbal medicines.

Conclusion

As a precursor to NGS developments, first-generation sequencing technologies and pioneering computing and bioinformatics tools generated the initial sequencing data and information within a structural and functional genomics framework. With NGS, high-throughput sequencing options are substantially cheaper, friendlier, and more flexible, allowing us to generate much more data on genomics and transcriptomics that can be used to further explore proteomics and metabolomics. A variety of NGS platforms have been

released. During the first three decades of sequencing, Sanger sequencing dominated, but cost and time were major obstacles. The emergence of the second generation sequencing in 2005 and subsequent years has set the stage for breaking through the limitations of the first-generation sequencing. Sequencing by synthesis and sequencing by ligation are the two approaches proposed so far for second-generation sequencing.

There are currently more advantages to “third-generation sequencing” platforms (compared to first- and second-generation NGS platforms), including longer run times, complete transcript sequencing, and faster turnaround times. Their high mismatch rate, however, limits their use in transcriptome sequencing. On the other hand, third-generation (3G) sequencing technology can be used in conjunction with NGS technology to repair errors and offer genotyping recognition. As the cost of third-generation sequencing decreases and the accuracy of the technology improves, third-generation sequencing will become more frequently used in transcriptome research for accurate and complete transcriptome sequencing. The economic crop research model should lead to the widespread use of transcriptome sequencing in traditional plants. The majority of plant research does not rely solely on RNA-Seq at the moment. As RNA-Seq technology develops, multi-omics-related techniques will play a significant role, along with metabolomics and proteomics. Plant research would be modernized with the advancement of transcriptomic approaches, including developing metabolomics and proteomics techniques.

NGS has become a science that integrates biological information systems with big data, but many challenges remain for NGS data acquisition, analysis, storage, interpretation and integration. In order to continue producing comprehensive, high-throughput data for analysis and production, new technologies and large-scale collaboration efforts will be needed in the future. With the advent of affordable benchtop sequencers and third-generation sequencing tools, smaller laboratories and individual scientists can participate in the genomic revolution and contribute new knowledge to structural genomics and functional genomics in the life sciences.

Author contributions

PT, DS, SM, and AS wrote the manuscript and RR edited the same for further improvement.

Acknowledgments

We are grateful thanks to the Director of Dayalbagh Educational Institute, Dayalbagh, Agra for the support.

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.2022.1030890/full#supplementary-material>

References

- Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., and Schilkey, F. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7, 11706. doi: 10.1038/ncomms11706
- Ai, Y., Zhang, Q., Wang, W., Zhang, C., Cao, Z., and Bao, M. (2016). Transcriptomic analysis of differentially expressed genes during flower organ development in genetic Male sterile and Male fertile *Tagetes erecta* by digital gene-expression profiling. *PLoS One* 11 (3), e0150892. doi: 10.1371/journal.pone.0150892
- Andrews, S. (2010) *FastQC: A quality control tool for high throughput sequence data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Anita, A., and Caterina, M. (2018). Whole transcriptome profiling of late-onset alzheimer's disease patients provides insights into the molecular changes involved in the disease. *Sci. Rep.* 8, 4282. doi: 10.1038/s41598-018-22701-2
- Bains, S., Thakur, V., Kaur, J., Singh, K., and Kaur, R. (2019). Elucidating genes involved in sesquiterpenoid and flavonoid biosynthetic pathways in *Saussurea lappa* by *de novo* leaf transcriptome analysis. *Genomics* 111, 1474–1482. doi: 10.1016/j.ygeno.2018.09.022
- Barrett, T., Wilhite, S. E., and Ledoux, P. (2013). NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193
- Bezruczyk, M., Zöllner, N. R., Kruse, C. P., Hartwig, T., Lautwein, T., Köhrer, K., et al. (2021). Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. *Plant Cell* 33 (3), 531–547. doi: 10.1093/plcell/koaa055
- Bhandari, M. S., Meena, R. K., Shamooin, A., Saroj, S., and Pandey, S. (2020). First *de novo* genome specific development, characterization, and validation of simple sequence repeat (SSR) markers in genus *Salvadora*. *Mol. Biol. Rep.* 47, 6997–7008. doi: 10.1007/s11033-020-05758-z
- Birney, E., Stamatoyannopoulos, J. A., and Dutta, A. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874
- Bolón-Canedo, V., Alonso-Betanzos, A., López-de-Ullibarri, I., and Cao, R. (2019). "Challenges and future trends for microarray analysis," in *Microarray bioinformatics* (New York, NY: Humana), 283–293.
- Chen, Y., Liu, Y. S., and Zeng, J. G. (2014). Progresses on plant genome sequencing profile. *Life Sci. Res.* 18, 66–74. doi: 10.16605/j.cnki.1007-7847.2014.01.006
- Clamp, M., Fry, B., Kamal, M., Xie, X. H., Cuff, J., Lin, M. F., et al. (2007). Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19428–19433. doi: 10.1073/pnas.0709013104
- Clark, T. A., Sugnet, C. W., and Ares, M. J. (2002). Genome wide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296, 907–910. doi: 10.1126/science.1069415
- Costa, V., Angelini, C., Feis, I. D., and Ciccodicola, A. (2010). Uncovering the complexity of transcriptomes with RNA-seq. *J. BioMed. Biotechnol.* 2010, 853916. doi: 10.1155/2010/853916
- Crowgey, E. L., and Mahajan, N. (2019). "Advancements in next-generation sequencing for detecting minimal residual disease," in *Minimal residual disease testing* (Switzerland: Springer, Cham), 159–192.
- Dinh, D. V., Syed, N. S., Mai, P. P., Van, T. B., Minh, T. N., and Thi, P. N. (2020). *De novo* assembly and transcriptome characterization of an endemic species of Vietnam, *Panax vietnamensis* ha et grushv., including the development of EST-SSR markers for population genetics. *BMC Plant Biol.* 20, 159–138. doi: 10.1186/s12870-020-02571-5
- Drygin, Y. F., Butenko, K. O., and Gasanova, T. V. (2021). Environmentally friendly method of RNA isolation. *Analytical Biochemistry* 620, 114113. doi: 10.1016/j.ab.2021.114113
- Feng, C. H., Hei, C. Y., Wang, Y., Zeng, Y. F., and Zhang, J. G. (2019). Phylogenetic position of *Chosenia arbutifolia* in the salicaceae inferred from whole chloroplast genome. *For. Res.* 32, 73–77. doi: 10.13275/j.cnki.lykxyj.2019.02.011
- Filichkin, S. A., Priest, H. D., and Givan, S. A. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., and Amit, I. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Graeff, M., Rana, S., Wendrich, J. R., Dorier, J., Eekhout, T., Fandino, A. C. A., et al. (2021). A single-cell morpho-transcriptomic map of brassinosteroid action in the arabidopsis root. *Mol. Plant* 14 (12), 1985–1999. doi: 10.1016/j.molp.2021.07.021
- Guo, J., Huang, Z., Sun, J., Cui, X., and Liu, Y. (2021). Research progress and future development trends in medicinal plant transcriptomics. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.691838
- Henschel, R., Lieber, M., Wu, L., Nista, P. M., Haas, B. J., and LeDuc, R. D.. (2012). "Trinity RNA-seq assembler performance optimization," in *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond*. (New York: The association for computing machinery).
- Hina, F., Yisilam, G., Wang, S., Li, P., and Fu, C. X. (2020). *De novo* transcriptome assembly, gene annotation and SSR marker development in the moon seed genus *Menispermum* (Menispermaceae). *Front. Genet.* 11. doi: 10.3389/fgen.2020.00380
- Hoopes, G. M., Hamilton, J. P., Kim, J., Zhao, D., Wiegert-Rininger, K., Crisovan, E., et al. (2018). Genome assembly and annotation of the medicinal plant *calotropis gigantea*, a producer of anticancer and antimalarial cardenolides. *G3: Genes Genomes Genet.* 8 (2), 385–391. doi: 10.1534/g3.117.300331
- Hou, D. Y., Shi, L. C., Yang, M. M., Li, J., Zhou, S., and Zhang, H. X. (2018). *De novo* transcriptomic analysis of leaf and fruit tissue of *Cornus officinalis* using illumina platform. *PLoS One* 13 (13), e0192610. doi: 10.1371/journal.pone.0192610
- Huang, N., Nie, F., Ni, P., Luo, F., Gao, X., and Wang, J. (2021). NeuralPolish: a novel nanopore polishing method based on alignment matrix construction and orthogonal bi-GRU networks. *Bioinformatics* 37 (19), 3120–3127. doi: 10.1093/bioinformatics/btab354
- Hwang, B., Lee, J. H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50 (8), 1–14. doi: 10.1038/s12276-018-0071-8
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., et al. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11 (2), 163–166. doi: 10.1038/nmeth.2772
- Jain, M. (2012). Next-generation sequencing technologies for gene expression profiling in plants. *Brief. Funct. Genomics* 11, 63–70. doi: 10.1093/bfpg/blr038
- Jayakodi, M., Lee, S. C., Lee, Y. S., Park, H. S., Kim, N. H., and Jang, W. J. (2015). Comprehensive analysis of *Panax ginseng* root transcriptomes. *BMC Plant Biol.* 15, 138. doi: 10.1186/s12870-015-0527-0

- Jayakodi, M., Lee, S. C., Park, H. S., Jang, W. J., Lee, Y. S., and Choi, B. S. (2014). Transcriptome profiling and comparative analysis of *Panax ginseng* adventitious roots. *J. Ginseng. Res.* 38, 278–288. doi: 10.1016/j.jgr.2014.05.008
- Kapoor, M., Mawal, P., Sharma, V., and Gupta, R. C. (2020). Analysis of genetic diversity and population structure in *Asparagus* species using SSR markers. *J. Genet. Eng. Biotechnol.* 18, 50. doi: 10.1186/s43141-020-00065-3
- Karsch-Mizrachi, I., Takagi, T., and Cochrane, G. (2018). International nucleotide sequence database collaboration. the international nucleotide sequence database collaboration. *Nucleic Acids Res.* 46, D48–D51. doi: 10.1093/nar/gkx1097
- Kumar, S., Razaq, S. K., and Vo, A. D. (2016). Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip Rev RNA* 7, 811–823. doi: 10.1002/wrna.1382
- Lade, S., Pande, V., Rana, T. S., and Yadav, H. K. (2020). Estimation of genetic diversity and population structure in *Tinospora cordifolia* using SSR markers. *3 Biotech.* 10, 413–425. doi: 10.1007/s13205-020-02300-7
- Lathe, W., Williams, J., Mangan, M., and Karolchik, D. (2008). Genomic data resources: challenges and promises. *Nat. Educ.* 1, 2.
- Lee, D. G., and Lee, S. H. (2021). Investigation of COGs (Clusters of orthologous groups of proteins) in 1,309 species of prokaryotes. *J. Life Sci.* 31 (9), 834–839. doi: 10.5352/JLS.2021.31.9.834
- Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. L., Smits, B. M., et al. (2012). An empirical bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29, 1035–1043. doi: 10.1093/bioinformatics/btt087
- Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115. doi: 10.1146/annurev-genom-083115-022413
- Liang, S. B., Liu, J. Y., Yang, J. T., Liu, J., Li, J. L., and Zhang, Y. M. (2017). Next-generation sequencing applications for crop genomes. *China Biotechnol.* 37, 111–120. doi: 10.13523/j.cb.20170216
- Liao, W. F., Mei, Z. N., Miao, L. H., Liu, P. L., and Gao, R. J. (2020). Comparative transcriptome analysis of root, stem, and leaf tissues of *Entada phaseoloides* reveals potential genes involved in triterpenoid saponin biosynthesis. *BMC Genomics* 21, 639. doi: 10.1186/s12864-020-07056-1
- Li, X. L., Bo, B., Wu, J., Deng, Q. Y., and Zhou, B. (2012). Transcriptome analysis of early interaction between rice and *Magnaporthe oryzae* using next-generation sequencing technology. *Hereditas* 34, 104–114. doi: 10.3724/SP.J.1005.2012.00102
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinf.* 12, 323. doi: 10.1186/1471-2105-12-323
- Li, H., Fu, Y., Sun, H., Zhang, Y., and Lan, X. (2017). Transcriptomic analyses reveal biosynthetic genes related to rosmarinic acid in *Dracocephalum tanguticum*. *Sci. Rep.* 7, 74. doi: 10.1038/s41598-017-00078-y
- Li, Y. M., Li, S. X., Li, X. S., and Li, C. Y. (2018). Transcriptome studies with the third-generation sequencing technology. *Life Sci. Instrum.* 16, 114–121.
- Liu, X., Bi, B., Xu, X., Li, B., Tian, S., Wang, J., et al. (2019). Rapid identification of a candidate nicosulfuron sensitivity gene (Nss) in maize (*Zea mays* L.) via combining bulked segregant analysis and RNA-seq. *Theor. Appl. Genet.* 132 (5), 1351–1361. doi: 10.1007/s00122-019-03282-8
- Liu, L., Li, Y., Li, S., Hu, N., He, Y. M., and Pong, R. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2008, 251364. doi: 10.1155/2012/251364
- Liu, M. H., Yang, B. R., Cheung, W. F., Yang, K. Y., Zhou, H. F., and Kwok, J. L. (2015). Transcriptome analysis of leaves, roots, and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. *BMC Genomics* 16, 265. doi: 10.1186/s12864-015-1477-5
- Liu, F. X., Yang, W. G., and Sun, Q. H. (2018). Transcriptome sequencing data analysis and high throughput GO annotation. *J. Anhui Agric. Univ.* 46, 88–91. doi: 10.13989/j.cnki.0517-6611.2018.31.027+100
- Liu, M. M., Zhu, J. H., Wu, S. B., Wang, C. K., Guo, X. Y., and Wu, J. W. (2018). *De novo* assembly and analysis of the *Artemisia argyi* transcriptome and identification of genes involved in terpenoid biosynthesis. *Sci. Rep.* 8, 1236–1243. doi: 10.1038/s41598-018-24201-9
- Li, X., Wang, J., Ming, L., Li, L., and Li, Z. (2016). Transcriptome analysis of storage roots and fibrous roots of the traditional medicinal herb *Callerya speciosa* (Champ.) ScHot. *PLoS One* 11, e0160338. doi: 10.1371/journal.pone.0160338
- Li, Y., and Xu, X. Y. (2019). Research progress of high-throughput sequencing technology. *China Med. Eng.* 27, 32–37. doi: 10.19338/j.issn.1672-2019.2019.03.009
- Li, X. Y., Zhou, J. W., Yan, Z. Y., and Chen, X. (2020). Sequencing and analysis of transcriptome to reveal regulation of gene expression in *Salvia miltiorrhiza* under moderate drought stress. *Zhong Cao Yao* 51, 1600–1608. doi: 10.7501/j.issn.0253-2670.2020.06.029
- Lockhart, D. J., and Winzler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature* 405 (6788), 827–836. doi: 10.1038/35015701
- Loke, K. K., Rahnamaie-Tajadod, R., Yeoh, C. C., Goh, H. H., and Noor, N. M. (2017). Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. *PeerJ* 5, e2938. doi: 10.7717/peerj.2938
- 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
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. *PLoS Comput. Biol.* 13, e1005457. doi: 10.1371/journal.pcbi.1005457
- Lu, X. (2013). “A comparison of transcriptome assembly software for next-generation sequencing technologies,” in *Ph.D. Thesis* (Gansu: University of Lanzhou).
- Luo, H., Sun, C., Sun, Y. Z., Wu, Q., Li, Y., and Song, J. Y. (2011). Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics* 12, S5. doi: 10.1186/1471-2164-12-S5-S5
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161 (5), 1202–1214. doi: 10.1016/j.cell.2015.05.002
- Ma, L., Dong, C., Song, C., Wang, X., Zheng, X., Niu, Y., et al. (2021). *De novo* genome assembly of the potent medicinal plant *Rehmannia glutinosa* using nanopore technology. *Computational and Structural Biotechnology Journal* 19, 3954–3963. doi: 10.1016/j.csbj.2021.07.006
- Madritsch, S., Burg, A., and Sehr, E. M. (2021). Comparing *de novo* transcriptome assembly tools in di- and autotetraploid non-model plant species. *BMC Bioinf.* 22 (1), 1–17. doi: 10.1186/s12859-021-04078-8
- Mamedov, N., and Nazim, N. (2012). Medicinal plants studies: history, challenges and prospective. *Aromat. Plants* 1, 1–2. doi: 10.4172/2167-0412.1000e133
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., and Bembel, L. A. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959
- Martin, J., Bruno, V. M., Fang, Z., Meng, X., Blow, M., Zhang, T., et al. (2010). Rnnotator: an automated *de novo* transcriptome assembly pipeline from stranded RNA-seq reads. *BMC Genomics* 11 (1), 1–8. doi: 10.1186/1471-2164-11-663
- Martin, D., Berriman, M., and Barton, G. J. (2004). GOTcha: a new method for prediction of protein function assessed by the annotation of seven genomes. *BMC Bioinf.* 5, 178. doi: 10.1186/1471-2105-5-178
- Ma, L. N., Yang, J. B., Ding, Y. F., and Li, Y. K. (2019). Research progress on three generations sequencing technology and its application. *China Anim. Husb. Vet. Med.* 46, 2246–2256. doi: 10.16431/j.cnki.1671-7236.2019.08.007
- Mei, C. G., Wang, H. C., Zan, L. S., Cheng, G., Li, A. L., and Zhao, C. P. (2016). Research progress on animal genome research based on high-throughput sequencing technology. *J. Northwest A&F Univ.* 44, 43–51. doi: 10.13207/j.cnki.jnwf.2016.03.007
- Mironova, V. V., Weinholdt, C., and Grosse, I. (2015). “RNA-seq data analysis for studying abiotic stress in horticultural plants,”. *Abiotic Stress Biol* 1, 197–220. doi: 10.1007/978-4-431-55251-2_14
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49 (D1), D412–D419. doi: 10.1093/nar/gkaa913
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., and Kanehisa, M. (2007). KAA: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Mortazavi, A., Williams, B. A., and McCue, K. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., and Gerstein, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349. doi: 10.1126/science.1158441
- Navin, N., and Hicks, J. (2011). Future medical applications of single-cell sequencing in cancer. *Genome Med.* 3, 1–12. doi: 10.1186/gm247
- Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., et al. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* 165 (4), 1012–1026. doi: 10.1016/j.cell.2016.03.023
- Pickrell, J. K., Marioni, J. C., and Pai, A. A. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464, 768–772. doi: 10.1038/nature08872
- Pradhan, S. K., Pandit, E., Nayak, D. K., Behera, L., and Mohapatra, T. (2019). Genes, pathways and transcription factors involved in seedling stage chilling stress tolerance in indica rice through RNA-seq analysis. *BMC Plant Biol.* 19 (1), 1–17. doi: 10.1186/s12870-019-1922-8

- Pragati, C., Muniya, R., Sangwan, R. S., Ravinder, K., Anil, K., and Vinod, C. (2018). *De novo* sequencing, assembly, and characterization of *Aloe vera* transcriptome and analysis of expression profiles of genes related to saponin and anthraquinone metabolism. *BMC Genomics* 19, 427. doi: 10.1186/s12864-018-4819-2
- Ranzoni, A. M., Strzelecka, P. M., and Cvejic, A. (2019). Application of single-cell RNA sequencing methodologies in understanding haematopoiesis and immunology. *Essays Biochem.* 63, 217–225. doi: 10.1042/EBC20180072
- Rastogi, S., Shah, S., Kumar, R., Vashisth, D., Akhtar, M. Q., and Dwivedi, U. N. (2019). *Ocimum* metabolomics in response to abiotic stresses: cold, flood, drought, and salinity. *PLoS One* 14, e0210903. doi: 10.1371/journal.pone.0210903
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* 74 (12), 5463–5467.
- Schoch, C. L., Ciuffo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., et al. (2020). NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database* 2020, 1–21. doi: 10.1093/database/baaa062
- Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2012). Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi: 10.1093/bioinformatics/bts094
- Shah, M., Alharby, H. F., Hakeem, K. R., Ali, N., Rahman, I. U., and Munawar, M. (2020). *De novo* transcriptomic analysis of *Lantana camara* L. revealed candidate genes involved in phenylpropanoid biosynthesis pathway. *Sci. Rep.* 10, 467–486. doi: 10.1038/s41598-020-70635-5
- Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498 (7453), 236–240. doi: 10.1038/nature12172
- Shendure, J. (2008). The beginning of the end for microarrays. *Nat. Methods* 5, 585–587. doi: 10.1038/nmeth0708-585
- Singh, P., Singh, G., Bhandawat, A., Singh, G., Parmar, R., and Seth, R. (2017). Spatial transcriptome analysis provides insights of key gene(s) involved in steroidal saponin biosynthesis in medicinally important herb *Trillium govanianum*. *Sci. Rep.* 7, 45295. doi: 10.1038/srep45295
- Singh, S. K., Wu, Y., Ghosh, J. S., Pattanaik, S., Fisher, C., Wang, Y., et al. (2016). RNA-Sequencing reveals global transcriptomic changes in *Nicotiana tabacum* responding to topping and treatment of axillary-shoot control chemicals. *Sci. Rep.* 5, 18148. doi: 10.1038/srep18148
- Strickler, S. R., Bombarely, A., and Mueller, L. A. (2012). Designing a transcriptome next-generation sequencing project for a non-model plant species. *Am. J. Bot.* 99, 257–266. doi: 10.3732/ajb.1100292
- Stubbington, M. J., Lönnberg, T., Proserpio, V., Clare, S., Speak, A. O., Dougan, G., et al. (2016). T Cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* 13 (4), 329–332. doi: 10.1038/nmeth.3800
- Sun, H. J., and Wei, H. J. (2018). The application of RNA-seq technology in the study of the transcriptome. *chin. Foreign Med. Res.* 16, 184–187. doi: 10.14033/j.cnki.cfmr.2018.0809
- Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23, 1282–1288. doi: 10.1093/bioinformatics/btm098
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6 (5), 377–382. doi: 10.1038/nmeth.1315
- Tan, X. M., Zhou, Y. Q., Chen, J., and Guo, S. X. (2015). Advances in research on diversity of endophytic fungi from medicinal plants. *Chin. Pharm. J.* 50, 1563–1580. doi: 10.11669/cpj.2015.18.001
- Touch, B. B., Laborde, R. R., and Xu, X. (2010). Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. *PLoS One* 5, e9317. doi: 10.1371/journal.pone.0009317
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nat. Biotechnol.* 32 (4), 381. doi: 10.1038/nbt.2859
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53. doi: 10.1038/nbt.2450
- Tyagi, P., and Ranjan, R. (2021). Comparative study of the pharmacological, phytochemical, and biotechnological aspects of *Tribulus terrestris* Linn. and *Pedalium murex* Linn: An overview. *Acta Ecologica Sinica*. doi: 10.1016/j.chnaes.2021.07.008
- Tyagi, P., Singh, A., Gupta, A., Prasad, M., and Ranjan, R. (2022). Mechanism and function of salicylate in plant toward biotic stress tolerance. *Emerging Plant Growth Regulat. Agriculture*. 131–164. doi: 10.1016/B978-0-323-91005-7.00018-7
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Wang, X., Hu, H., Wu, Z., Fan, H., Wang, G., Chai, T., et al. (2021). Tissue-specific transcriptome analyses reveal candidate genes for stilbene, flavonoid and anthraquinone biosynthesis in the medicinal plant *Polygonum cuspidatum*. *BMC Genomics* 22 (1), 1–17. doi: 10.1186/s12864-021-07658-3
- Wang, Y. S., Shahid, M. Q., Ghouri, F., and Baloch, F. S. (2020). *De novo* assembly and annotation of the juvenile tuber transcriptome of a *Gastrodia elata* hybrid by RNA sequencing: detection of SSR markers. *Biochem. Genet.* 58, 914–934. doi: 10.1007/S10528-020-09983-W
- Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., and Jiao, Y. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7, 11708. doi: 10.1038/ncomms11708
- Wang, C., Zhu, J., Liu, M., Yang, Q. S., Wu, J. W., and Li, Z. G. (2018). *De novo* sequencing and transcriptome assembly of *Arisaema heterophyllum* blume and identification of genes involved in isoflavonoid biosynthesis. *Sci. Rep.* 8, 17643. doi: 10.1038/s41598-018-35664-1
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., and Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 11551162. doi: 10.1038/s41587-019-0217-9
- Wu, Y. Q., Guo, J., Zhou, Q., Xin, Y., Wang, G. B., and Xu, L. A. (2018). *De novo* transcriptome analysis revealed genes involved in flavonoid biosynthesis, transport, and regulation in *Ginkgo biloba*. *Ind. Crop Prod.* 124, 226–235. doi: 10.1016/j.indcrop.2018.07.060
- Xie, Y., Wu, G., Tang, J., Luo, R., Jordan, P., and Liu, S. (2014). SOAPdenovo-trans: *de novo* transcriptome assembly with short RNA-seq reads. *Bioinformatics* 12, 1660–1666. doi: 10.1093/bioinformatics/btu077
- Xu, Y. (2018). “Transcriptome analysis of freezing tolerance mechanism for Tibetan saussurea laniceps callus,” in *Ph.D. Thesis* (Beijing: Beijing Forestry University).
- Xu, X., Crow, M., Rice, B. R., Li, F., Harris, B., Liu, L., et al. (2021). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell* 56 (4), 557–568. doi: 10.1016/j.devcel.2020.12.015
- Xu, W. R., Li, R. M., Zhang, N. B., Ma, F., Jiao, Y. T., and Wang, Z. P. (2014). Transcriptome profiling of *Vitis amurens*, an extremely cold-tolerant Chinese wild vitis species, reveals candidate genes and events that potentially connected to cold stress. *Plant Mol. Biol.* 86, 527–541. doi: 10.1007/s11103-014-0245-2
- Xu, B., Zhang, W. Q., Feng, X. X., Wang, C. Y., Zhang, H. S., and Xu, H. T. (2014). Application progress of transcriptomic sequencing technology in maize. *J. Maize Sci.* 22, 67–72. doi: 10.13597/j.cnki.maize.science.2014.01.014
- Yang, J., Guo, Z., Luo, L., Gao, Q., Xiao, W., Wang, J., et al. (2021). Identification of QTL and candidate genes involved in early seedling growth in rice via high-density genetic mapping and RNA-seq. *Crop J.* 9 (2), 360–371. doi: 10.1016/j.cj.2020.08.010
- Yan, J. L., Qian, L. H., Zhu, W. D., Qiu, J. R., Lu, Q. J., and Wang, X. B. (2020). Integrated analysis of the transcriptome and metabolome of purple and green leaves of *Tetragonia hemsleyana* reveals gene expression patterns involved in anthocyanin biosynthesis. *PLoS One* 15, e0230154. doi: 10.1371/journal.pone.0230154
- Yousef, M., Ülgen, E., and Sezerman, O. U. (2021). CogNet: classification of gene expression data based on ranked active-subnetwork-oriented KEGG pathway enrichment analysis. *PeerJ Comput. Sci.* 7, e336. doi: 10.7717/peerj-cs.336
- Yu, G. (2020). “Gene ontology semantic similarity analysis using GOSemSim,” in *Stem cell transcriptional networks* (New York, NY: Humana), 207–215.
- Yuan, X., Li, K., Huo, W., and Lu, X. (2018). *De novo* transcriptome sequencing and analysis to identify genes involved in the biosynthesis of flavonoids in *Abrus mollis* leaves. *Russ. J. Plant Physiol.* 65, 333–344. doi: 10.1134/S1021443718030147
- Zhang, H. (2019). “The review of transcriptome sequencing: principles, history and advances,” in *IOP conference series: Earth and environmental science*, vol. 332, No. 4. (Bristol, UK: IOP Publishing), 042003.
- Zhang, T. Q., Chen, Y., Liu, Y., Lin, W. H., and Wang, J. W. (2021). Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat. Commun.* 12 (1), 1–12. doi: 10.1038/s41467-021-22352-4
- Zhang, H., Tan, J., Zhang, M., Huang, S., and Chen, X. (2020). Comparative transcriptomic analysis of two bottle gourd accessions differing in fruit size. *Gene* 11 (4), 359. doi: 10.3390/genes11040359
- Zhang, T. Q., Xu, Z. G., Shang, G. D., and Wang, J. W. (2019). A single-cell RNA sequencing profiles the developmental landscape of arabidopsis root. *Mol. Plant* 12 (5), 648–660. doi: 10.1016/j.molp.2019.04.004

Zhang, D. Y., Zhang, T. X., and Wang, G. X. (2016). Development and application of second-generation sequencing technology. *Environ. Sci. Technol.* 39, 96–102. doi: 10.3969/j.issn.1003-6504.2016.09.017

Zhao, Q. Y., Wang, Y., Kong, Y. M., Lou, D., Li, X., Hao, P., et al. (2011). Optimizing de novo transcriptome assembly from short-read RNA-seq data: a comparative study. *BMC Bioinf.* 12, 644–652. doi: 10.1186/1471-2105-12-S14-S2

Zhao, X., Yu, H., Kong, L., Liu, S., and Li, K. (2014). Comparative transcriptome analysis of two oysters, *Crassostrea gigas* and *Crassostrea hongkongensis* provides insights into adaptation to hypo-osmotic conditions. *MolEcolResour* 14, 139–149.

Zhao, B., Katuwawala, A., Oldfield, C. J., Dunker, A. K., Faraggi, E., Gsponer, J., et al. (2021). DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Res.* 49 (D1), D298–D308. doi: 10.1093/nar/gkaa931



OPEN ACCESS

EDITED BY

Md. Anowar Hossain,
University of Rajshahi, Bangladesh

REVIEWED BY

Mian Arif,
Nuclear Institute for Agriculture and
Biology, Pakistan
Hongju Zhu,
Zhengzhou Fruit Research Institute,
Chinese Academy of Agricultural
Sciences (CAAS), China
Umesh K. Reddy,
West Virginia State University,
United States

*CORRESPONDENCE

Hongyu Liu
✉ hylu@neau.edu.cn
Peng Gao
✉ gaopeng_neau@163.com
Feishi Luan
✉ luanfeishi@neau.edu.cn

SPECIALTY SECTION

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

RECEIVED 02 September 2022

ACCEPTED 20 December 2022

PUBLISHED 12 January 2023

CITATION

Amanullah S, Li S, Osae BA, Yang T,
Abbas F, Gao M, Wang X, Liu H, Gao P
and Luan F (2023) Primary mapping of
quantitative trait loci regulating
multivariate horticultural phenotypes
of watermelon (*Citrullus lanatus* L.).
Front. Plant Sci. 13:1034952.
doi: 10.3389/fpls.2022.1034952

COPYRIGHT

© 2023 Amanullah, Li, Osae, Yang,
Abbas, Gao, Wang, Liu, Gao and Luan.
This is an open-access article
distributed under the terms of the
Creative Commons Attribution License
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Primary mapping of quantitative trait loci regulating multivariate horticultural phenotypes of watermelon (*Citrullus lanatus* L.)

Sikandar Amanullah^{1,2}, Shenglong Li^{1,2}, Benjamin Agyei Osae^{1,2},
Tiantian Yang^{1,2}, Farhat Abbas³, Meiling Gao⁴,
Xuezheng Wang^{1,2}, Hongyu Liu^{1,2*},
Peng Gao^{1,2*} and Feishi Luan^{1,2*}

¹College of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin, China, ²Key Laboratory of Biology and Genetic Improvement of Horticulture Crops (Northeast Region), Ministry of Agriculture and Rural Affairs, Harbin, China, ³College of Horticulture, South China Agricultural University, Guangzhou, China, ⁴College of Life Sciences, Agriculture and Forestry, Qiqihar University, Qiqihar, China

Watermelon fruits exhibit a remarkable diversity of important horticultural phenotypes. In this study, we initiated a primary quantitative trait loci (QTL) mapping to identify the candidate regions controlling the ovary, fruit, and seed phenotypes. Whole genome sequencing (WGS) was carried out for two differentiated watermelon lines, and 350 Mb (96%) and 354 Mb (97%) of re-sequenced reads covered the reference *de novo* genome assembly, individually. A total of 45.53% non-synonymous single nucleotide polymorphism (nsSNPs) and 54.47% synonymous SNPs (sSNPs) were spotted, which produced 210 sets of novel SNP-based cleaved amplified polymorphism sequence (CAPS) markers by depicting 46.25% co-dominant polymorphism among parent lines and offspring. A biparental F_{2:3} mapping population comprised of 100 families was used for trait phenotyping and CAPS genotyping, respectively. The constructed genetic map spanned a total of 2,398.40 centimorgans (cM) in length and averaged 11.42 cM, with 95.99% genome collinearity. A total of 33 QTLs were identified at different genetic positions across the eight chromosomes of watermelon (Chr-01, Chr-02, Chr-04, Chr-05, Chr-06, Chr-07, Chr-10, and Chr-11); among them, eight QTLs of the ovary, sixteen QTLs of the fruit, and nine QTLs of the seed related phenotypes were classified with 5.32–25.99% phenotypic variance explained (PVE). However, twenty-four QTLs were identified as major-effect and nine QTLs were mapped as minor-effect QTLs across the flanking regions of CAPS markers. Some QTLs were exhibited as tightly localized across the nearby genetic regions and explained the pleiotropic effects of multigenic nature. The flanking QTL markers also depicted significant allele specific contributions and accountable genes were predicted for respective traits. Gene Ontology (GO) functional enrichment was categorized in molecular function (MF), cellular components (CC), and biological process (BP); however, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were classified into three main

classes of metabolism, genetic information processing, and brite hierarchies. The principal component analysis (PCA) of multivariate phenotypes widely demonstrated the major variability, consistent with the identified QTL regions. In short, we assumed that our identified QTL regions provide valuable genetic insights regarding the watermelon phenotypes and fine genetic mapping could be used to confirm them.

KEYWORDS

watermelon (*Citrullus lanatus* L.), ovary, fruit, seed, genetic markers, QTL

Introduction

Watermelon (*Citrullus lanatus* L.) is an annual fruit crop that belongs to the Cucurbitaceae family, and it is a highly cultivated fruit with more than a thousand varieties (Pan et al., 2020).

Plants are mostly grown in tropical to temperate climates under high temperature requirements of about >25°C (77°F) to thrive (Walters et al., 2021). China is the major consumer and producer of watermelon, and the cultivation area accounts for nearly half (46.04%) of the world's watermelon planting area. Among the world's watermelon producing countries, China ranks first with 60 million tons production (Grumet et al., 2021).

The better quality and sweeter tasting watermelons are not only the result of natural selection (Chomicki and Renner, 2015), but they also depend upon artificial selection during their adaptation to the diverse environments (Pan et al., 2020). Since the start of watermelon domestication, the cultivated watermelons (citron, dessert, and egusi) have been classified as sub-species of the main species (*Citrullus lanatus*) (Fursa, 1972), exhibiting remarkable phenotypic diversity (Sandlin et al., 2012). Although there is some cross-ability among them, the genome dataset recommends their separation into three dissimilar species (Renner et al., 2014): (1) *C. lanatus* (Thunb.) Matsum. & Nakai is the dessert watermelon (also known as *C. lanatus* subsp. *vulgaris*), (2) *C. amarus* Schrad. is the citron watermelon (also known as *C. lanatus* subsp. *lanatus*), and (3) *C. mucospermus* Fursa is the egusi watermelon (also called as *C. lanatus* subsp. *mucospermus*) (Paris, 2015).

Watermelon genotypes have short and long ovaries with differentiated weight (Weetman, 1937; Osae et al., 2022), and flowers of some botanical varieties are monoecious or andromonoecious (Poole and Grimball, 1938; Aguado et al., 2020). The developed ovary and mature fruit display a high correlation since pre-anthesis. However, the obvious structure of mature fruit is reflected by gradual cell division and cell size elongation during each developmental stage (McKay, 1936; Weetman, 1937; Dou et al., 2018b). Fruit size and shape

indexes vary within elongated, blocky, and rounded fruits effectively classified by representing a highly quantitative genetic architecture regulated by the contribution of polygenic architecture of allelic variants (Wehner et al., 2001; Gusmini and Wehner, 2006; Gusmini and Wehner, 2007; Sandlin et al., 2012; Kim et al., 2015).

Further, watermelon fruit-related quality traits are extremely connected with each other and significantly fascinate the consumer's attention. Fruit weight varies from 1 kilogram (kg) to more than 10 kg and mainly depends upon the fruit size and shape, affecting the total crop yield (Gusmini and Wehner, 2006; Osae et al., 2022). Fruit flesh firmness is a standard quantitative trait that determines the edible quality of watermelon, and variations are genetically inherited with genotypes and environments. It is jointly regulated through polygenes and multifarious metabolic networks (Zhu et al., 2017) and the flesh cell size is increased by enhanced vacuolation at the initial days after pollination (DAP) (Sun et al., 2020). Fruit rind texture is highly diversified in numerous cultivars, which explains the good relationship with postharvest life. Watermelon cultivars with high rind-hardness are less prone to cracking and have better resistance to long-term storage (Liao et al., 2019; Yang et al., 2021). It was also shown that flesh firmness variation is an uneven and multifaceted phenomenon that is particularly shifted through inherited genetics (Khadiji-Khub, 2015), and endogenous lignin accumulation in peel stone cells form the hard ultrastructure of rind (Gao et al., 2013; Yang et al., 2022). An ethylene bio-synthesis related transcription factors "Md-ACO1, Md-ACS1, and MADS-box" contributed in the fruit ripening stages, nutrient metabolism, and hormone signal transduction, that mainly trigger the internal respiratory mechanism and led to a low firmness level in the fruit (Costa et al., 2005; Costa et al., 2008; Moreno et al., 2008; Costa et al., 2010; Vegas et al., 2013; Liao et al., 2019).

Fruit rind appearance can be gray, striped, and solid, or can be light-green, medium green, or dark green (Liang et al., 2022). Rind stries pattern can be blotchy, wavy, narrow, medium, or wide, depending on their presence on the rind surface (Zhang et al., 2018; Maragal et al., 2022). Multiple genetic basis of

watermelon rind stripe pattern and rind color have been observed through differential gene architecture (Guner and Wehner, 2004; Dou et al., 2018a; Liu et al., 2020; Wang et al., 2022). Furthermore, the watermelon flesh color is a primary determinant of edible quality and is concerned with the carotenoids accumulation in internal chromoplast cells (Tadmor et al., 2005; Fang et al., 2020). Many accessions have different flesh color gradients, e.g., white, salmon-yellow, orange, red, canary yellow, pale green, and are thought to be polygenic (Gusmini and Wehner, 2006; Liu et al., 2015; Pei et al., 2021; Liang et al., 2022). The salmon-yellow color of the flesh is mainly developed by accumulation of pro-lycopene (tetra-cis-lycopene); orange develops from pro-lycopene and rarely from β -carotene, red from lycopene; and canary yellow watermelon results from accumulation of small amounts of xanthophylls and β -carotene (Tadmor et al., 2005; Bang et al., 2010; Branham et al., 2017; Subburaj et al., 2019). The natural variation in carotenoid accumulation takes place among the heirloom and exotic watermelon accessions, which induce an extensive range of flesh colors regulated by multiple genes (Fang et al., 2022).

Seed is an integral part of the plant life cycle that determines the vigorous growth and development of crop plants (Xiang et al., 2002). The size and shape of seed greatly vary in different crops; however, these were always considered as primary target for breeding selection throughout domestication (Gómez, 2004; Moles et al., 2005; Xia et al., 2018). The wild-type watermelons usually bear small and rounded-shaped seeds, while the improved cultivars bear much larger and variegated-shaped seeds (Guo et al., 2020). Most of the variations in seed size and shape, seed oil content, seed coat thickness, seed weight, and seed thickness are the effective outcomes of natural and artificial selection during adaptation to different environmental localities (Meru and McGregor, 2013; Guo et al., 2020), and are thought to be influenced by polygenic alleles in a moderate-type dominance fashion (Baboli and Kordi, 2010; Tian et al., 2012). However, the genetic basis of seed-related content is less known in fruit and vegetable crops and still needs much more attention for an in-depth understanding.

The primary mapping of multifaceted QTLs is a traditional molecular technique that has been well-employed for the identification of candidate genomic regions controlling various crop traits based on different types of genetic markers (Olsen and Wendel, 2013). Different generations of DNA-based genetic markers (RFLPs, RAPDs, AFLPs, SSRs, SNPs, and CAPSs) and derived mapping populations have been introduced and utilized for effective genetic mapping of major loci controlling watermelon traits, respectively (Hashizume et al., 1996; Levi et al., 2001; Levi et al., 2006). Whole genome sequencing approach coupled with the published *de novo* reference genome assembly of watermelon has assisted in developing the base-by-base SNP markers and high-resolution genetic linkage mapping (Guo et al., 2013). In recent years, SNP based codominant CAPS markers have been emerged as an efficient

DNA markers for dissecting the major-effect QTLs of regulating the important qualitative and quantitative traits of melon and watermelon (Bang et al., 2007; Liu et al., 2015; Cheng et al., 2016; Liu et al., 2016; Li et al., 2017; Zhang et al., 2018; Liu et al., 2019; Luan et al., 2019; Wang et al., 2019; Sun et al., 2020; Amanullah et al., 2020; Amanullah et al., 2021; Zhang et al., 2021; Pei et al., 2021; Osae et al., 2021; Yang et al., 2021; Amanullah et al., 2022; Osae et al., 2022; Lv et al., 2022).

Until now, many researchers have performed significant molecular mapping studies, but the detailed genetic basis of many unexplored cultivars has not been completely resolved. The current study was aimed at primary mapping of novel genetic regions controlling watermelon ovary, fruit, and seed phenotypes by using novel SNP-derived CAPS markers and genetic segregation analysis in a developed biparental $F_{2:3}$ mapping population.

Materials and methods

Plant materials and mapping population

Two comparative watermelon parent lines “W1-38 as P_1 (cultivated female parent line, *Citrullus lanatus* L.) and PI542119 as P_2 (male parent line, forage watermelon, *Citrullus* var. *citroides*)” were particularly chosen as experimental materials based on the primarily distinguished ovary, fruit, and seed phenotypes (Figure 1, Supplementary Table S1). These two parent lines were then crossed to develop their F_1 offspring and $F_{2:3}$ mapping population, respectively. A field experiment was carried out in a big plastic greenhouse at XiangYang Agricultural Farm of Northeast Agricultural University, Harbin, China.

In the first year of the experiment, plants of both parent materials were cultivated, total genomic DNA (GDNA) was extracted, whole genome sequencing was performed, and newly exported SNP-CAPS markers were validated for molecular genotyping, respectively. The raw data of DNA sequencing (DNA-seq) was uploaded to the online Sequence Read Archive (SRA) database (Accession: PRJNA878948, containing 2 biosamples) of the National Center for Biotechnology Information (NCBI). In the second year of experiment, seedlings of P_1 (15-plants), P_2 (15-plants), F_1 (15-plants), and 100 $F_{2:3}$ mapping families (5-plants of each family) were raised and one month old seedlings (true-leaves stages) were shifted into greenhouse. To ensure the healthy growth of plants, a mixture of 60% loamy soil, 30% compost, and 10% potting mixture (peat moss, perlite, and vermiculite) was used in the raised beds of the greenhouse. All the plants of the respective mapping families were grown by maintaining the row spacing (60 cm) and plant spacing (70 cm) in the planting geometry of a complete randomized block design (CRBD), and common horticultural techniques were subsequently applied, respectively.

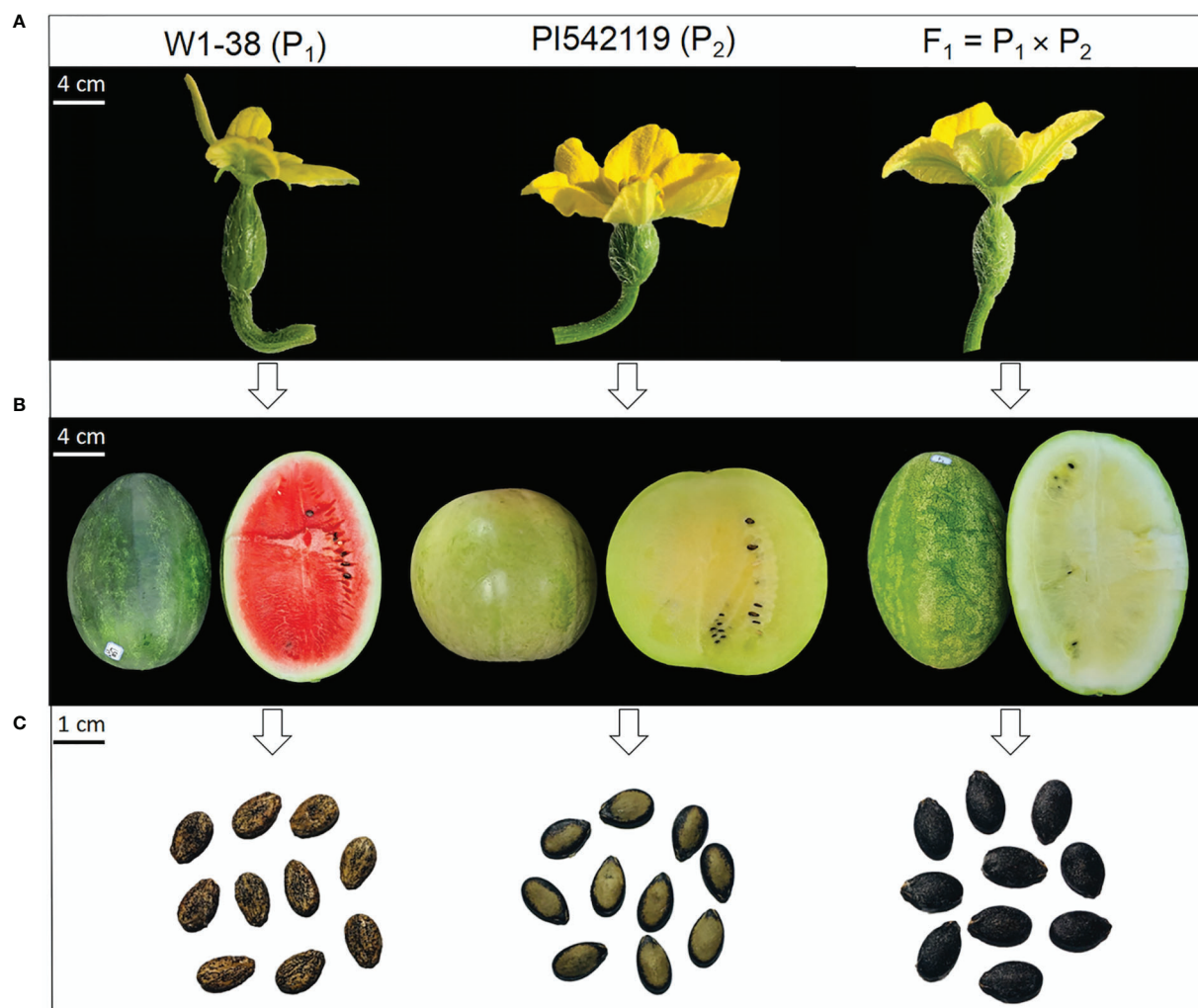


FIGURE 1
Primary phenotypes of comparative watermelon parent lines and developed F1 off-spring. (A) Ovary traits. (B) Fruit traits. (C) Seed traits, respectively.

Evaluation of phenotypes

The ovary, fruit, and seed phenotypes of P_1 , P_2 , F_1 , and $F_{2,3}$ mapping population were evaluated as earlier reported method (Pei et al., 2021; Liang et al., 2022; Maragal et al., 2022; Osae et al., 2022). For the ovary phenotypes, the anthesis period of each plant was observed on a daily basis, then fresh ovaries were plucked just before the day of flower opening. Ovary weight (OWt) was freshly weighed in grams (gm), ovary length (OL) was measured from the proximal end to the distal end, and ovary width (OW) was checked at the widest part from one edge to another edge. OL and OW were measured in millimeter (mm) and ovary shape index (OSI) was the ratio of ovary length to width (OL/OW).

For the fruit phenotypes, flowers were pollinated and mature watermelon fruits were harvested between 35 and 45 days after

pollination (DAP). One fruit from each plant was harvested and fruit weight (FWt) was freshly weighed in kilogram (kg) units by using a portable digital weighing machine having high measuring accuracy. Fruit length (FL) was measured from top (stem end to lower end) to bottom (lower end) section, fruit width (FW) was checked from equatorial fruit diameter in centimeter (cm) units, and fruit shape index (FSI) was evaluated by the ratio of FL/FW. Fruit rind thickness (FRT) was checked by observing the distance between exocarp and mesocarp layers and recorded in millimeter (mm) units. Fruit flesh firmness (FFF) was tested by operating the digital fruit firmness tester (FR-5120) having a high-precision sensor with peak hold. The tip of the meter was steeply inserted at five different points of flesh (right, left, mid-point, top, bottom) and the firmness indicator mean was expressed in kg/cm^2 . Brix of fruits (FBR) was quantified from the squeezed juice samples and

checked in percentage (%) using the hand-held refractometer (Atago Co., Ltd., Tokyo, Japan). The qualitative fruit traits were observed based on their visual phenotypes. Fruit rind color (FRC) was checked by observing the outer rind surface color “dark-green (DG), light-green (LG), and intermediate (I)” and scoring as 3 (DG), 1 (LG), and 2 (I). Fruit rind stripes (FRS) were observed based on their pattern differences on rind surfaces as “striped wavy (W) and non-striped blotchy (B)”, and visually scored as 0 (W), 1 (B). Fruit flesh color (FFC) was judged by the naked eye using the differentiated color scoring rate of “(3) red (R), (1) pale green (PG), and (2) yellow (Y)”, respectively.

For the seed phenotypes, seeds were extracted from the flesh cavity of each harvested fruit, then gently washed, and dried in partial sunlight. The weight of 50-seeds (SWt) of each fruit was weighed in grams by using the small digital weighing scale. Seed thickness (ST), seed length (SL), and seed width (SW) phenotypes were measured at their widest and longest axis, and the seed shape index (SSI) was calculated by dividing the SL over the SW (SL/SW). The seed coat of each seed was gently removed and the seed coat thickness (SCT) was also measured. The morphometric data of ST, SL, SW, and SCT was recorded in millimeter (mm) units by using a digital vernier caliper.

Statistical data analysis

The multivariate phenotypic datasets were computed on Microsoft Excel (v2016) and statistical analysis were performed. The graphic illustration of analysis was done by using the R language tool (v3.2.3) coupled with the interface of RStudio (v1.0.143) (Wickham, 2009; R Core Team, 2018). The normality of the frequency distribution was checked based on the Shapiro-Wilk test at a significance *p*-value of <0.05. The representative qualitative and quantitative datasets with no missing values were used for the major variability by means of principal component analysis (PCA), respectively.

Genomic sequencing and SNP mining

A sufficient amount (0.2 g) of juvenile leaves were sampled from the seedlings of comparative genotypes and high-quality GDNA was isolated by using the cetyl trimethyl ammonium bromide (CTAB) method (Allen et al., 2006). The density of isolated GDNA was checked on 1% agarose gel electrophoresis, and quickly stored at an ultra-low temperature (-80°C) before being used for molecular genotyping experiments. The quantified GDNA of both parent materials was fragmented by using the Diagenode Bioruptor Sonication device, and a sequencing library of 200–300 base pairs (bp) was pooled up by using the polymerase chain reaction (PCR) amplification (Amanullah et al., 2020). Whole genome sequencing of 10×

genomic depth per sample was performed by using the High-throughput Illumina Sequencing™ 2000 technology at Beijing Genomics Institutes (BGI), Guangdong, P. R. China.

The quality of paired-end re-sequencing reads of comparative watermelon lines was checked and aligned over the downloaded *de novo* assembled genomic directory of watermelon (97103, v2 genome) using the perl script of Burrows-Wheeler Aligner (BWA) software package (v0.7.15-r1140) (Li and Durbin, 2009). Then, flexible generic sequence-alignment and map (SAM) text files and binary-alignment and map (BAM) index reads were sorted using the SAMtools program (v1.15.1) (<http://www.htslib.org/>). The whole genome filtered sequenced files were aligned and core sets of SNPs were mined using the bioinformatic tools “Genome Analysis Toolkit (GATK, v3.1-1) and VarScan (v2.0)” (McKenna et al., 2010; Ruffalo et al., 2011). SNPs were exported in variant call format (VCF) files by using the latest SnpEff (v5.1) (Cingolani et al., 2012) and annotated within intron and exon regions based on their genetic variant effects, respectively.

CAPS markers validation

Whole genome CAPS markers were derived based on major SNP sequences, as reported in our earlier study (Amanullah et al., 2022). In brief, a total of 20–25 physical positions of SNP sequences “before and after 500 base pairs (bp)” of both parent lines were aligned across the whole genome chromosomal length using SNP2CAPS software (Thiel et al., 2004). The suitable SNP sequences were chosen and converted into CAPS markers and DNA-based standard PCR primers were exported using the upgraded Primer Premier software (v6.25). The best quality primers were oligo-synthesized by Sangon Biotechnology LTD., and then codominant polymorphic primers were screened based on distinct bands using the optimal CAPS PCR reactions in 10 µL mixture as previously stated (Amanullah et al., 2022), e.g., 0.5 µL of each primer sequence (forward + reverse), 1 µL PCR buffer, 0.1 µL *Taq* DNA-polymerase, 0.15 µL dNTPs, 1 µL GDNA, and 6.75 µL deionized water. The enzyme digested PCR products were cleaved on gel electrophoresis, and the digested bands were captured on the Invitrogen iBright Imaging Systems.

QTL mapping and putative genes prediction

A watermelon linkage map was constructed and primary QTL mapping was done using QTL IciMapping software (Meng et al., 2015). In brief, the dataset of coded allelic bands of 100 F₂ mapping population was recorded and arranged across the chromosomal length of watermelon genome. The recombination frequency of the linkage map was estimated among the adjacent markers based on

genetic orders and intervals in centimorgans (cM). QTLs of ovary, fruit, and seed phenotypes were mapped on each chromosome, using the 1 cM sliding scale and 1000 permutation testing ($p = 0.05$). The threshold of a logarithm of odds (LOD) value of 2.5 and a genome-wide Type I (false-positive) error rate were used for checking the confidence intervals and genetic effects of identified QTLs between potential markers. The mapped QTLs with high LOD values (above 3) and phenotypic variance explained (PVE, >10%) were designated as major QTLs, and QTLs with low LODs and PVE (<10%) were designated as minor QTLs. Finally, the mapped QTLs were abbreviated as follows; the name of the trait, its position on the chromosome, and the number of QTLs, as reported earlier (Amanullah et al., 2022).

For the prediction of putative genes and their comprehensive genomic annotation, the physical sequences of adjacent QTL markers and comparative files of re-sequenced watermelon lines were pairwise aligned across the consensus directory of the watermelon genome (97103, v2), and genes numbers were visually tracked using the scalable visualization tool “Integrative Genomics Viewer (IGV, v2.12.2)” (Robinson et al., 2017). All the predicted genes were annotated based on the Gene Ontology (GO) function enrichments and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, as earlier described method (Yang et al., 2022).

Results

Analysis of WGS and SNPs identification

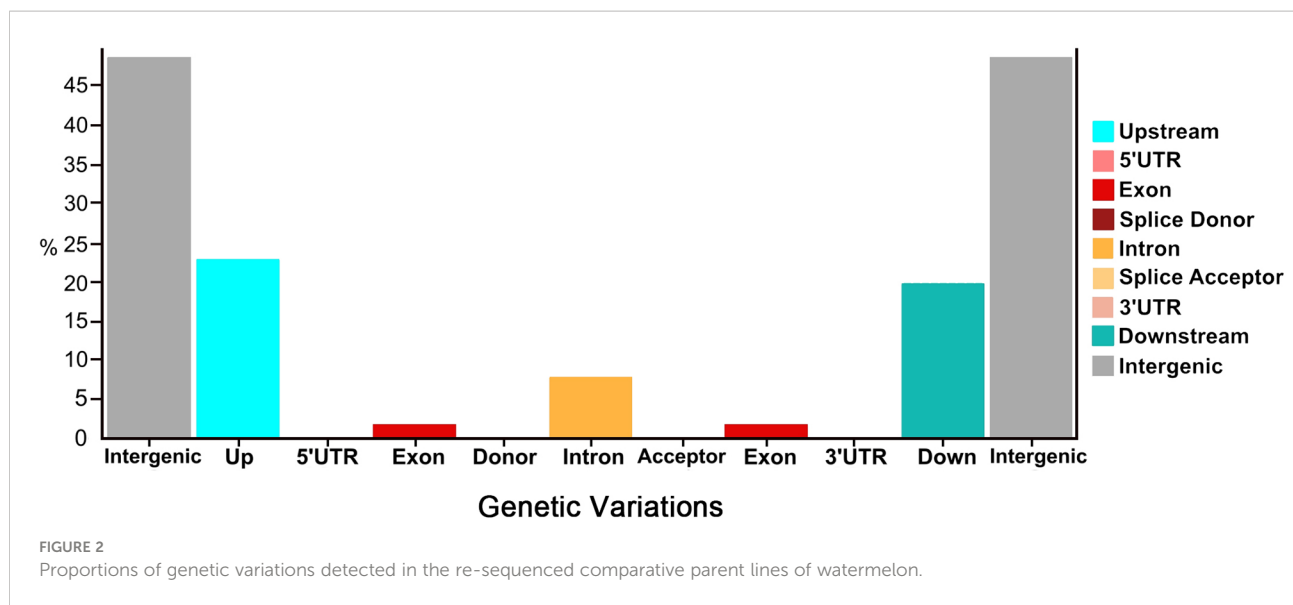
A total of 16.40 gigabytes (GB) of molecular data was obtained from whole-genome resequencing of two

comparative parent lines and exposed 350 Mb and 354 Mb of genomic coverage across the watermelon genome directory (97103, v2), respectively. The mapping statistics of the WGS can be seen in [Supplementary Table S2](#). Further, the detected SNPs and CAPS loci pairs were subsequently filtered through bio-informatics analysis, and the SNP distribution is shown in [Table 1](#). A total of 352,177,534 bp of genomic length was effectively spanned across the whole-genome chromosomes (Chr-01 to Chr-11); among them, Chr-02 depicted the highest coverage with 36.80 Mb (36,805,829 bp) length, Chr-04 showed the minimum coverage of 26.10 Mb (26,100,705 bp), and other chromosomes demonstrated the differentiated coverage of genetic length (Mb) over the reference genome assembly.

Overall, a total of 45.53% nsSNPs (missense) and 54.47% sSNPs (non-sense (53.94%) and silent mutations (0.53%)) were identified by their functional classes. These genetic variants were categorized into differential proportions with their number of produced effects by type and region ([Figure 2](#)). Regarding the genetic variants effects by type, the highest percentage was depicted in intergenic_region (48.44%), upstream_region (22.51%), and downstream_region (19.63%), while the minimum variants were exhibited in intron_region (7.75%) and exon_region (1.67%), followed by very less percentages of splice_site_acceptor, splice_site_donor, 5_prime_UTR_variant, and 3_prime_UTR_variant. Regarding the genetic variants effects by region, the highest percentage was depicted in intergenic_region (48.50%), upstream_region (22.54%), and downstream_region (19.65%), while minimum type of effects were present in intron_region (7.66%) and exon_region (1.65%), followed by very less percentages of splice_site_acceptor, splice_site_donor, 5_prime_UTR_variant, and 3_prime_UTR_variant, respectively.

TABLE 1 Distribution of SNPs and derived CAPS loci pairs across the total genetic length of watermelon genome.

| Chromosomes | Genomic length | Total SNPs | SNPs rate/kb | CAPS loci |
|---|--------------------|------------------|--------------|----------------|
| Chr-01 | 35,825,788 | 406,622 | 11 | 18,350 |
| Chr-02 | 36,805,829 | 411,170 | 11 | 13,848 |
| Chr-03 | 30,762,151 | 331,710 | 11 | 16,145 |
| Chr-04 | 26,100,705 | 262,652 | 10 | 14,015 |
| Chr-05 | 34,767,876 | 400,668 | 12 | 18,309 |
| Chr-06 | 28,405,350 | 215,003 | 08 | 11,162 |
| Chr-07 | 30,829,010 | 338,621 | 11 | 16,276 |
| Chr-08 | 27,200,117 | 311,345 | 11 | 10,578 |
| Chr-09 | 36,616,462 | 421,504 | 12 | 20,720 |
| Chr-10 | 34,089,233 | 370,015 | 11 | 17,231 |
| Chr-11 | 30,775,013 | 315,340 | 10 | 15,517 |
| Total | 352,177,534 | 3,784,650 | 11 | 172,151 |
| The bold values denote the total calculation of each column data. | | | | |



Analysis of CAPS markers

A total of 172,151 CAPS loci pairs were detected based on the appropriate cleaved sites of restriction endonucleases (RE) onto the whole-genome chromosomes. Overall, a good average of CAPSs was also observed across the remaining chromosomes; however, a maximum of 20,720 CAPSs were observed within the genetic length of Chr-09 and a minimum of 10,578 CAPSs were observed across Chr-08. A total of six different restriction enzymes (*EcoR* I, *BsaH* I, *Msp* I, *Hind* II, *BamH* I, *Pst* I) were corresponded for cleaving the DNA within or adjacent to that site, and yielded the suitable pairs of CAPS markers sequences ($n=454$). The amplified PCR and digested products were assessed and a total of 210 sets of codominant adjacent CAPS markers were confirmed (Supplementary Table S2), by depicting 46.25% polymorphism among the different base pairs (bp) of P_1 , P_2 , and F_1 , respectively.

Analysis of constructed linkage map

In total, 210 pairs of codominant CAPS markers were genotyped within biparental F_2 mapping families ($n=100$) and linkage map was constructed (Figure 3). A large portion of markers were localized over the developed linkage map (Chr-01 to Chr-11) according to their fragmented length (bp) and their physical positions in the sequences, and they showed perfect genotypic association and genome collinearity ($R^2 = 0.924$ to 0.995). A total of 195 markers (92.85%) displayed the perfect segregation ratio (p -value of >0.05) and just 15 markers (7.14%) exhibited biased segregation. The constructed linkage map enclosed a total length of 2,398.40 cM, having an average of 11.42 cM. However, whole-genome linkage group (LG) length

varied from 147.68 cM (Chr-03) to 289.23 cM (Chr-02) length. The total number of codominant CAPS markers varied from a minimum of 11 (Chr-05 with 23.73 cM) to a maximum of 26 (Chr-04 with 195.31 cM length); however, an average of 10–15 markers were normally positioned on most of the chromosomes.

Further, a heat map of 11 linkage groups (LGs) was generated based on pair-wise recombination values that illustrated the genomic collinearity of markers. The visualized linkage relations in the heat map exhibited the relationship between recombination of markers on each chromosome and this was used to identify the potential marker ordering. The order of markers on line and row was arranged according to their genetic distance. The closer the distance between different markers, the lower the recombination rate (indicated by the yellow color) was observed; however, a higher recombination rate is indicated by the purple color (Figure 4). The linkage rectangles (upper-left and lower-right) in the heat map generally indicate that the construction of our genetic map was accurate since the linkage groups were easily visualized.

Phenotypic variation analysis

The multivariate quantitative and qualitative datasets of ovary traits (OWt, OL, OW, OSI), fruit traits (FWt, FL, FW, FSI, FRS, FRC, FFC, FFF, FBR, FRT), and seed traits (SWt, SL, SW, SSI) were analyzed among the comparative watermelon lines (P_1 and P_2), F_1 (off-spring), and $F_{2,3}$ mapping families (Figures 5–7; Supplementary Tables S4–S5). A normal frequency distribution, strong transgressive segregation, significant correlation coefficients, and strong patterns of explained variability were observed among the phenotypic datasets, respectively.

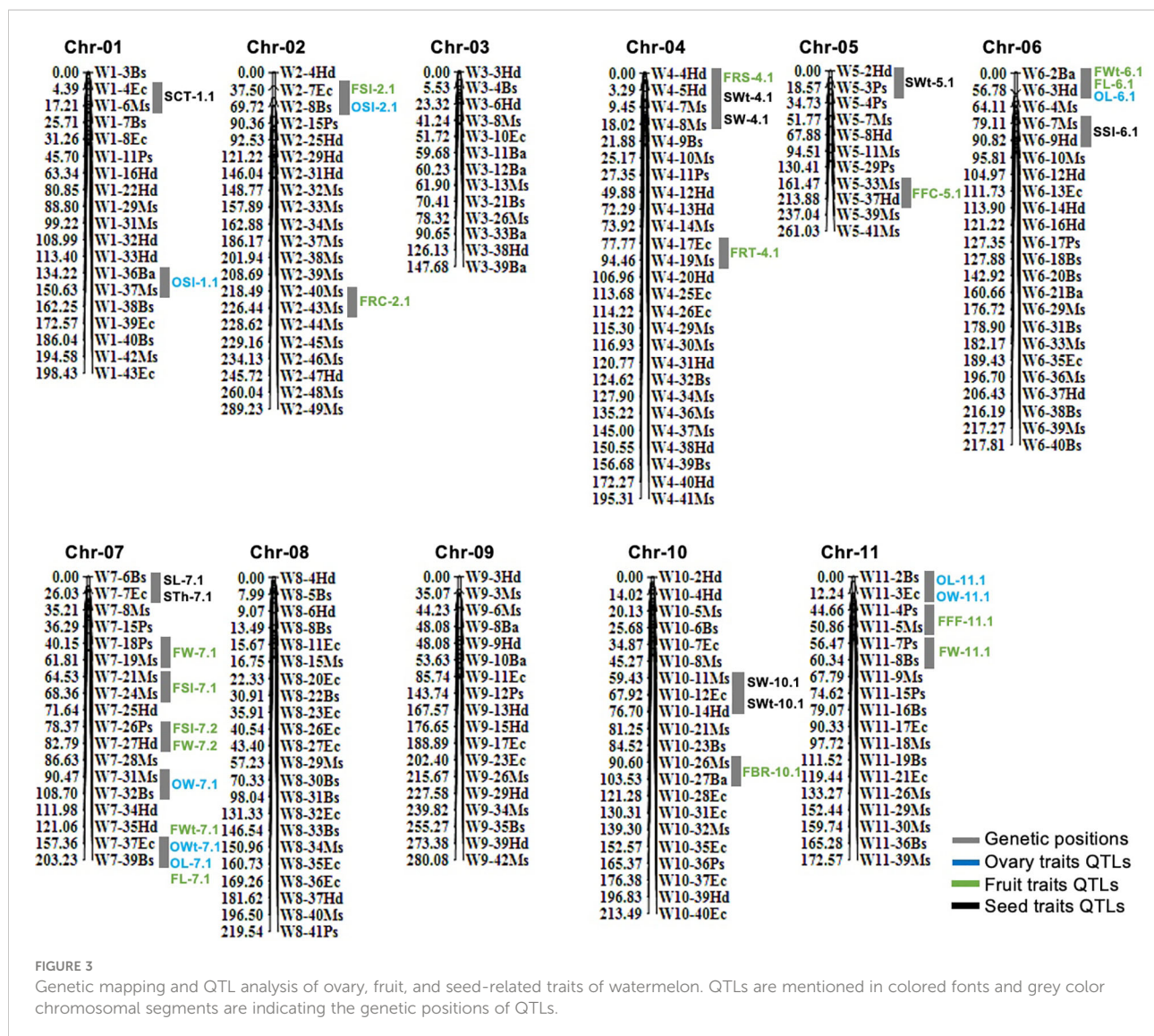


FIGURE 3

Genetic mapping and QTL analysis of ovary, fruit, and seed-related traits of watermelon. QTLs are mentioned in colored fonts and grey color chromosomal segments are indicating the genetic positions of QTLs.

Analysis of ovary phenotypes

For the OWt, the mean values of parent lines (P_1 and P_2) were 0.73 ± 0.35 and 1.55 ± 0.30 , and F_1 offspring showed the different mean value of 0.88 ± 0.11 , respectively. The overall OWt mean value of $F_{2.3}$ mapping population was 1.10 ± 0.34 and major variability ranged from 0.30 to 1.90 gm (Figure 5A). For the OL, the mean values of P_1 , P_2 , and F_1 were differentially noticed as 16.13 ± 1.22 , 10.20 ± 0.35 , and 12.43 ± 1.26 , respectively. The overall OL mean of $F_{2.3}$ mapping population was 14.83 ± 2.76 and variations ranged from 8.75 to 21.00 mm (Figure 5B). For the OW, the mean values of P_1 , P_2 , and F_1 were also different (8.42 ± 0.34 , 10.49 ± 0.11 , and 7.45 ± 0.21), respectively. The overall OW mean values of $F_{2.3}$ mapping population was 10.49 ± 2.19 and a varied range was seen with minimum of 6.29 to maximum of 15 mm (Figure 5C). The OSI of both parent lines and F_1 was also dissimilar (1.92 ± 0.13 , 1.05 ± 0.37 ,

and 1.67 ± 0.12); however, the overall mean value of $F_{2.3}$ mapping population was 1.44 ± 0.24 with a range of 0.98–2.14 (Figure 5D), respectively. Overall, the ovary associated phenotypes showed transgressive segregation and normal frequency distributions, indicating the inheritance of quantitative genetics with partial polygenic phenomena.

Analysis of fruit phenotypes

The mean values of the FWT in parent lines (P_1 and P_2) were quite different (4.32 ± 0.14 and 5.73 ± 0.15), and their F_1 offspring also showed more value (8.96 ± 0.11) than parent lines. The FWT mean value of $F_{2.3}$ mapping population was 4.60 ± 1.62 and variations were ranged from 1.85 to 8.15 kg. A transgressive segregation and a normal frequency distribution were observed,

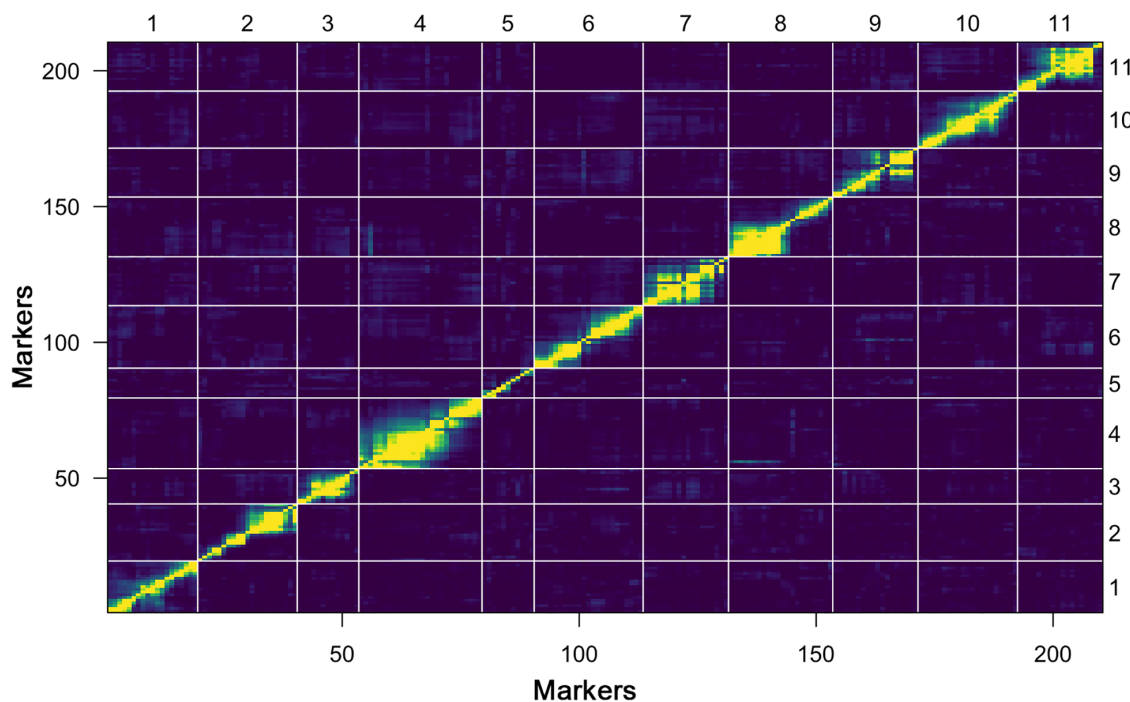


FIGURE 4
Plot of estimated pair-wise recombination fractions and LOD scores (upper-left and lower-right rectangles) of constructed genetic map.

and the major genetic variation effect in $F_{2,3}$ fruit with medium to heavy weighted fruits was seemed to be inherited through F_1 offspring (Figure 6A). FFF mean values of both parents were 1.29 ± 0.03 and 3.90 ± 0.13 , and F_1 offspring showed dissimilar mean value (3.37 ± 0.07). Among the $F_{2,3}$ population, FFF mean value of fruit flesh rupturing force was ranged from 1.85 to 5.30 kg/cm^2 . A normal distribution frequency with lower to higher firmness values also illustrated the genetics of varied flesh texture (soft and hard) in fruits of $F_{2,3}$ populations (Figure 6B). The flesh firmness variation was seemed as normally inherited by the genetic effects of wild male parent line with hard flesh firmness.

FRT was different in P_1 , P_2 , and F_1 offspring (11.66 ± 0.58 and 8.16 ± 0.29 , and 11.33 ± 0.58), respectively. However, the mean value of an $F_{2,3}$ mapping population was noticed as 8.34 ± 2.88 , with a varied range of 4 to 16 mm, exhibiting a normal quantitative frequency distribution (Figure 6C). Further, according to the visual observation of dissected fruits, most of the fruits of $F_{2,3}$ mapping population showed more rind thickness and seemed that the cultivated female parent line (P_1 with more rind thickness) and F_1 depicted the major dominant effect of FRT inheritance; however, a few fruits showed less rind thickness. FBR was dissimilar in both parents (10.72 ± 0.20 and 3.80 ± 0.87) and their F_1 offspring exhibited moderate sweetness level with mean value of 6.07 ± 0.11 . The mean value of FBR of $F_{2,3}$ population was noticed as 5.80 ± 1.63 , with a varied range of 3 to 10%, respectively. The normal frequency distribution of Brix % explained a quantitative characteristic in $F_{2,3}$ mapping

population (Figure 6D), indicating the inherited nature of the female and sweet cultivated parent line (P_1).

FL means of both parent lines were 26.17 ± 0.29 and 22.83 ± 0.76 and F_1 offspring showed a different mean value of 33.83 ± 0.78 . The overall FL mean value of $F_{2,3}$ mapping population was 22.92 ± 4.07 and variations ranged from 14 to 31.50 cm, exhibiting the uniform frequency and transgressive segregation (Figure 6E). FW means were also different in both parents (18.05 ± 0.51 , 21.23 ± 0.75) and F_1 (23.51 ± 0.50), respectively. However, the FW mean value was 17.16 ± 3.41 and the observed normal frequency distribution was ranged 10–23 cm, explaining the transgressive segregation in $F_{2,3}$ population (Figure 6F). For both FL and FW traits, F_1 fruits showed heterosis by describing the superior phenotypes relative to the parent lines and genetic effects of dominance were observed for obvious length and width in $F_{2,3}$ family fruits, correspondingly. The FSI (FL/FW) of P_1 , P_2 , and F_1 was manually deliberated and their mean values were 1.30 ± 0.38 , 1.08 ± 0.36 , 1.50 ± 0.25 ; however, the $F_{2,3}$ mapping population mean value was 1.36 ± 0.25 , having a range from 0.86 to 2.13, respectively (Figure 6G).

FRC of both parent lines exhibited different rind colors “dark-green (DG) and light green (LG)” and developed F_1 offspring fruits showed an intermediate (I) color. The genetic inheritance of rind color in the $F_{2,3}$ family fruits was based upon homozygous patterns in parent genotype of P_1 (homozygous with dark green color), P_2 (homozygous light green color), and F_1 (heterozygous with intermediate color), e.g., 21 fruits with

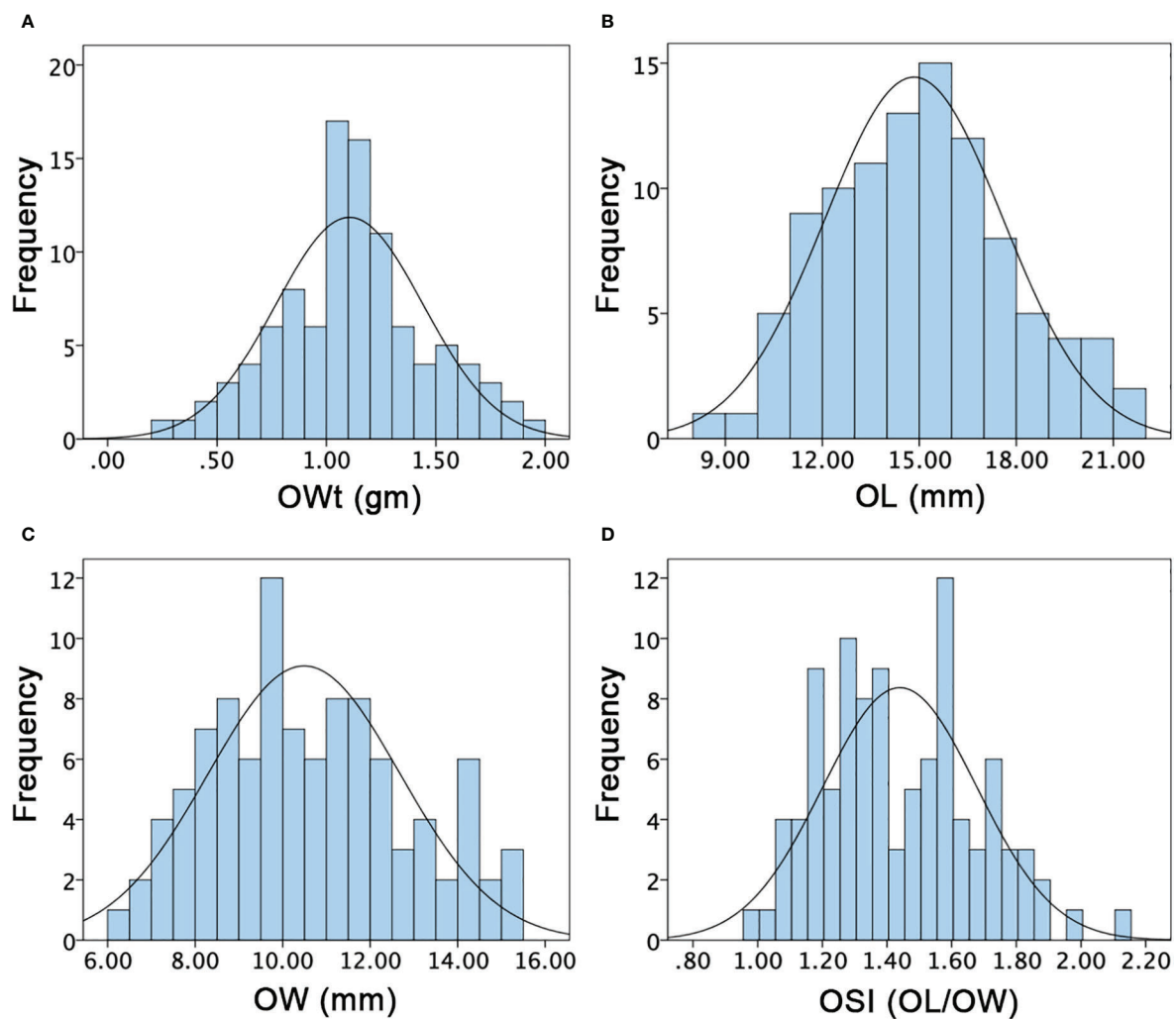


FIGURE 5

Histograms of the frequency distribution of ovary-related phenotypes in a developed $F_{2:3}$ mapping population. (A) Ovary weight. (B) Ovary length. (C) Ovary width. (D) Ovary shape index.

dark-green, 53 fruits with intermediate, and 26 fruits with light-green, exhibiting the 1:2:1 segregation ratio at P -value (0.12) and χ^2 value (4.39) (Supplementary Table S5). For the FFC, both parent lines illustrated dissimilar flesh colors “red (R), pale-green (PG)”, and their resultant F_1 offspring showed a yellow (Y) color. In the developed $F_{2:3}$ mapping population, most of the dissected fruits exhibited different and irregular flesh colors in the center part and placental tissues at the cross-sectional portion. So, the visible flesh color covering the maximum portion was considered as the dominant color and finally three colors were categorized, exhibiting the fitted genetic segregation of 1:2:1 at P -value (0.54) and χ^2 value (1.25), e.g., 21 fruits with red flesh, 56 fruits with yellow, and 23 fruits with pale-green flesh, proposing the major dominance of yellow flesh color

(Supplementary Table S5). For the FRS appearance, fruits of P_1 parent line exhibited wavy striped (W) and P_2 parent line showed non-striped blotchy (B) type appearances; however, F_1 fruits also showed wavy striped appearance but somewhat different from the female parent (P_1). In the fruits of the developed $F_{2:3}$ mapping population, genetic analysis showed that a total of 49 fruits were with homozygous pattern of wavy stripes, 49 fruits were with wavy striped pattern related to F_1 (striped but somewhat different from P_1), and 30 fruits were with homozygous pattern of non-striped blotchy type, exhibiting the 1:2:1 segregation ratio (Supplementary Table S5). Thus, we assumed that the dominance effect of striped rind appearance on non-striped appearance was generally inherited and regulated by single locus. Overall, the analyzed datasets of FRC, FFC, and FRS

suggested a largely but simply genetic inheritance pattern in the developed $F_{2,3}$ mapping population.

Analysis of seed phenotypes

SWt of both parent lines and F_1 were different and mean values were noticed as 4.20 ± 0.22 , 5.40 ± 0.26 , and 5.15 ± 0.23 , based on their different genetic inheritance, respectively. In the developed $F_{2,3}$ mapping population, the SWt mean was 4.62 ± 1.02 and ranged from 1.90 to 8.60 gm, by depicting a uniform quantitative distribution and transgressive segregation (Figure 7A). Overall, SWt genetic inheritance in mapping populations revealed close kinship with the wild type male parent line (P_2). STh exhibited different mean values (1.89 ± 0.40 , 2.73 ± 0.95 , and 2.19 ± 0.16) in P_1 , P_2 , and F_1 , respectively. The STh mean value of $F_{2,3}$ mapping population was 2.92 ± 0.33 , and the ranged from 2.32 to 3.95 mm, indicating a uniform quantitative distribution and transgressive segregation (Figure 7B). Overall, genetic inheritance of STh in mapping populations showed an intimate relationship with the wild type male parent line (P_2). SCT of both parent lines and F_1 offspring had also distinct mean values (0.42 ± 0.03 , 0.63 ± 0.02 , and 0.53 ± 0.02), respectively. The SCT mean value of $F_{2,3}$ mapping population was 0.61 ± 0.08 , with a range of minimum 0.30 to maximum 0.95 mm. A uniform quantitative distribution and transgressive segregation was noticed (Figure 7C), disclosing the genetic effects of equal characteristics of parents and F_1 offspring.

SL means of both parent lines were 9.64 ± 0.30 and 10.50 ± 0.75 , and F_1 had 8.65 ± 0.22 . The overall SL mean value of $F_{2,3}$ mapping population was 10.44 ± 1.21 , a uniform distribution was observed, and variations ranged from 7.00 to 13.70 mm. SW means were also different as 18.05 ± 0.51 , 21.23 ± 0.75 , 23.51 ± 0.50 , respectively. However, the SW mean value was 6.51 ± 0.82 and uniform frequency distribution was ranged 4.50 to 9.30 mm, that explained the transgressive segregation in $F_{2,3}$ population (Figures 7D, E). It was found that obvious SL and SW characteristics of both parent lines were mutually shifted characteristics in developed F_1 and $F_{2,3}$ mapping populations. SSI (SL/SW) of P_1 , P_2 , and F_1 was manually deliberated and their mean values were 1.67 ± 0.05 , 1.57 ± 0.03 , 1.62 ± 0.46 ; however, the $F_{2,3}$ mapping population mean value was 1.61 ± 0.13 , showing the range from 1.25 to 2.02 (Figure 7F), respectively.

In addition, the visualized biplot of principal component analysis (PCA) of multivariate phenotypic datasets explained a total of 43.10% shared major variability patterns and striking associations (Figure 8). The first dimension of principal component (Dim-1) extensively summarized the 26.90% explained variances among the seventeen phenotypic traits, and the second dimension of principal component (Dim-2) partially explained the 16.20% of three differentiated phenotypic traits. Overall, positive and linearly connected variables of ovary, fruit, and seed related phenotypes were seen at obtuse & acute angles, respectively.

Analysis of QTLs/genes

A total of 33 QTLs (eight ovary QTLs, sixteen fruit QTLs, and nine seed QTLs) were classified that were randomly pinpointed on different genetic position among the whole-genome chromosomes (Figure 3); however, a QTL cluster was identified on Chr-07 and none of the QTL was observed on Chr-03 and Chr-09. Among the detected QTLs, a total of twenty four QTLs were identified as major-effect QTLs and nine QTLs were identified as minor-effect QTLs that explained the different LOD score values, PVE% (Table 2; Supplementary Figure S1), and significant SNP allelic effects for specific contributions (Supplementary Figure S2). The predicted genes and their detailed GO terms and KEGG pathway enrichment information are given in Supplementary Tables S6–S8, respectively.

QTLs of ovary phenotypes

For the OWt, one major QTL (OWt-7.1) was detected at the bottom-end position on Chr-07. This QTL was also found to be closely related to the QTLs for fruit length and ovary length (FL-7.1 and OL-7.1), indicating a strong relationship. QTL of OWt-7.1 justified the individual genetic effect for ovary weight, with a total of 18.13% PVE, LOD score of 5.06, additive effect of 0.21, and dominance effect of -0.05 . The genetic position of OWt-7.1 was spotted at 203 cM between the confidence interval of CAPS markers (W7-37Ec~W7-39Bs) situated at 157.36 cM and 203.23 cM and exhibited a genetic interval of 45.87 cM. However, the adjacent physical positions (30359068~31916975 bp) of markers exhibited a total of 1.56 Mb of interval that depicted a total of 168 putative genes.

For the OL, a total of three QTLs (one major QTL “OL-6.1” and two minor QTLs “OL-7.1, OL-11.1”) were differentially spotted at genetic positions of three distinct chromosomes (Chr-06, Chr-07, and Chr-11). On Chr-06, a major QTL (OL-6.1) was identified along with the fruit trait QTLs (FWt-6.1 and FL-6.1) and justified the individual genetic effect for ovary length, with a total of 25.99% PVE, LOD score of 4.22, negative additive effect of -0.03 , and dominance effect of -3.32 . The genetic position of OL-6.1 was spotted at 37 cM between the identified CAPS markers (W6-2Ba~W6-3Hd) positioned at 0.00 cM and 56.78 cM; however, the physical positions (750440~1473640 bp) of detected adjacent markers disclosed a total of 723.20 kb interval that depicted a total of 71 putative genes. On Chr-07, a minor QTL (OL-7.1) was tightly located with other QTLs (FWt-7.1, FL-7.1, OWt-7.1), and justified the individual trait effect with 8.91% PVE, LOD score of 3.60, positive additive effect of 1.48, and dominance effect of -0.10 . The genetic position of OL-7.1 was spotted at 203 cM between the flanking markers (W7-37Ec~W7-39Bs) situated at 157.36 cM and 203.23 cM, which exhibited the genetic interval of 45.87 cM. However, the physical position (30359068~31916975 bp) of markers exhibited a total of 1.56

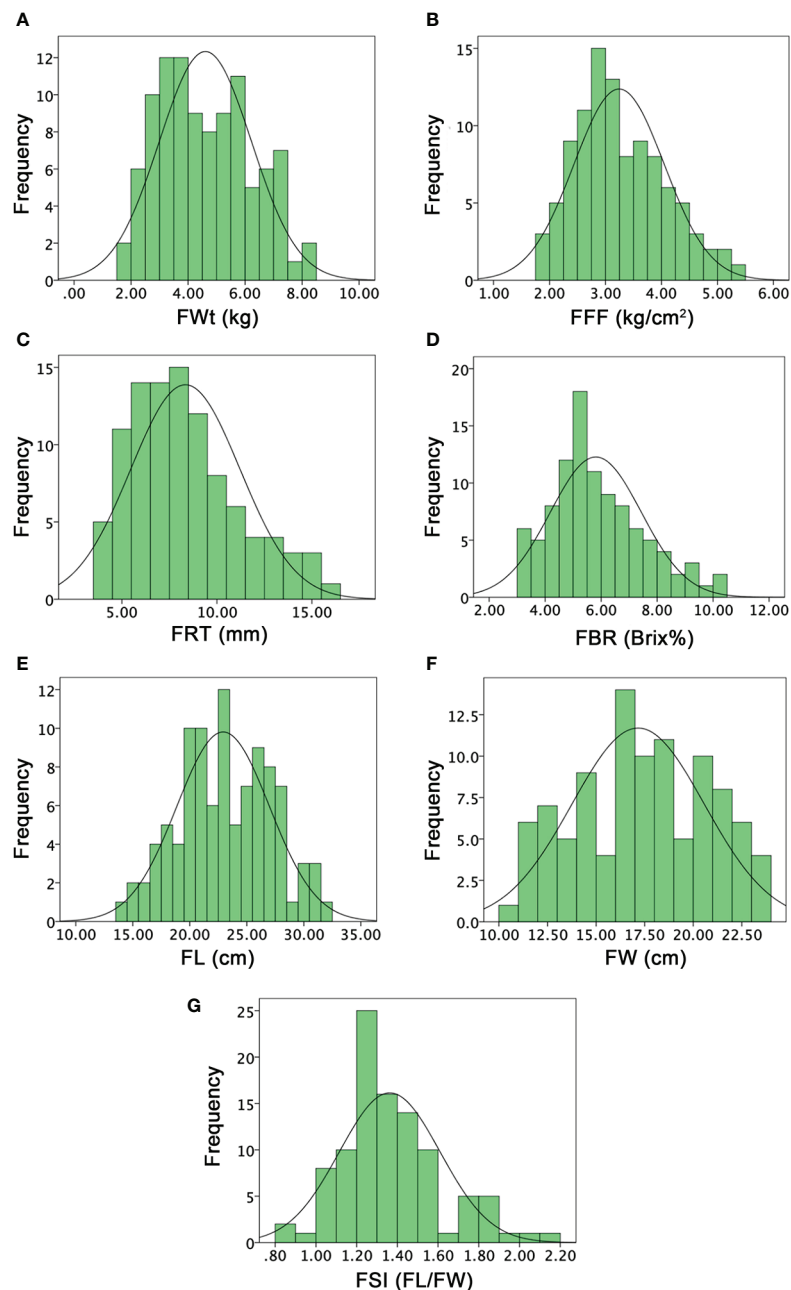


FIGURE 6

Histograms of the frequency distribution of fruit-related phenotypes in a developed F2:3 mapping population. (A) Fruit weight. (B) Fruit flesh firmness. (C) Fruit rind thickness. (D) Fruit Brix. (E) Fruit length. (F) Fruit width. (G) Fruit shape index.

Mb interval that depicted a total of 168 putative genes. On Chr-11, a minor QTL (OL-11.1) was identified along with a single QTL of ovary width (OW-11.1), that explained the individual trait effect with 5.94% PVE, LOD score of 2.85, negative additive effect of -0.46 , and dominance effect of -1.49 . The low PVE% might be due to the quantitative nature of measured traits. The genetic position of OL-11.1 was spotted at the start of chromosomal segment at 203 cM between the flanking

markers (W11-2Bs~W11-3Ec) situated at 0.00 cM and 12.24 cM, which spanned the moderate genetic interval of 12.24 cM. However, the physical position (793295~1543619 bp) of markers exhibited a total of 750.33 kb of interval that depicted a total of 73 putative genes.

For the OW, two major QTLs (OW-7.1 and OW-11.1) were identified on Chr-07 and Chr-11, and collectively explained 27.76% of the phenotypic variations. On Chr-07, OW-7.1

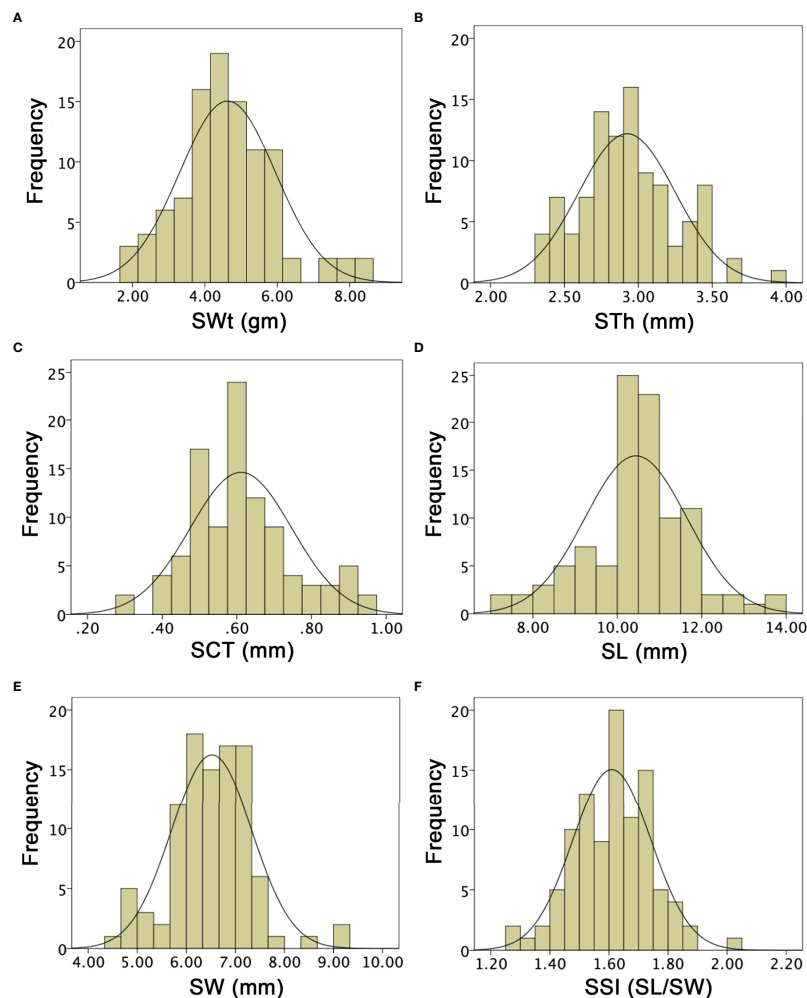


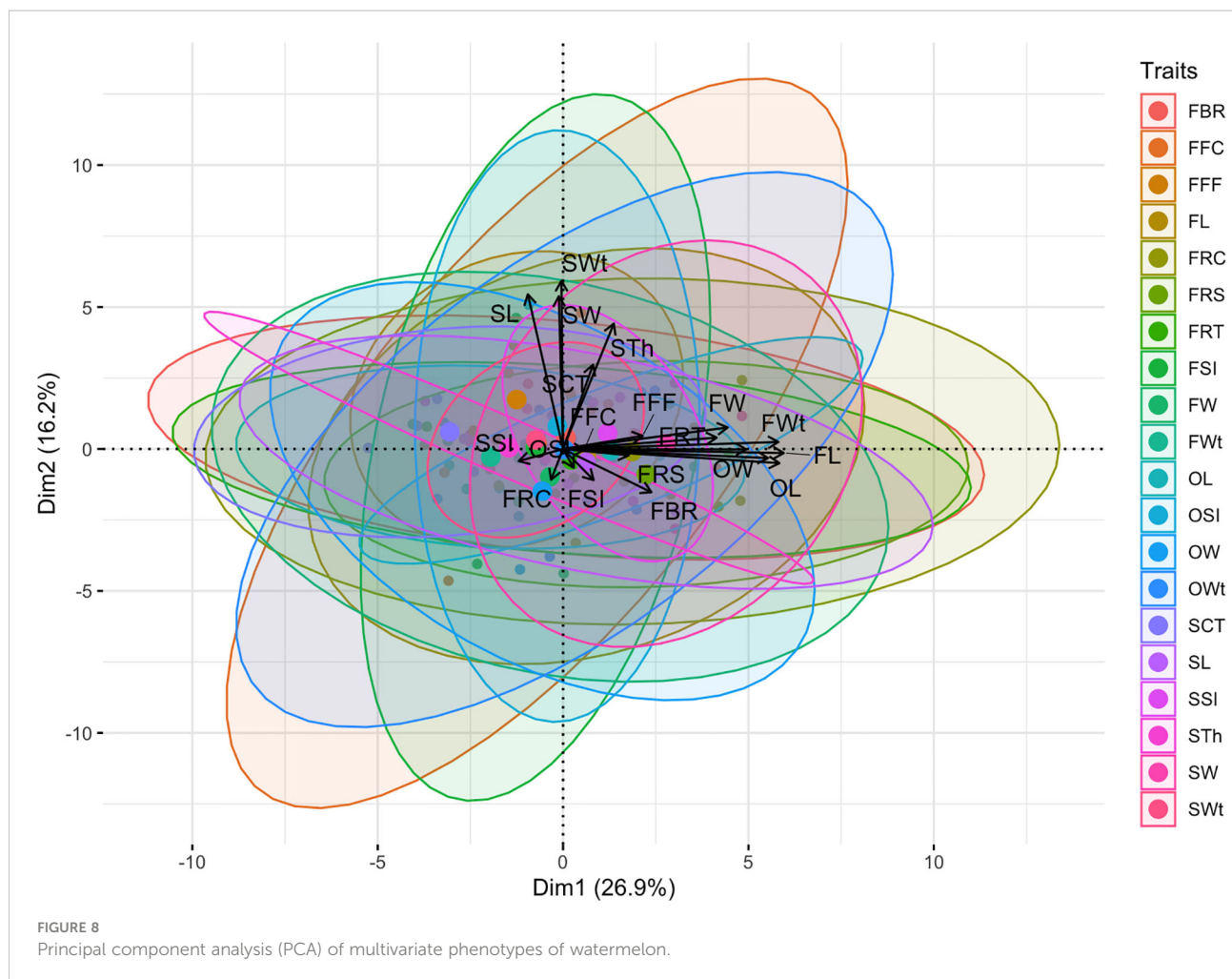
FIGURE 7

Histograms of the frequency distribution of seed-related phenotypes in a developed F2:3 mapping population. (A) Seed weight. (B) Seed thickness. (C) Seed coat thickness. (D) Seed length. (E) Seed width. (F) Seed shape index.

explained total 14.06% PVE, with LOD score of 3.52, negative additive effect of -0.31 , and dominance effect of 1.69 . The genetic position of OW-7.1 was marked at 101 cM between the flanking markers (W7-31Ms~W7-32Bs) situated at 90.47 cM and 108.70 cM, and spanned 18.23 cM interval. However, the physical position ($25579724\sim26363747$ bp) of pointed markers exhibited a total of 784.02 kb of interval that depicted a total of 61 putative genes. On Chr-11, OW-11.1 explained an individual effect of 13.70% PVE, with LOD score of 3.67 , negative additive effect of -0.01 , and dominance effect of -1.65 . The genetic position of OW-11.1 was marked at 10 cM between the flanking markers (W11-2Bs~W11-3Ec) situated at 0.00 cM and 12.24 cM, which spanned a genetic interval of 12.24 cM. However, the physical position ($793295\sim1543619$ bp) of flanking markers exhibited a total of 750.33 kb of interval that depicted a total of 73 putative genes.

For the OSI, two major QTLs (OSI-1.1 and OSI-2.1) were detected on Chr-01 and Chr-02, and showed 25.32% phenotypic

variation for shape indexes. On Chr-01, OSI-1.1 explained individual genetic effect of 10.61% PVE, with LOD score of 2.88 , positive additive effect of 0.06 , and dominance effect of 0.15 . The genetic position of this QTL was situated at 138 cM between the flanking markers (W1-36Ba~W1-37Ms) situated at 134.22 cM and 150.63 cM, and spanned total 16.41 cM. However, the physical positions ($32266077\sim33187192$ bp) exhibited 921.12 kb interval that depicted a total of 134 putative genes. On Chr-02, another OSI-2.1 explained individual genetic effects of 14.17% PVE, with a LOD score of 3.86 , negative additive effect of -0.12 , and dominance effect of -0.11 . The genetic position of this QTL was located at 67 cM between the flanking markers (W2-7Ec~W2-8Bs) situated at 37.50 cM and 69.72 cM, spanning 32.22 cM; however, the physical positions ($4733877\sim5684571$ bp) exhibited a 950.70 kb of interval that depicted a total of 73 putative genes. Regarding the overall inherited quantitative genetic of ovary related traits, the detected positive and negative additive effects



of mapped QTLs exhibited multiple ovary characteristics (OWt, OL, OW, OSI) in a developed $F_{2,3}$ mapping population and mainly signified the mutual heredity of both parent lines.

QTLs of fruit phenotypes

For the FWt, two major QTLs (FWt-6.1 and FWt-7.1) were mapped on different chromosomes “Chr-06 and Chr-07” and collectively explained 31.28% of the phenotypic variance for fruit weight morphology. On Chr-06, FWt-6.1 was positioned at the start of the chromosomal segment along with fruit length and ovary length QTLs (FL-6.1 and OL-6.1), which explained phenotypic variance with 15.61% PVE, with LOD value of 3.87, negative additive effect of -0.05 , and dominance effect of 2.67. The genetic position of this QTL was situated at 29 cM between the flanking markers (W6-2Ba~W6-3Hd) positioned at 0.00 cM and 56.78 cM, spanning a wide genetic distance of about 56.78 cM. However, the physical positions (750440~1473640 bp) exhibited a 723.20 kb interval that depicted a total of 71 putative

genes. On Chr-07, FWt-7.1 was positioned at the bottom end of the chromosomal segment, along with the QTLs of ovary weight, ovary length, and fruit length (OWt-7.1, OL-7.1, FL-7.1), respectively. This QTL “FWt-7.1” explained phenotypic variance with 15.67% PVE, with LOD value of 3.04, negative additive effect of -0.16 , and dominance effect of 2.65. The genetic position of this QTL was situated at 179 cM between the flanking markers (W7-37Ec~W7-39Bs) positioned at 157.36 cM and 203.23 cM, spanning a wide genetic distance about 45.87 cM, but adjacent physical positions (30359068~31916975 bp) exhibited a 1.56 Mb interval that showed a total 168 putative genes.

For the FL, one major QTL (FL-6.1) and one minor QTL (FL-7.1) were detected on genetic positions of different chromosomes (Chr-06 and Chr-07), and collectively explained 25.80% of the phenotypic variations for fruit length characteristics. On Chr-06, QTL “FL-6.1” was detected with LOD score of 3.16, negative additive effect of -2.08 , dominance effect of 50.24, and explained individual phenotypic effect with 15.67% PVE. The genetic position of this QTL was situated at 27

TABLE 2 Genetic effects of mapped QTLs affecting multivariate phenotypes of watermelon.

| QTLs | Chr. | Position (cM) | Adjacent markers | Position (bp) | LOD score | PVE (%) | Add effect | Dom effect |
|----------|------|---------------|-------------------|-------------------|-----------|---------|------------|------------|
| OWt-7.1 | 07 | 203 | W7-37Ec~W7-39Bs | 30359068~31916975 | 5.06 | 18.13 | 0.21 | -0.05 |
| OL-6.1 | 06 | 37 | W6-2Ba~W6-3Hd | 750440~1473640 | 4.22 | 25.99 | -0.03 | 3.32 |
| OL-7.1 | 07 | 203 | W7-37Ec~W7-39Bs | 30359068~31916975 | 3.60 | 8.91 | 1.48 | -0.10 |
| OL-11.1 | 11 | 0 | W11-2Bs~W11-3Ec | 793295~1543619 | 2.85 | 5.94 | -0.46 | -1.49 |
| OW-7.1 | 07 | 101 | W7-31Ms~W7-32Bs | 25579724~26363747 | 3.52 | 14.06 | -0.31 | 1.69 |
| OW-11.1 | 11 | 10 | W11-2Bs~W11-3Ec | 793295~1543619 | 3.67 | 13.70 | -0.01 | -1.65 |
| OSI-1.1 | 01 | 138 | W1-36Ba~W1-37Ms | 32266077~33187192 | 2.88 | 10.61 | 0.06 | 0.15 |
| OSI-2.1 | 02 | 67 | W2-7Ec~W2-8Bs | 4733877~5684571 | 3.86 | 14.17 | -0.12 | -0.11 |
| FRS-4.1 | 04 | 03 | W4-4Hd~W4-5Hd | 2031314~2731617 | 2.51 | 9.75 | 0.28 | 0.34 |
| FRC-2.1 | 02 | 219 | W2-40Ms~W2-43Ms | 32909754~34162281 | 3.13 | 15.12 | -0.28 | -0.38 |
| FWt-6.1 | 06 | 29 | W6-2Ba~W6-3Hd | 750440~1473640 | 3.87 | 15.61 | -0.05 | 2.67 |
| FWt-7.1 | 07 | 179 | W7-37Ec~W7-39Bs | 30359068~31916975 | 3.04 | 15.67 | -0.16 | 2.65 |
| FL-6.1 | 06 | 27 | W6-2Ba~W6-3Hd | 750440~1473640 | 3.16 | 20.48 | -2.08 | 50.24 |
| FL-7.1 | 07 | 203 | W7-37Ec~W7-39Bs | 30359068~31916975 | 2.60 | 5.32 | 19.34 | 1.95 |
| FW-7.1 | 07 | 46 | W7-18Ps~W7-19Ms | 13585253~14389028 | 3.29 | 10.93 | -10.96 | 20.93 |
| FW-7.2 | 07 | 81 | W7-26Ps~W7-27Hd | 20772533~21571067 | 2.60 | 7.13 | -9.68 | 15.46 |
| FW-11.1 | 11 | 57 | W11-7Ps~W11-8Bs | 4623563~5404406 | 4.61 | 14.22 | -22.61 | -10.43 |
| FSI-2.1 | 02 | 53 | W2-7Ec~W2-8Bs | 4733877~5684571 | 3.76 | 16.95 | -0.15 | -0.20 |
| FSI-7.1 | 07 | 68 | W7-21Ms~W7-24Ms | 16745635~19169162 | 2.67 | 5.82 | 0.06 | -0.13 |
| FSI-7.2 | 07 | 79 | W7-26Ps~W7-27Hd | 20772533~21571067 | 3.04 | 6.42 | 0.06 | -0.14 |
| FFC-5.1 | 05 | 193 | W5-33Ms~W5-37Hd | 28704441~32310869 | 2.60 | 7.42 | 0.50 | 0.67 |
| FFF-11.1 | 11 | 18 | W114-Ps~W11-5Ms | 2340516~3081651 | 2.57 | 9.01 | -0.33 | -0.13 |
| FRT-4.1 | 04 | 78 | W4-17Ec~W4-19Ms | 10841468~12199409 | 3.02 | 14.48 | -1.40 | -0.69 |
| FBR-10.1 | 10 | 91 | W10-26Ms~W10-27Ba | 21923221~22813961 | 2.94 | 13.98 | 0.08 | 1.22 |
| SWt-4.1 | 04 | 7 | W4-5Hd~W4-7Ms | 2731617~4061374 | 2.91 | 13.85 | -0.58 | -2.00 |
| SWt-5.1 | 05 | 15 | W5-2Hd~W5-3Ps | 886916~1795952 | 2.61 | 10.72 | 1.23 | -0.02 |
| SWt-10.1 | 10 | 69 | W10-12Ec~W10-14Hd | 9647097~11399426 | 2.89 | 11.54 | -1.52 | 0.34 |
| STh-7.1 | 07 | 14 | W7-4Ba~W7-6Bs | 3978006~4794459 | 2.51 | 11.87 | -0.18 | 0.09 |
| SL-7.1 | 07 | 7 | W7-6Bs~W7-7Ec | 3978006~4794459 | 6.77 | 23.82 | -0.86 | 0.34 |
| SW-4.1 | 04 | 13 | W4-7Ms~W4-8Ms | 4061374~4741494 | 2.79 | 13.78 | -0.18 | -0.59 |
| SW-10.1 | 10 | 64 | W10-11Ms~W10-12Ec | 8770056~9647097 | 2.73 | 12.28 | -0.52 | 0.30 |
| SSI-6.1 | 06 | 83 | W6-7Ms~W6-9Hd | 4421727~5895809 | 2.76 | 13.41 | 0.07 | -0.07 |
| SCT-1.1 | 01 | 13 | W1-4Ec~W1-6Ms | 2773065~4611685 | 3.54 | 16.80 | -0.07 | 0.07 |

cM between the flanking markers (W6-2Ba~W6-3Hd) positioned at 0.00 cM and 56.78 cM, spanning a wide genetic distance (cM). The adjacent physical positions (750440~1473640 bp) showed a 723.20 kb interval that depicted a total of 71 putative genes. On Chr-07, QTL “FL-

7.1” was positioned at the end of corresponded chromosomal region and detected with LOD score of 2.60, positive additive effect of 19.34, dominance effect of 1.95, and explained individual trait effect with 15.67% PVE. The genetic position of this QTL was situated at 203 cM between the flanking markers

(W7-37Ec~W7-39Bs) positioned at 157.36 cM and 203.23 cM, spanning a wide genetic distance of about 45.87 cM, but adjacent physical positions (30359068~31916975 bp) exhibited a 1.56 Mb interval that contained a total of 168 putative genes.

For the FW, a total of three QTLs (one minor QTL “FW-7.2” and two major QTLs “FW-7.1 and FW-11.1”) were detected on different genetic positions of Chr-07 and Chr-11, and collectively explained 32.28% of the phenotypic variances for fruit width descriptions. On Chr-07, a major QTL (FW-7.1) was found with LOD score of 3.29, negative additive effect of -10.96, dominance effect of 20.93, and a 10.93% PVE, individually. The genetic position of this QTL was located at 46 cM between the flanking markers (W7-18Ps~W7-19Ms) positioned at 40.15 cM and 61.81 cM, spanning total 21.66 cM genomic interval, but adjacent physical positions (13585253~14389028 bp) showed 803.78 kb interval, displaying a total 7 putative genes. Further, a minor QTL (FW-7.2) was noticed with LOD score of 2.60, negative additive effect of -9.68, dominance effect of 15.46, and a 7.13% PVE, individually. The genetic position of this QTL was located at 81 cM between the flanking markers (W7-26Ps~W7-27Hd) positioned at 78.37 cM and 82.79 cM, spanning delimited genomic interval of 4.42 cM, but adjacent physical positions (20772533~21571067 bp) showed a 798.54 kb interval, showing a total of 35 putative genes. On Chr-11, a single major QTL (FW-11.1) was identified with LOD score of 4.61, negative additive effect of -22.61, and dominance effect of -10.43 and individually depicted 14.22% PVE. The genetic position of this QTL was situated at 57 cM between the flanking markers (W11-7Ps~W11-8Bs) positioned at 56.47 cM and 60.34 cM, spanning a shortened genomic interval (3.87 cM), but adjacent physical positions (4623563~5404406 bp) exhibited a 780.85 kb interval, disclosing a total 88 putative genes.

For the FSI, a total of three QTLs “one major QTL (FSI-2.1) and two minor QTLs (FSI-7.1 and FSI-7.2)” were localized on the genetic positions of two differential chromosomes (Chr-02 and Chr-07). On Chr-02, FSI-2.1 QTL was positioned along with OSI-2.1 and exhibited a good genetic connection between ovary and fruit shape. This QTL explained individual genetic effects of 16.95% PVE, with LOD score of 3.76, negative additive effect of -0.15, and dominance effect of -0.20. The genetic position of FSI-2.1 was found at 53 cM between the flanking markers (W2-7Ec~W2-8Bs) situated at 37.50 cM and 69.72 cM, that spanned 32.22 cM; however, the physical position (4733877~5684571 bp) exhibited 950.70 kb interval that depicted a total 73 putative genes. On Chr-07, FSI-7.1 QTL was located separately but near to the genetic position of FW-7.1 and explained the individual effect with 5.82% PVE, positive additive effect of 0.06, and negative dominance effect of -0.06. The genetic position of FSI-7.1 was found at 68 cM between the flanking markers (W7-21Ms~W7-24Ms) situated at 64.53 cM and 68.36 cM, spanning a shortened genetic interval of 3.83 cM; however, the physical position (16745635~19169162 bp) exhibited a 2.20 Mb interval, enclosing a total of 21 putative genes. Another QTL

(FSI-7.2) seemed as tightly localized as the FW-7.2 QTL, perhaps signifying that fruit shape was mainly determined by fruit width. The genetic position of FSI-7.2 was situated at 79 cM between the flanking markers (W7-26Ps~W7-27Hd) positioned at 78.37 cM and 82.79 cM, and spanned a shortened genetic interval of 4.42 cM. The physical position (20772533~21571067 bp) exhibited a 798.54 kb interval, revealing a total of 35 putative genes.

For the FRT, one major QTL (FRT-4.1) was mapped at Chr-04 and explained an individual genetic effect with 14.48% PVE, LOD score of 3.02, negative additive and dominance effects (-1.40 and -0.69). The genetic position of this QTL was situated in the middle genetic section of Chr-04, at 78 cM between the flanking markers (W4-17Ec~W4-19Ms) positioned at 77.47 cM and 94.46 cM, spanning a total genetic interval of 16.69 cM. However, the physical position (10841468~12199409 bp) exhibited a 1.36 Mb interval that showed a total of 18 genes. For the FFF trait, one minor QTL (FFF-11.1) was mapped on Chr-11 and this QTL explained an individual genetic effect with 9.01% PVE, LOD score of 2.57, negative additive and dominance effect (-0.33 and -0.13). The genetic position of this QTL was situated at 18 cM between the flanking markers (W11-4Ps~W11-5Ms) positioned at 44.66 cM and 50.86 cM, covering a minimum of 6.20 cM, and the physical position (2340516~3081651 bp) exhibited total 741.14 kb interval, depicting 78 putative genes. For the FBR trait, one major QTL (FBR-10.1) was mapped at Chr-10 and explained individual phenotypic effects with 13.98% PVE, LOD score of 2.94, and positive additive and dominance effects of 0.08 and 1.22. The genetic position of this QTL was situated at 91 cM between the flanking markers (W10-26Ms~W10-27Ba) positioned at 90.60 cM and 103.53 cM, spanning a genetic interval of 12.93 cM. The physical position (21923221~22813961 bp) exhibited an 890.74 kb interval that showed a total of 43 genes.

For the FFC, one minor QTL (FFC-5.1) was mapped to a wide chromosomal region of Chr-02 and explained individual genetic effects with 7.42% PVE, LOD score of 2.60, positive additive and dominance effects (0.50 and 0.67). The genetic position of this QTL was situated at 193 cM between the flanking markers (W5-33Ms~W5-37Hd) positioned at 161.47 cM and 213.88 cM, spanning a broad range of genetic interval (52.41 cM); however, the physical positions (28704441~32310869 bp) exhibited a 3.61 Mb interval that depicted a total of 475 genes. For the FRC trait, one major QTL (FRC-2.1) was mapped at the bottom end of Chr-02. This QTL explained individual genetic effects of 15.12% PVE, with LOD score of 3.13, negative additive effect of -0.28, and dominance effect of -0.38. The genetic position of this QTL was situated at 219 cM between the flanking markers (W2-40Ms~W2-43Ms) positioned at 218.49 cM and 226.44 cM, spanning just 9.80 cM. However, the physical position (32909754~34162281 bp) exhibited a 1.25 Mb interval that depicted a total of 157 putative genes. For the FRS trait, one minor QTL (FRS-4.1) was identified at the start position of Chr-

04. This QTL explained individual genetic effects of 9.75% PVE, with LOD score of 2.51, positive additive effect of 0.28, and dominance effect of 0.34. The genetic position of this QTL was situated at 3 cM between the flanking markers (W4-4Hd~W4-5Hd) positioned at 0.00 cM and 3.29 cM, spanning a total of 3.29 cM. However, the physical positions (2031314~2731617 bp) exhibited a 700.30 kb interval and depicted a total of 16 putative genes.

QTLs of seed phenotypes

For the SWt, a total of three major QTLs (SWt-4.1, SWt-5.1, and SWt-10.1) were detected on three different chromosomes (Chr-04, Chr-05, and Chr-10). Interestingly, these QTLs shared a strong relationship with seed width and seed length QTLs (SW-4.1 and SW-10.1), indicating that seed weight was primarily inherited by seed width characteristics. On Chr-04, SWt-4.1 QTL was mapped with LOD score of 2.91, negative additive and dominance effects (−0.58 and −2.00), and mainly explained the phenotypic variation for reduced seed weight with 13.85% PVE. The genetic position of SWt-4.1 QTL was situated at 7 cM between the flanking markers (W4-5Hd~W4-7Ms) placed at 3.29 cM and 9.45 cM, covering 6.16 cM; however, the physical position (4061374~4741494 bp) unveiled a total 680.12 kb interval that depicted 32 putative genes. On Chr-05, SWt-5.1 QTL was pinpointed with LOD score of 2.61, positive additive effect of 1.23, and a negative dominance effect of −0.02, and mainly explained the phenotypic variation for more seed weight with 10.72% PVE. The genetic position of SWt-5.1 QTL was situated at 15 cM between the flanking markers (W5-2Hd~W5-3Ps) sited at 0.00 cM and 18.57 cM; however, the physical position (886916~1795952 bp) unveiled a total 909.04 kb interval predicted for 135 genes. On Chr-10, SWt-10.1 QTL was pinpointed with LOD score of 2.89, negative additive effect of −1.52 and a positive dominance effect of 0.34, and explained the phenotypic variation for less seed weight with 11.54% PVE. The genetic position of SWt-10.1 QTL was situated at 69 cM between the flanking markers (W10-12Ec~W10-14Hd) situated at 67.92 cM and 76.70 cM, and the physical positions (9647097~11399426 bp) revealed a 1.75 Mb interval having 51 genes.

For the STh and SL, two major QTLs (STh-7.1 and SL-7.1) were pinpointed on Chr-07. They shared a common genetic location by explaining the combined characteristics of reduced seed thickness as well as seed length and seemed to determine the genetic inheritance by somewhat moderate-sized seeds having less seed thickness, respectively. For the STh, one major QTL (STh-7.1) was pinpointed with LOD score of 2.51, negative additive effect of −0.18, positive additive effect of 0.09, and 11.87% PVE. For the SL, a major QTL (SL-7.1) was detected with LOD score of 6.77, negative additive effect of −0.86, positive dominance effect of 0.34, and 23.82% PVE, separately. The

genetic positions of both QTLs were found at 14 cM and 7 cM between the adjacent flanking markers (W7-6Bs~W7-7Ec) situated at 0.00 cM and 26.03 cM, and their physical positions (3978006~4794459 bp) exhibited an interval of 1.59 Mb, which contained 46 genes.

For the SW, two major QTLs (SW-4.1 and SW-10.1) were found on two different chromosomes (Chr-04 and Chr-10). These two QTLs were found to be close to the seed weight QTLs on those chromosomes, showing that seed weight and seed width are closely related. On Chr-04, a major QTL (SW-4.1) justified the individual genetic effect for seed weight, with a total of 13.78% PVE, LOD score of 2.79, negative additive and dominance effects of −0.18 and −0.59. The genetic position of SW-4.1 was spotted at 13 cM between the flanking sections of two CAPS markers (W4-7Ms~W4-8Ms) situated at 9.45 cM and 18.02 cM, and exhibited the genetic interval of 8.57 cM. But the physical positions of markers next to each other (4061374~4741494 bp) only showed 8 putative genes over a total of 680.12 kb. On Chr-10, another major QTL (SW-10.1) explained the genetic effects with 12.28% PVE, LOD score of 2.73, negative additive effect of −0.52, and dominance effect of 0.30. The genetic location of SW-10.1 was found to be 64 cM between the CAPS markers W10-11Ms and W10-12Ec, which were at 59.43 cM and 67.92 cM, with a genetic gap of 8.49 cM. However, the adjacent physical positions (8770056~9647097 bp) of markers exhibited a total of 877.04 kb of interval that depicted just 36 putative genes.

For the SSI, one major QTL (SSI-6.1) was mapped over Chr-06 and explained individual genetic effects with 13.41% PVE, with LOD score of 2.76, positive additive effect of 0.07, and negative dominance effect of −0.07. The genetic position of this QTL was situated at 83 cM between the flanking markers (W6-7Ms~W6-9Hd) positioned at 79.11 cM and 90.82 cM, spanning about 11.71 cM; however, the adjacent physical positions (4421727~5895809 bp) exhibited a total 1.47 Mb interval and showed 146 putative genes. For the SCT trait, one major QTL (SCT-1.1) was separately mapped to the chromosomal region of Chr-01 and explained individual genetic effects with 16.80% PVE, with LOD score of 3.54, negative additive effect of −0.07, and positive dominance effect of 0.07. The genetic position of this QTL was situated at 13 cM between the flanking markers (W1-4Ec~W1-6Ms) situated at 4.39 cM and 17.21 cM, covering 12.82 cM; however, the adjacent physical positions (2773065~4611685 bp) exhibited total 1.84 Mb interval and displayed a total of 187 putative genes.

GO and KEGG enrichment analysis of predicted genes

In total, 580 genes for four ovary traits, 1172 genes for ten fruit phenotypes, and 641 genes for six seed traits were identified among the identified flanking QTL regions ([Supplementary](#)

Table S6), and SNP allelic effects underlying identified QTLs also shown the significant allele specific contributions (Supplementary Figure S2), respectively.

According to the pair-wise sequence analysis of parent line sequences and reference genome assembly, most of the genes showed candidate mutations within the CDS coding regions. Moreover, the genes were subsequently predicted and analyzed to check their categorized synteny of potential mechanisms mediating the ovary, fruit, and seed traits of comparative experimental material.

A schematic representation of the VENN diagram significantly illustrated the categorized genes (Figure 9A), and a total of 312 genes were observed to show the decisive functions for regulating the dynamic traits of ovary and fruit, and only 8 genes exhibited a connection between fruit and seed traits, respectively. Gene Ontology function enrichment analysis exhibited the functional distribution of ovary, fruit, and seed traits linked genes (Supplementary Table S7). According to the bio-informatic analysis, GO functional enrichment was mainly categorized into molecular function (MF), biological process (BP), and cellular components (CC), and GO terms with a mean P -value of <0.05 were recognized as considerably enriched (Figure 9B).

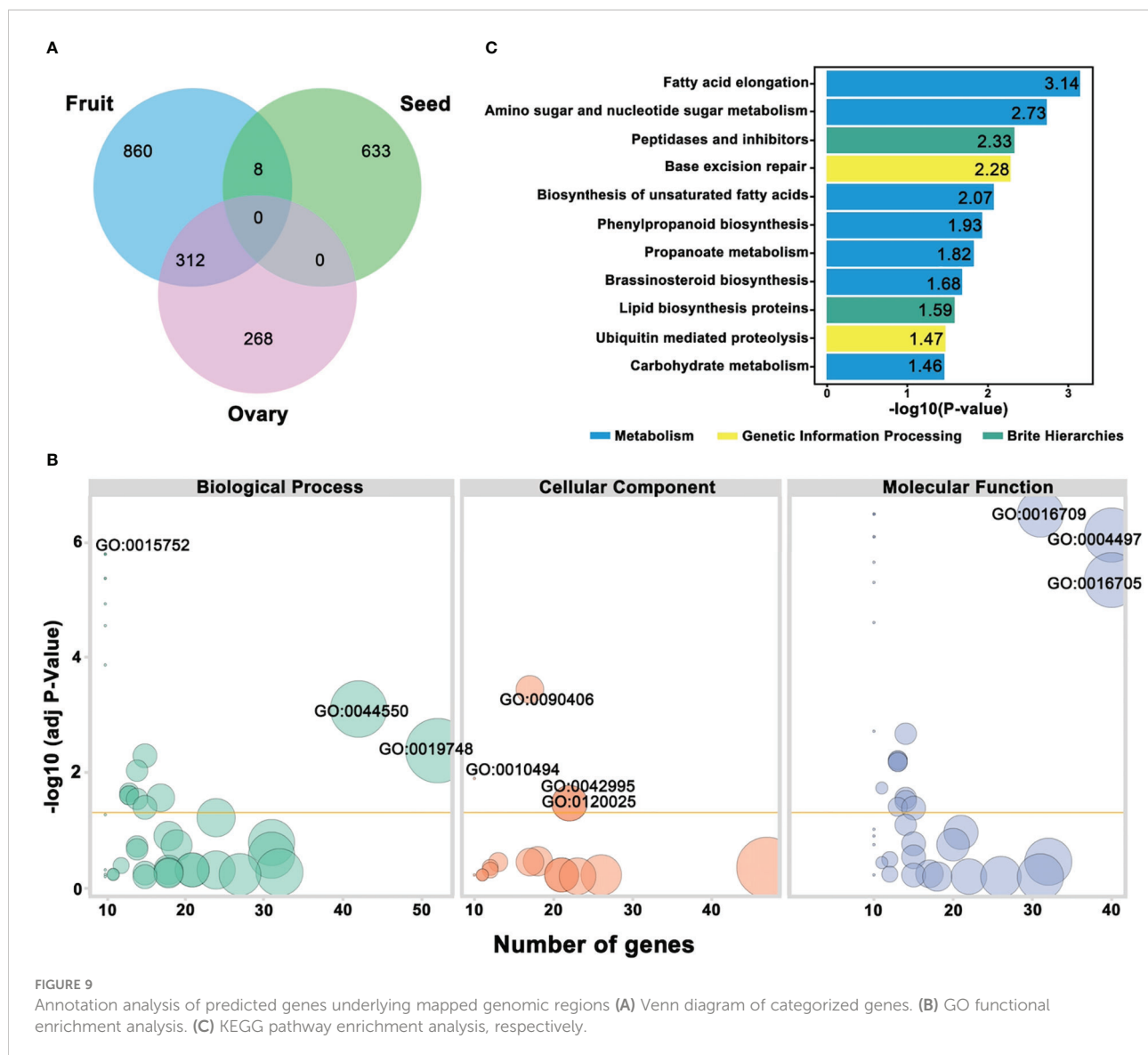
Regarding the GO biological process, the identified genes were divided into 19 significant GO terms with different values of $-\log_{10}(P\text{-value})$; among them, GO:0015752 (D-ribose transmembrane transport), GO:0044550 (secondary metabolite biosynthetic process), and GO:0019748 (secondary metabolic process) were found with highly enriched GO terms. The low numbers of GO terms were cellular response to metal ion, glucose transmembrane transport, hexose transmembrane transport, and glucose import. However, few other genes were found with a moderate number of GO terms. In the GO cellular component, 4 GO terms were exhibited as highly enriched, e.g., GO:0090406 (pollen tube) GO:0010494 (cytoplasmic stress granule), GO:0042995 (cell projection), and GO:0120025 (plasma membrane bounded cell projection), and the remaining were detected with low to moderate enrichment. All the genes were divided into 24 significant GO terms in the GO molecular function. Three of these terms were significantly enriched; GO:0016709 (oxidoreductase activity acting on paired donors with incorporation or reduction of molecular oxygen of NAD(P)H as one donor, and incorporation of one atom of oxygen), GO:0004497 (D-xylose transmembrane transporter activity), and GO:0016705 (oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen). The GO terms with the lowest significant enrichment were hexose transmembrane transporter activity, glucose transmembrane transporter activity, phosphatidic acid binding, carbohydrate: proton symporter activity, carbohydrate: cation symporter activity, oxidoreductase activity, hydro-lyase

activity, poly(A) binding, solute: cation symporter activity, and symporter activity; however, the remaining GO terms were noticed with moderate enrichment.

To better understand the important gene contributions in the key metabolic and signal transduction pathways regulating the ovary, fruit, and seed phenotypes of comparative parental lines, the identified genes were evaluated in the KEGG database, and the first 11 pathways with 3 main classes were designated at significant $-\log_{10}(P\text{-value})$ and used for visualized plotting (Figure 9C; Supplementary Table S8). The most significantly enriched pathways were observed with fatty acid elongation (Ko00062), amino sugar and nucleotide sugar metabolism (Ko00520), peptidase and inhibitors (Ko01002), lipid biosynthesis proteins (Ko01004), followed by Ko03410 (Base excision repair) and Ubiquitin mediated proteolysis, and the main classes were metabolism, genetic information processing, and brite hierarchies, which might be involved in the differentiation of the horticultural phenotypes of comparative parental lines, respectively.

Discussion

Watermelon is an important fruit in the cucurbitaceae family with a wide range of quantitative and qualitative characteristics. In this experiment, two extremely divergent watermelon parent lines (W1-38 and PI542119) were used for genomic sequencing and a total of 3,784,650 SNPs and 172,151 CAPS loci pairs were detected; however, transition type SNPs were noticed higher than transversion type SNPs. A total of 210 sets of novel SNP-CAPS markers exhibited a moderate level of polymorphism (46.25%), and the obtained results are fairly comparable with the earlier reported genetic mapping studies in watermelon (Liu et al., 2015; Liu et al., 2016; Amanullah et al., 2021; Osae et al., 2021; Amanullah et al., 2022; Osae et al., 2022). Our constructed genetic map had a total length of 2,398.40 cM, an average interval length of 11.42 cM, and most physical intervals ranging from a minimum of 680.12 kb to a maximum of 950.70 kb, indicating lower recombination rate (Figures 3, 4). Furthermore, a few CAPS markers displayed a relatively large genetic interval ranging from 1.5 Mb to 3.3 Mb and appeared to be deviated in the linkage equilibrium across the reference genome. The density of whole-genome markers with fewer genetic distances and the size of mapping populations are major concerns for perfect linkage mapping and QTL analysis (Pereira et al., 2018; Liang et al., 2022). So, we assumed that our developed genetic map still needs to be improved by delimiting the unsuitable genetic intervals by incorporating the high density markers that would provide the additional accuracy for QTLs/genes mapping.



Ovary QTLs

A few QTLs of ovary-related specific traits (weight, width, length, and shape index) have been successfully analyzed in a few Cucurbitaceae fruits, including cucumber (Wei et al., 2016; Wang et al., 2020), melon (Ramamurthy and Waters, 2015; Amanullah et al., 2020; Amanullah et al., 2021), and squash (Kamiska et al., 2018). In watermelon, it was reported that development of ovary shape/size is a pre-anthesis genetic phenomenon developed at the initial development stage of the ovary, which similarly leads to the different obvious fruit shapes induced by polygenic control (Dou et al., 2018b). Due to pre-anthesis genetic inheritance in cultivars with different genetic backgrounds (Legendre et al., 2020), the long ovaries cause the relative absolute shape of the fruit to be long, and the round ovaries cause the fruit to be round.

For the genetic mapping of watermelon ovary traits, a molecular mapping study classified two ovary weight QTLs, two ovary length QTLs, and three ovary width QTLs, which were positioned on Chr-01, Chr-03, Chr-08, and Chr-09 (Osae et al., 2022) and explained about 8.87% to 20.82% PVE, by defining the polygenic architecture and conferring the obvious ovary-fruit shape index. In our study, ovary traits related to major and minor QTLs were mapped between the delimited adjacent regions of genetic markers positioned on Chr-01, Chr-02, Chr-06, Chr-07, and Chr-11, which explained 8.91% to 25.99% PVE, respectively (Figure 3; Table 2). The detected positive and negative additive effects of our mapped QTLs exhibited the variation of multiple ovary characteristics (OWt, OL, OW, and OSI) in the developed $F_{2,3}$ mapping population and mainly signified the mutual genetic heredity of both parent lines. We also discovered that our mapped QTLs regions contradicted

those previously mapped QTLs (Dou et al., 2018b; Legendre et al., 2020; Osae et al., 2022), as shown in Supplementary Table S9, respectively.

To the best of our knowledge, our identified QTL segments indicated the new genetic regions with strong pleiotropic effects for controlling the ovary traits. Furthermore, transgressive segregation was observed for all ovary traits in our study (Figure 5), and co-QTLs on Chr-02, Chr-06, and Chr-11 strongly supported the existence of synteny modulating between genetic positions in the watermelon genome. It is supposed that allelic fashions of both parent lines produced the genetic effects for reduced size and lengthy ovaries variation. Interestingly, the detected QTLs were also fitted with quantitative genetics of ovary associated traits and their mechanisms was seemed to be triggered by numerous genes.

Fruit QTLs

Fruit weight (FWt) is very important for making a good commercial profit (Amanullah et al., 2021). The morphological divergences of different cultivars range from fruit weight in terms of gm to kg (Osae et al., 2022; Liang et al., 2022). In few earlier studies, QTLs of watermelon FWt have been mapped over Chr-09 and Chr-03 by using the developed biparental and RIL mapping families resulting from the crossing of cultivated-type and wild-type parent lines (Fan et al., 2000; Sandlin et al., 2012; Yang et al., 2021; Liang et al., 2022; Osae et al., 2022). In this study, we similarly incorporated biparental $F_{2:3}$ mapping population and mapped just two major QTLs of fruit weight (FWt-6.1 and FWt-7.1) positioned over Chr-06 and Chr-07, which justified 15.61~15.67% PVE for FWt variation (less and more), respectively (Figure 3; Table 2). These QTL results contradict the earlier published studies (Supplementary Table S9) and strongly suggest the genetic divergence in contrast to the parental lines and their derived experimental populations. So, we hypothesized that our new FWt QTLs might signify the new reliable mapped genomic regions for controlling the variation in FWt of watermelon. For the genetic regulation of FWt mechanisms, the *LC* (an important member of the *WOX*, *YABBY*, and *FAS* families) was significantly known for gradual variation in FWt (Huang et al., 2013), and the *FW2.2/CNR* was similarly classified as a major locus of FWt that encodes the protein for relative regulation of cell number regulators (*CNR*) (Wu et al., 2018). Our identified genes for FWt might exhibit the gene expression profiling but strong validation is necessary by further fine mapping study.

Fruit size (length, width, and shape) variations in watermelon have been classified into elongated, round, blocky, or oval shapes (McKay, 1936; Wehner et al., 2001). Many studies have identified QTLs that control the majority of fruit variation in various biparental mapping populations, RILs, and natural populations under various environmental conditions (Lu et al., 2009; Sandlin et al., 2012; Ren

et al., 2014; Liu et al., 2014; Kim et al., 2015; Liu et al., 2015; Liu et al., 2016; Cheng et al., 2016; Lu et al., 2016; Dou et al., 2018b; Maragal et al., 2019; Legendre et al., 2020). The inclusive number of QTLs regulating the watermelon shape/size have been reviewed across all previously published QTL results (Pan et al., 2020), and 9 inclusive QTLs have been reported across 7 distinct chromosomal regions of Chr-02, Chr-03, Chr-05, Chr-07, Chr-08, Chr-09, and Chr-10. The FSI QTL “CIFS-3.1” was discovered on Chr-03, exhibiting stable genetic effects for the regulation of FSI in segregating biparental populations of watermelon. However, four QTLs (CIFS-2.3, CIFS-3.3, CIFS-4.1, and CIFS-8.1) could express stable interactions in more than one experimental mapping population or environmental location of the 15 inclusive FSI QTL (Sandlin et al., 2012). In this study, we pinpointed a total of eight QTLs with multiple-effects for fruit length, fruit width, and fruit shape index across the genomic intervals positioned on four different chromosomes (Chr-02, Chr-06, Chr-07, and Chr-11) (Figure 3; Table 2). Two QTLs of fruit length (FL-6.1 and FL-7.1) were expressed as major-effect and minor-effect QTLs, which explained the fruit length variations (short size and large size fruits) with 20.48% and 5.82% PVE, respectively. Three QTLs of fruit width were identified (FW-7.1 “major-effect”, FW-7.2 “minor-effect”, and FW-11.1 “major-effect”), and contributed to the fruit width variations with 7.13% to 14.22% PVE. Three QTLs of fruit shape variations were further classified (FSI-2.1 “major-effect”, FSI-7.1 “minor-effect”, and FSI-7.2 “minor-effect”), and these QTLs contributed to most of the fruit width variations with 5.82~16.95% PVE. We noticed that our identified QTLs and their genetic positions are inconsistent with the previously published results, as shown in Supplementary Table S9, respectively.

Regarding the genomic co-linearity, the fruit size/shape related QTLs were noticed as tightly co-localized with ovary size/shape related QTLs, respectively. The co-localized QTLs (OL-6.1 and FL-6.1, OL-7.1 and FL-7.1, OSI-2.1, and FSI-2.1) significantly demonstrated that obvious shapes of long and wide fruits have high connectivity since the cell structure development at the ovary establishment stage. We also noticed that oblong shaped fruits of the $F_{2:3}$ population have a genetic resemblance with the female watermelon parent (P_1 , with an oblong shape) and moderate and rounded fruits have a resemblance with the male parent (P_2 , rounded shape). Our identified fruit size/shape QTL results similarly suggested an inherited quantitative genetics and transgressive segregation (Figure 6), which is supposed to be controlled by polygenic architecture and is mainly regulated by dominant allelic fashions of comparative watermelon lines that primarily triggers the clear fruit shape variations of oblong and rounded fruit growth throughout the dynamic growth stages in fruits of $F_{2:3}$ families.

Similarly, in melon, a pre-anthesis genetic structure explains the prominent polygenic regulatory mechanism (Ramamurthy and Waters, 2015; Amanullah et al., 2021). In cucumber (*Cucumis sativus*), a strong association was reported between the establishment of an ovary and fruit shape (Weng et al., 2015; Wei et al., 2016). However, the fruit size/shape is mainly determined by cell division and cell expansion during the

vegetative and reproductive growth stages. The fruit shape index is triggered by a predominantly genetic mechanism in tomato, which has been identified at various developmental stages (Lippman and Tanksley, 2001; Eduardo et al., 2007; van der Graaff et al., 2009; van der Knaap et al., 2014; van der Knaap and Ostergaard, 2018). The near isogenic line (NILs) of tomato with the allelic nature of lengthy ovaries produced more elongated fruits than small-shaped NILs (Frary et al., 2000), and the modeling of FSI at flowering stage suggested a pleiotropic effect with a major drag effect of QTL and appeared to be handled by the OVATE gene family (Ku et al., 2000). Actually, fruit length and width increment is not a continual development near the proximal distal axis; but it depends upon the gradual cell division process that occurs during the ovary formation (van der Knaap and Ostergaard, 2018).

Overall, our identified results are in accordance with the few earlier published studies; e.g., it has been stated that differences concerning the elongate/oblong and rounded watermelon shapes can be identified by identical ovaries at pre-anthesis stages, and a single gene with incomplete dominance locus “O”, differential genotypes “OO, oo, Oo,” are mainly responsible for the obvious shapes of elongated, rounded, and blocky shapes (Weetman, 1937; Poole and Grimball, 1945; Tanaka et al., 1995). But, the quantitative genetics of watermelon FSI variations have been similarly stated (Gusmini and Wehner, 2005; Gusmini and Wehner, 2007; Kumar and Wehner, 2013), e.g., the genetic locus “ClFSI-3.2” harboring the O gene has been significantly validated with the homologous SUN gene in tomato (Dou et al., 2018b). Recently, a newly identified allelic fashion revealed that the rounded watermelon shape is produced by the deletion of a 159 bp region in the CDS coding sequence of the *Clat11257* gene (Maragal et al., 2019; Legendre et al., 2020). In our study, we also identified few more genes for FL, FW, and FSI, but strong validation should be required for validation.

Fruit rind thickness (FRT) is mainly associated with resistance or susceptibility to splitting/cracking (Fan et al., 2000; Liao et al., 2019; Yang et al., 2021). Until now, molecular basis studies of watermelon fruit rind thickness have received little attention, and few genetic mapping studies have been conducted. In the recent molecular study of watermelon (Yang et al., 2021), a single major QTL of rind thickness (RTH-2.1) was mapped on Chr-02, which explained 14.74% of the phenotypic variations for the fruits with less rind thickness and signified the allelic dominance of the parental line with less rind thickness. These results were in line with the earlier published results, where genetic segregation analysis exhibited that the rind thickness is controlled by a major-effect locus positioned on Chr-09 (Fan et al., 2000). But in this study, we mapped a single major QTL (FRT-4.1) to the genetic location of Chr-04. This explained the 14.74% phenotypic variation for fruits with thinner rinds (Figure 3; Table 2).

As far as we know, our identified QTL region of FRT contradicted to the earlier reported studies (Supplementary Table S9), thus exhibiting the novel genetic loci controlling fruit rind hardness. Further, a frequent uniform distribution of genetic segregation was observed in fruits of $F_{2:3}$ plant families (Figure 6). In the other previous study, extremely significant and positive associations were observed for individual fruit weight, cracking, and rind thickness of cherry fruit (Yamaguchi et al., 2014). Brinjal fruit rinds with high firmness showed suitable resistance to fruit cracking due to their thicker peels and affected the prolonged storage and shelf life (Liu et al., 2007). Regarding the genetically and physiologically understandings, it was reported that differential watermelon rind thickness level is similarly interconnected with reliable rind hardness, cracking resistance, and susceptibility (Li et al., 2016; Liao et al., 2019), as well as fruit weight (Yang et al., 2021). It has recently been reported that fruits with rind thickness variations bear dissimilar types of cell size and shapes, particularly due to the presence and absence of lignin accumulation in the rind cell walls (Gao et al., 2013; Yang et al., 2021). However, it was stated that class III peroxidase genes are primarily responsible for regulating the internal structure of lignin accumulation and cell wall structure (Yang et al., 2022).

Fruit flesh firmness (FFF) is a primary attribute of the premium quality and shelf life of edible fruits. A significant changes in flesh firmness involves a series of natural and complex physiological changes that trigger the metabolism of cell wall, cellulose, hemicellulose, and pectin (Brummell et al., 1999; Girard et al., 2012; Shi et al., 2013; Yoko et al., 2014; Sun et al., 2020), and mainly regulated by numerous genes and metabolic networks (Brummell et al., 1999; Brummell and Harpster, 2001; Liao et al., 2019). In contrast, fruit flesh softening (loss of firmness) is the ultimate effect of the respiratory process and ethylene bio-synthesis factor (Liao et al., 2019). Until now, few genetic mapping studies effectively mapped the major locus of regulating the flesh firmness in watermelon fruit. Juarez et al. (2013) developed novel SNP markers, constructed a genetic linkage map using the biparental F_2 generation, and discovered a major QTL region of controlling the watermelon flesh firmness in the 9th linkage group (LG). Liu et al. (2014) re-sequenced two comparative parental lines and identified the candidate region harboring the important genes regulating the edge flesh firmness on the 9th LG of the watermelon linkage map, using a derived F_2 plant population. Lu et al. (2016) performed molecular mapping and traced the major genes of edge flesh firmness on the 4th, 6th, and 8th LGs of watermelon. Gao et al. (2016) used simple sequence repeat markers and a closely linked QTL marker to identify the gene controlling the firmness of watermelon flesh. Gao (2018) performed primary genetic mapping and detected a physical interval of 4.7 Mb, harboring a potential gene for controlling watermelon flesh firmness. Sun et al. (2020) used a rapid method

of BSA-sequencing and preliminary mapped the two genetic regions (1.53 Mb on Chr-02 and 195 kb on Chr-08) for regulating flesh firmness. In the current study, one minor-effect QTL (FFF-11.1) was mapped on Chr-11, which explained an individual genetic effect with 9.01% PVE for less firmness of fruit flesh (Table 2), with uniform segregation (Figure 6). The physical position of this QTL was located between the 2340516~3081651 bp, exhibiting a total of 741.14 kb interval (Figure 3). We also noticed that our identified QTL results are inconsistent with the previously published research, as shown in Supplementary Table S9, respectively.

For the genetic understanding of flesh firmness regulation, it was speculated that flesh textural properties are controlled by polygonal architecture, which is the interconnected activities of proteins encoding cell-wall transformation. In watermelon, the transcription factor (MADS-box) is significantly involved in the biological processes of the transformation of plants at vegetative and reproductive growth stages “photoperiodism, pollen development, and floral organs formation, photosynthesis and nutrient metabolism, fruit development stages, maturation stages, and hormonal signal transduction pathways” (Hu et al., 2005; Liu et al., 2010; Huang et al., 2012; Li et al., 2015; Sun et al., 2020). The endogenous crude fiber and pectin content were reported as the main reasons for flesh firmness variations among the botanical groups of wild-type as well as cultivated watermelon, but the metabolic pathway differentiation of pectin and crude fiber might be a fundamental reason for divergent flesh firmness during watermelon domestication (Liu et al., 2013). It was similarly shown that the breakdown of cell walls is tightly connected with the genetic factor of fruit softening due to the ethylene-dependent accumulation of sucrose, which is regulated by the *CmMYB113* factor in melon (Gao et al., 2021). Furthermore, genes encoding glyoxysomal malate synthase, β -D-xylosidase, chloroplastic anthranilate phosphoribosyltransferase (*MELO3C011963*), and histidine kinase (*MELO3C020055*) were discovered to be involved in regulating flesh firmness in the natural population of melon (Nimmakayala et al., 2016).

Watermelon fruit is primarily consumed due to its high Brix content and health benefits (Zhang et al., 2006). The major genetic loci (QTLs) of Brix% have been identified in a few earlier studies; e.g., three minor but consistent QTLs were identified on Chr-08 that accounted for 6.87%, 5.14%, and 5.27% PVE (Sandlin et al., 2012; Liu et al., 2015), and two QTLs exhibited the main loci controlling gene expression for Brix% (Guo et al., 2006). In a recent genetic mapping study, two significant co-localized QTLs (BCC-2.1 and BCC-5.1) were identified for center flesh Brix% and three co-localized QTLs (BCE-2.1, BCE-2.2, and BCE-5.1) were identified for Brix% in the edge part of the flesh (Liang et al., 2022). The detected QTLs tightly shared the mutual genetic contributions and suggested that there might be a single locus for regulating the Brix% in the whole fruit flesh. Even though the Brix QTL (BRX-2.1) has been found on

Chr-02 between 17,657,266 and 18,454,759 bp and shown to be a stable QTL with a strong genetic effect on the Brix% value (Sandlin et al., 2012; Ren et al., 2014; Ren et al., 2018). In this study, we found one major QTL (FBR-10.1), which was mapped at a 890.74 kb genetic interval on Chr-10, which explained the individual phenotypic effect with 13.98% PVE for Brix% (Table 2; Figure 3), and a transgressive segregation of FL and Brix% in fruits of $F_{2:3}$ families (Figure 6) signified that elongated oblong-shaped fruits have a higher Brix% value than small rounded fruits. Overall, our detected QTLs were traced back to previously reported chromosomal regions, possibly indicating a new genomic region for sugar level regulation and biosynthesis due to the divergent genetic backgrounds of watermelon cultivars. For the genetic regulation, it was determined that *Cla000264* “*CITST2*” is the candidate gene regulating Brix% value (Ren et al., 2018). Hence, our mapped QTL regions of watermelon Brix% are novel results (Supplementary Table S9), but delimitation of the mapped region is necessary for candidate gene validation.

Fruit flesh color (FFC) is a pivotal trait and most of the cultivated watermelon fruits display a sweet-tasting red flesh color that is reported to be dominant over the other flesh color categories due to epistatic effects and different gene expression profiling for different pigment synthesis (Henderson et al., 1998; Wehner, 2007; Cazzonelli and Pogson, 2010; Bang et al., 2010; Nakkanong et al., 2012; Grassi et al., 2013; Liu et al., 2015; Lv et al., 2015; Wang et al., 2019; Fang et al., 2022). The genetic locus of controlling red flesh color was mapped on Chr-02 and Chr-08 of watermelon, using a comprehensive linkage map (Hashizume et al., 2003). Then, a new major QTL region “FC-4.1” controlling the red flesh color was identified on Chr-04, with 34.68% PVE (Liu et al., 2015). This QTL was mainly associated with endogenous accumulation of lycopene content. Furthermore, an in-depth genetic mapping study validated the QTL “FC-4.1” and explained the 392,077-bp region on Chr-04, controlling the dominant red flesh color in a six-generation population of watermelon (Wang et al., 2019). Another major effect QTL “FC-10.1” region (about 519 kb) was mapped on Chr-10, which signified the regulation of differentiated pale green flesh colors associated with the major accumulation of endogenous green chlorophyll content in the flesh (Pei et al., 2021). Recently, a QTL mapping study effectively pinpointed the main genetic locus (645-kb interval) locus on Chr-04, controlling the FFC in $F_{2:3}$ mapping population (Liang et al., 2022). In our study, we mapped just one minor QTL (FFC-5.1) at a wide chromosomal region of Chr-05 that explained the individual genetic effect with 7.42% PVE (Table 2; Figure 3), and this QTL was situated at the physical positions of 28704441~32310869 bp, exhibiting the big genetic interval (3.61 Mb). This mapped locus result was noticed as consistent with et al. (2003) but inconsistent with other published results as Supplementary Table S9, respectively.

The key gene for red flesh color in watermelon is *CILCYB*, which is annotated as lycopene beta-cyclase, was finely mapped

on Chr-04 using three genetic populations. In addition, two SNP non-synonymous mutations were found in the coding region of *CILCYB* among red flesh and yellow flesh watermelon accessions (Liu et al., 2015; Wang et al., 2019). Regarding the genetic regulatory mechanism, it has been reported in many studies that the *LCYB* gene (*Cl97C04G070940*) exhibits a relative expression profiling for differentiating the pink to red flesh color gradients (Gusmini and Wehner, 2006; Perkins-Veazie et al., 2006; Bang et al., 2007; Bang et al., 2010; Kang et al., 2010; Guo et al., 2019; Wang et al., 2019). Two flesh color QTLs were identified on Chr-02 and Chr-04, and map-based cloning was performed based on the white-fleshed line and red-fleshed line. The *Cl005011* gene was considered a lycopene β -cyclase (*LCYB*) and candidate gene in the genomic region of Chr-04, and another gene was narrowed down to a region of 1,200 kb on Chr-02 (Zhang et al., 2014). Subsequently, Zhang and colleagues conducted cloning and transgenic analyses of the *LCYB* gene, and the findings revealed that the abundance of the *CILCYB* protein rather than the *CILCYB* transcription level was negatively correlated with lycopene accumulation (Zhang et al., 2020). *CIPHT4;2* (annotated as a function of the chromoplast-localized phosphate transporter) determines the development of flesh color through carotenoid accumulation, and it also controls the level of sweetness, which is regulated by the transcription factors *ClbZIP1* and *ClbZIP2* (Zhang et al., 2017). A single dominant gene, Y^{scr} , was suggested to produce the scarlet red flesh color rather than the coral red flesh color (Gusmini and Wehner, 2006). Another genetic study also revealed that Y^{scr} produces the scarlet red flesh color and was fine-mapped in a 150-Kb region on Chr-06 (Li et al., 2020). Branham et al. (2017) mapped a major QTL (FC.1) associated with β -carotene accumulation for orange flesh on Chr-01 (2.4 Mb interval). *PSY* gene expression profiling regulate the phytoene synthase which may cause the paler and orangeyellow flesh colors to occur in response to maximum endogenous carotenoids synthesis (Guo et al., 2015; Branham et al., 2017; Fang et al., 2022), and *Cl97C10G185970* was recently annotated as a plastid lipid-associated protein for regulating the differentiated pale green flesh color (Pei et al., 2021). However, in our study, we observed the contrasted minor QTL region at Chr-5, that is located at somewhat big genetic interval (Figure 3) that strongly needs to be delimited for identifying the candidate genes for regulating the differentiated pale green flesh color variants.

Fruit rind color (FRC) is a significant factor for evaluating the mature quality of watermelon fruit. The rinds of most commercial/edible watermelons are solid-green (G), light-green (G), or yellow (go). It has been reported that $G > gs > g$ (dominant to recessive) has the allelic relationships of the traits for rind patterns (Weetman, 1937; Porter, 1937; Poole, 1944; Barham, 1956; Henderson, 1989; Henderson, 1991; Henderson, 1992; Rhodes and Zhang, 1995; Rhodes and Dane, 1999; Guner and Wehner, 2004; Kumar and Wehner, 2011). The solid-green

rind was identified as completely dominant over the striped light green (*gs*) and partially dominant over the unique type of light-green, gray or yellowish-green (Johnson and Buckley, 1991; Yang et al., 2015; Kim et al., 2015). Some molecular mapping studies identified a stable major-effect genetic region located on Chr-08, controlling the watermelon fruit rind color (Park et al., 2016; Li et al., 2019). In this QTL mapping study, one major QTL (FRC-2.1) was mapped at the bottom end of Chr-02 that explained an individual genetic effect of 15.12% PVE, and the physical positions (32909754~34162281 bp) exhibited a 1.25 Mb interval (Figure 3). Regarding the genetic regulation, a few genetic studies revealed that watermelon fruit rind coloration is controlled by just a single gene that is located on Chr-08; however, dark-green rinds represented dominant genetic effects over light-green rinds (Park et al., 2016; Li et al., 2018). It was differentially stated that a potential gene (*CICG08G017810* “*CICGMenG*”) is situated on Chr-08, encoding the 2-phytyl-1,4-beta-naphthoquinone methyltransferase (Li et al., 2019). The chromosomal position of Chr-08 was traced with approximately a 262 kb of deletion in the genome assembly of watermelon (97103, v1.0) (Li et al., 2019); but, the identified gene “*CICGMenG*” position of FFC was differentially located within 29,869,645 to 29,901,009 bp at Chr-08 of the improved genome “Charleston Gray”, and exhibited an additional 34 genes. Furthermore, a recent study mapped the FRC regulating locus was positioned from 24,184, 248 to 24,644,537 bp between adjacent SNP markers at the bottom end of Chr-08 (Liang et al., 2022). Thus, our genetic mapping results are inconsistent with the abovementioned studies (Supplementary Table S9), and the genetic location controlling the rind coloration needs to be narrowed down for gene annotation and functional validation.

Many watermelon cultivars possess varying degrees of fruit rind stripes (FRS) with narrow, wide, wavy and blotchy patterns (Yang et al., 2015; Maragal et al., 2022). Many QTL studies have been performed by using the developed biparental mapping populations (F_2 and F_3) and recombinant inbred lines. These studies identified that the differentiated watermelon rind strip pattern is generally controlled by a stable major effect genetic locus positioned on Chr-06 (Park et al., 2016; Zhang et al., 2018; Wu et al., 2019; Guo et al., 2019; Yue et al., 2021; Wang et al., 2022; Liang et al., 2022) and Chr-09 (Maragal et al., 2022), except the minor variations in genetic positions found in this study. Our genetic mapping study exhibited the contrasted QTL results as compared to the earlier reported QTLs. Herein, one minor QTL (FRS-4.1) was mapped at the start end of Chr-04 that explained an individual genetic effect of 9.75% PVE. The physical positions (2031314~2731617 bp) exhibited a 700.30 kb interval (Figure 3) that depicted a total of 16 putative genes.

For the genetic regulation, it was first reported that the single gene locus “S” mainly controls the watermelon FRS (Weetman, 1937). However, FRS regulation also depends upon the divergences in genetic background of watermelon cultivars.

Some alleles at the *g* locus positioned on Chr-08 control the wide-stripe (g^W), medium-stripe (g^M), narrow-stripe (g^N), and solid-light green or light-green (*g*), with a dominance order of $G > g^W > g^M > g^N > g$ (Kim et al., 2015; Lou and Wehner, 2016; Zhang et al., 2018). In a recent study, linkage mapping was done by using the two differentially derived mapping populations and novel stable QTLs/genes regions were spotted on Chr-09. The comparative genomic analysis revealed two candidate genes “*Cla97C09G175170* and *Cla97C09G175150*” regulating the blotchy stripes and wavy type rind stripe pattern, and sequence analysis of the *Cla97C09G175170* gene exhibited the 3 bp deletion on the 11th exon associated with strip color (Maragal et al., 2022). To the best of our knowledge, the genetic locus (FRS-4.1) for wavy type rind stripes trait has been identified for the first time in our study (Supplementary Table S9), but more molecular research is needed to further clarify the results variations and identify gene homologs contributing to differentiating the rind stripe of watermelon.

Seed QTLs

Seed is an integral part of the plant life cycle and a significant determinant of growth and development. Seed genetics have been extensively studied in major food crops that are consumed directly or indirectly as grains or seeds, such as rice, wheat, and soybean, but few genetic studies are available for vegetable crops (Amanullah et al., 2021). In watermelon, a wide range of seed size variation is present among the various cultivars (Guo et al., 2020). The inheritance of seed size/shape was first reported, and small-to-large sized seeds were found to be dominant over medium-sized seeds (Poole et al., 1941). Later, the quantitative nature of seed size/shape variation was further observed in a few studies.

Until now, a total of 61 QTLs (twenty seed length QTLs, nineteen seed width QTLs, 3 seed thickness QTLs, and nineteen QTLs of seed weight) have been identified with major variations in PVE% across the different chromosomal positions of the whole genome chromosome of watermelon “Chr-01, Chr-03, Chr-06, Chr-07, Chr-10, LG-04, and LG-09” using numerous mapping populations derived from crossing of wild type and cultivar parent lines (Prothro et al., 2012a; Prothro et al., 2012b; Sandlin et al., 2012; Ren et al., 2012; Meru and McGregor, 2013; Liu et al., 2014; Kim et al., 2015; Zhou et al., 2016; Li et al., 2018; Meru et al., 2018; Luan et al., 2019; Guo et al., 2020; Li et al., 2021; Osae et al., 2021; Liang et al., 2022). In this study, three major QTLs of SWt (SWt-4.1, SWt-5.1, and SWt-10.1), two co-localized QTLs of STh and SL (STh-7.1 & SL-7.1), two major QTLs of SW (SW-4.1 and SW-10.1), one major QTL (SSI-6.1), and one major QTL (SCT-1.1) were detected on six differential chromosomes (Chr-01, Chr-04, Chr-05, Chr-06, Chr-07, and Chr-10) (Table 2; Figure 3). The QTL results of Chr-04, Chr-06, Chr-7, and Chr11 are in line with the earlier published QTL results of the above mentioned studies; but, we observed that the

mapped QTL of seed weight (SWt-5.1) is a novel QTL because it wasn't reported earlier (Supplementary Table S9). Further, our identified co-QTLs of seed width, seed length, and seed thickness traits on Chr-07 and Chr-10 also signified the stable and major genetic effects. These QTLs were also consistent and supported the contribution of uniform distribution of segregation also displayed the mutual quantitative inheritance of comparative watermelon parent lines within seeds of $F_{2:3}$ families (Figure 7).

Regarding the understanding of seed genetic mechanisms, it has been stated that the large-to-medium length of seed is controlled by two major genes (*l* and *s*), and small-sized seeds seem dominant over medium-sized seeds. Later, two genes (*Ti* for tiny seed and *ts* for tomato seed) were also reported for small-sized seeds (Poole et al., 1941; Zhang et al., 1994; Tanaka et al., 1995; Zhang, 1996). Later, Zhang and Zhang (2011) reported that a recessive gene pair significantly determines the size of seed, but additional modifiers are also involved in the seed size regulations. Watermelon seed size was considered as a quality trait and supposed to be controlled by a single dominant gene (Kim et al., 2015). Fine genetic mapping revealed the major-effect locus of controlling the seed size/shape “qSS-6.1”, harboring a total of three candidate genes (*Cla009291*, *Cla009301*, and *Cla009310*) (Li et al., 2018). The *Cla009291* gene was exhibited as encoding the MDR protein “mdtK” and was differentially expressed in the seed development stages of large and small seeded lines. The *Cla009310* gene exhibited a major SNP in the 1st exon region and was presumed to be a candidate gene that encodes an unknown protein for regulating the seed size/shape between comparative watermelon lines. Further, *Cla009301* was found as the homolog of SRS3 (SMALL AND ROUND SEED) for a BY-kinesin-like protein, known as the seed size regulator in rice (Kanako et al., 2010). Further, the relative expression profiling of the *Cla007520* gene was noticed to be higher at the early stage of seed formation, but the *Cla007520* gene seemed to belong to the CPP protein family and its related amino acid components were similar to those of the *AtTSO1*, responsible for promoting cell proliferation initiation during ovule development (Luan et al., 2019). However, seed size and shape, as well as their associated traits, are elastic in nature and change in different environmental locations (Fisher et al., 2017). Thus, our reported seed-related QTL segments might harbor some potential genes that would exhibit significant expression profiling but need to be narrowed down, respectively. Furthermore, the identified genes of all traits were categorized into the molecular function, biological process, cellular components, and GO terms with significant enrichment (Figure 9; Supplementary Tables S7,S8). There were a few important GO enrichment terms and KEGG pathways which signified the polygenic phenomenon for regulatory mechanisms of ovary, fruit, and seed phenotypes. In addition, some explicit genes were also indicating the same regulatory pathways for different traits, and those genes were presented in tightly co-localized QTL regions with high LOD scores and major PVE%.

A comprehensive gene descriptions have been reported in our current study but there is still no direct evidence about stable QTLs/genes controlling these traits of watermelon due to extreme divergences in genetic backgrounds.

In summary, we assumed that our identified genes might harbor expression profiling for the genetic regulation of watermelon phenotypes, but gene functional validation is necessary based on fine genetic mapping of the mapped QTL region. Further, our constructed linkage map and mapped QTL regions will provide an important genetic basis for comparative genetic mapping and marker-assisted selection (MAS) in watermelon.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA878948.

Author contributions

SA designed and completed the molecular genetics and breeding experiments, did the formal analysis, drafted the manuscript, and reviewed & edited the manuscript. SL helped in performing the formal analysis. BO, TY participated in the field sampling. FA assisted the theoretical guidance. XW provided the seeds of the watermelon parent lines. MG assisted in the practical guidance. HL, PG and FL supervised the research project, reviewed & edited the manuscript. All authors contributed to the article and approved the submitted version.

References

- Aguado, E., Garc'ia, A., Iglesias-Moya, J., Romero, J., Wehner, T. C., Go'mez-Guillamo'n, M. L., et al. (2020). Mapping a partial andromonoecy locus in *Citrullus lanatus* using BSA-seq and GWAS approaches. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01243
- Allen, G. C., Flores-Vergara, M. A., Krasnynski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384
- Amanullah, S., Gao, P., Osae, B. A., Saroj, A., Yang, T., Liu, S., et al. (2021). Genetic linkage mapping and QTLs identification for morphology and fruit quality related traits of melon by SNP based CAPS markers. *Sci. Hortic.* 278, 109849. doi: 10.1016/j.scienta.2020.109849
- Amanullah, S., Osae, B. A., Yang, T., Li, S., Abbas, F., Liu, S., et al. (2022). Development of whole genome SNP-CAPS markers and preliminary QTL mapping of fruit pedicel traits in watermelon. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.879919
- Amanullah, S., Saroj, A., Osae, B. A., Liu, S., Liu, H., Gao, P., et al. (2020). Detection of putative QTL regions associated with ovary traits in melon using SNP-CAPS markers. *Sci. Hortic.* 270, 109445. doi: 10.1016/j.scienta.2020.109445
- Baboli, Z. M., and Kordi, A. A. S. (2010). Characteristics and composition of watermelon seed oil and solvent extraction parameters effects. *J. Am. Oil Chem. Soc.* 87, 667–671. doi: 10.1007/s11746-010-1546-5
- Bang, H., Davis, A. R., Kim, S., Leskovar, D. I., and King, S. R. (2010). Flesh color inheritance and gene interactions among canary yellow, pale yellow, and red watermelon. *J. Am. Soc. Hortic. Sci.* 135 (4), 362–368. doi: 10.21273/JASHS.135.4.362
- Bang, H., Kim, S., Leskovar, D., and King, S. (2007). Development of a codominant CAPS marker for allelic selection between canary yellow and red watermelon based on SNP in lycopene β -cyclase (*LCYB*) gene. *Mol. Breed.* 20, 63–72. doi: 10.1007/s11032-006-9076-4
- Barham, W. S. (1956). A study of the royal golden watermelon with emphasis on the inheritance of the chlorotic condition characteristic of this variety. *Proc. Am. Soc. Hortic. Sci.* 67, 487–489.
- Branham, S., Vexler, L., Meir, A., Tzuri, G., Frieman, Z., Levi, A., et al. (2017). Genetic mapping of a major codominant QTL associated with β -carotene accumulation in watermelon. *Mol. Breed.* 37, 146. doi: 10.1007/s11032-017-0747-0
- Brummell, D. A., and Harpster, M. H. (2001). Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. *Plant Mol. Biol.* 47, 311–340. doi: 10.1007/978-94-010-0668-2_18
- Brummell, D. A., Harpster, M. H., and Dunsmuir, P. (1999). Differential expression of expansion gene family members during growth and ripening of tomato fruit. *Plant Mol. Biol.* 39 (1), 161–116. doi: 10.1023/a:1006130018931
- Cazzonelli, C. I., and Pogson, B. J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trends Plant Sci.* 15 (5), 266–274. doi: 10.1016/j.tplants.2010.02.003

Funding

This research was financially supported by Taishan Industrial Leading Talents Project (No. LJNY202112) and China Agriculture Research System of MOF and MARA (No. CARS-25).

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.2022.1034952/full#supplementary-material>

- Cheng, Y., Luan, F., Wang, X., Gao, P., Zhu, Z., Liu, S., et al. (2016). Construction of a genetic linkage map of watermelon (*Citrullus lanatus*) using CAPS and SSR markers and QTL analysis for fruit quality traits. *Sci. Hortic.* 202, 25–31. doi: 10.1016/j.scienta.2016.01.004
- Chomici, G., and Renner, S. S. (2015). Watermelon origin solved with molecular phylogenetics including linnaean material: another example of museumics. *New Phytol.* 205, 526–532. doi: 10.1111/nph.13163
- 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. *Fly* 6. doi: 10.4161/fly.19695
- Costa, F., Peace, C. P., Stella, S., Serra, S., Musacchi, S., Bazzani, M., et al. (2010). QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (*Malus domestica* borkh.). *J. Exp. Bot.* 61, 3029–3039. doi: 10.1093/jxb/erq130
- Costa, F., Stella, S., Van de Weg, W. E., Guerra, W., Cecchinell, M., Dallavia, J., et al. (2005). Role of the genes md-ACO1 and md-ACS1 in ethylene production and shelf life of apple (*Malus domestica* borkh.). *Euphytica* 141, 181–190. doi: 10.1007/s10681-005-6805-4
- Costa, F., Van de Weg, W. E., Stella, S., Dondini, L., Pratesi, D., Musacchi, S., et al. (2008). Map position and functional allelic diversity of md-Exp7, a new putative expansion gene associated with fruit softening in apple (*Malus domestica* borkh.) and pear (*Pyrus communis*). *Tree Genet. Gen.* 4, 575–586. doi: 10.1007/s11295-008-0133-5
- Dou, J. L., Lu, X. Q., Ali, A., Zhao, S. J., Zhang, L., He, N., et al. (2018a). Genetic mapping reveals a marker for yellow skin in watermelon (*Citrullus lanatus* L.). *PloS One* 13, e0200617.
- Dou, J. L., Zhao, S. J., Lu, X. Q., He, N., Zhang, L., Ali, A., et al. (2018b). Genetic mapping reveals a candidate gene (*ClFS1*) for fruit shape in watermelon (*Citrullus lanatus* L.). *Theor. Appl. Genet.* 131, 947–958. doi: 10.1007/s00122-018-3050-5
- Eduardo, I., Arus, P., Monforte, A. J., Obando, J., Fernandez-Trujillo, J. P., Martinez, J. A., et al. (2007). Estimating the genetic architecture of fruit quality traits in melon using a genomic library of near isogenic lines. *J. Am. Soc. Hortic. Sci.* 132, 80–89. doi: 10.21273/JASHS.132.1.80
- Fang, X., Gao, P., Luan, F., and Liu, S. (2022). Identification and characterization roles of *Phytoene synthase* (*PSY*) genes in watermelon development. *Genes* 13 1189. doi: 10.3390/genes13071189
- Fang, X., Liu, S., Gao, P., Liu, H., Wang, X., Luan, F., et al. (2020). Expression of *CIPAP* and *CIPSY1* in watermelon correlates with chromoplast differentiation carotenoid accumulation, and flesh color formation. *Sci. Hortic.* 270, 109437. doi: 10.1016/j.scienta.2020.109437
- Fan, M., Xu, Y., Zhang, H., Ren, H. Z., Kang, G. B., Wang, Y. J., et al. (2000). QTL mapping and genetic effect analysis of watermelon fruit traits. *Acta Gen.* 27 (10), 902–910.
- Fisher, J., Bensal, E., and Zamir, D. (2017). Bimodality of stable and plastic traits in plants. *Theor. Appl. Genet.* 130, 1915–1926. doi: 10.1007/s00122-017-2933-1
- Frary, A., Nesbitt, T. C., Grandillo, S., van der Knaap, E., Cong, B., Liu, J., et al. (2000). fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science* 289, 85–88. doi: 10.1126/science.289.5476.85
- Fursa, T. B. (1972). On the taxonomy of the genus *Citrullus* schad. *Bot. Zh.* 57, 31–34.
- Gao, L. (2018). Transcriptome analysis and fine mapping of major genes controlling flesh firmness and sour flesh in watermelon. Beijing. *Chin. Acad. Agric. Sci.*
- Gao, G., Duan, X. Y., Jiang, H. C., Yang, F., and Qi, H. Y. (2021). *CmMYB113* regulates ethylene-dependent sucrose accumulation in postharvest climacteric melon fruit. *Postharvest Biol. Technol.* 181, 111682. doi: 10.1016/j.postharvbio.2021.111682
- Gao, M. L., Yuan, C. Z., and Wang, Y. F. (2013). Comparative study on the microstructure of watermelon peel. *Jiangsu Agric. Sci.* 41, 133–135.
- Gao, L., Zhao, S. J., Lu, X. Q., He, N., Zhu, H. J., Dou, J. L., et al. (2016). Linkage analysis on flesh firmness of watermelon [*Citrullus lanatus* (Thunb.) mansf.] by SSR molecular marker. *J. Plant Genet. Resour.* 17, 866–870.
- Girard, A. L., Mounet, F., Lemaire-Chamley, M., Gaillard, C., Elmorjani, K., Vivancos, J., et al. (2012). Tomato *GDSL1* is required for cutin deposition in the fruit cuticle. *Plant Cell* 24 (7), 3119–3134. doi: 10.1105/tpc.112.101055
- Gómez, J. M. (2004). Bigger is not always better: conflicting selective pressure on seed size in quercus ilex. *Evolution* 58, 71–80. doi: 10.1111/j.0014-3820.2004.tb01574.x
- Grassi, S., Piro, G., Lee, J. M., Zheng, Y., Fei, Z. J., Dalessandro, G., et al. (2013). Comparative genomics reveals candidate carotenoid pathway regulators of ripening watermelon fruit. *BMC Genom.* 14 (1), 781. doi: 10.1186/1471-2164-14-781
- Grumet, R., McCreight, J. D., McGregor, C., Weng, Y., Mazourek, M., Reitsma, K., et al. (2021). Genetic resources and vulnerabilities of major cucurbit crops. *Genes* 12 1222. doi: 10.3390/genes12081222
- Guner, N., and Wehner, T. C. (2004). The genes of watermelon. *Hortic. Sci.* 39, 1175–1182.
- Guo, S. G., Xu, Y., Zhang, H. Y., and Gong, G. Y. (2006). QTL analysis of soluble solids content in watermelon under different environments. *Mol. Plant Breed.* 4, 393–398.
- Guo, Y., Gao, M., Liang, X., Xu, M., Liu, X., Zhang, Y., et al. (2020). Quantitative trait loci for seed size variation in cucurbits – a review. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00304
- Guo, S., Sun, H., Zhang, H., Liu, J., Ren, Y., Gong, G., et al. (2015). Comparative transcriptome analysis of cultivated and wild watermelon during fruit development. *PloS One* 10, e0130267. doi: 10.1371/journal.pone.0130267
- Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–58. doi: 10.1038/ng.2470
- Guo, S., Zhao, S., Sun, H., Wang, X., Wu, S., Lin, T., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nat. Genet.* 51 (1646) 1623. doi: 10.1038/s41588-019-0518-4
- Gusmini, G., and Wehner, T. C. (2005). Foundations of yield improvement in watermelon. *Crop Sci.* 45, 141–146. doi: 10.2135/cropsci2005.0141a
- Gusmini, G., and Wehner, T. C. (2006). Qualitative inheritance of rind pattern and flesh color in watermelon. *J. Hered.* 97, 177–185. doi: 10.1093/jhered/esj023
- Gusmini, G., and Wehner, T. C. (2007). Heritability and genetic variance estimates for fruit weight in watermelon. *Hortic. Sci.* 42, 1332–1336. doi: 10.21273/HORTSCI.42.6.1332
- Hashizume, T., Shimamoto, I., Harushima, Y., Yui, M., Sato, T., Imai, T., et al. (1996). Construction of a linkage map for watermelon (*Citrullus lanatus* (Thunb.) matsum & nakai) using random amplified polymorphic DNA (RAPD). *Euphytica* 90 (3), 265–273. doi: 10.1007/BF00027475
- Hashizume, T., Shimamoto, I., and Hirai, M. (2003). Construction of a linkage map and QTL analysis of horticultural traits for watermelon [*Citrullus lanatus* (THUNB.) MATSUM & NAKAI] using RAPD, RFLP and ISSR markers. *Theor. Appl. Genet.* 106 (5), 779–785. doi: 10.1007/s00122-002-1030-1
- Henderson, W. R. (1989). Inheritance of orange flesh color in watermelon. *Cucurbit Genet. Cooperative Rep.* 12, 59–63.
- Henderson, W. R. (1991). Gene list for watermelon. *Cucurbit Genet. Cooperative Rep.* 14, 129–138.
- Henderson, W. R. (1992). Corrigenda to 1991 watermelon gene list (CGC14:129–138). *Cucurbit Genet. Cooperative Rep.* 15, 110.
- Henderson, W. R., Scott, G. H., and Wehner, T. C. (1998). Interaction of flesh color genes in watermelon. *J. Hered.* 89, 50–53. doi: 10.1093/jhered/89.1.50
- Huang, F., Chi, Y. J., and Yu, D. Y. (2012). Research advances of MADS-box genes in plants. *J. Nanjing Agric. Univ.* 35, 9–18.
- Huang, Z., Van Houten, J., Gonzalez, G., Xiao, H., and van der Knaap, E. (2013). Genome-wide identification, phylogeny and expression analysis of *SUN*, *OFF* and *YABBY* gene family in tomato. *Mol. Genet. Genom.* 288, 111–129. doi: 10.1007/s00438-013-0733-0
- Hu, L. F., Jin, Z. Q., and Xu, B. Y. (2005). Effect of MADS-box genes to fruit development and ripeness. *Mol. Plant Breed.* 3, 415–420.
- Johnson, C. E., and Buckley, B. (1991). A genetic study of rind color patterns in watermelon. *Hortic. Sci.* 26 (5), 484–498. doi: 10.21273/HORTSCI.26.5.496d
- Juarez, B., King, J. J., Bachlawa, E., Mills, J. M., and Wentzell, A. M. (2013). Patent No: WO2013033611A1 (St. Louis, MS: Seminis Vegetable Seeds, Inc). Patent Application Publication, Pub.No.: US 2013/0055466A1.
- Kanako, K., Shigeru, K., Katsuyuki, O., Yuki, A., Tsuyu, A., Izumi, K., et al. (2010). A novel kinesin 13 protein regulating rice seed length. *Plant Cell Physiol.* 51, 1315–1329. doi: 10.1093/pcp/pcq092
- Kang, B. S., Zhao, W. E., Hou, Y. B., and Tian, P. (2012). Expression of carotenogenic genes during the development and ripening of watermelon fruit. *Sci. Hortic.* 124, 368–375. doi: 10.1016/j.scienta.2010.01.027
- Kaźmińska, K., Hallmann, E., Rusaczek, A., Korzeniewska, A., Sobczak, M., Filipczak, J., et al. (2018). Genetic mapping of ovary colour and quantitative trait loci for carotenoid content in the fruit of *Cucurbita maxima* duchesne. *Mol. Breed.* 38 (9), 114. doi: 10.1007/s11032-018-0869-z
- Khadiji-Khub, A. (2015). Physiological and genetic factors influencing fruit cracking. *Acta Physiol. Plant* 37 1718. doi: 10.1007/s11738-014-1718-2
- Kim, H., Han, D., Kang, J., Choi, Y., Levi, A., Lee, G. P., et al. (2015). Sequence characterized amplified polymorphism markers for selecting rind stripe pattern in watermelon (*Citrullus lanatus* L.). *Hortic. Environ. Biotech.* 56, 341–349. doi: 10.1007/s13580-015-0017-1
- Ku, H. M., Grandillo, S., and Tanksley, S. D. (2000). fw8.1, a major QTL, sets the pattern of tomato carpel shape well before anthesis. *Theor. Appl. Genet.* 101, 873–878. doi: 10.1007/s001220051555
- Kumar, R., and Wehner, T. C. (2013). Quantitative analysis of generations for inheritance of fruit yield in watermelon. *Hortic. Sci.* 48, 844–847. doi: 10.21273/HOTSCI.48.7.844

- Kumar, R., and Wehner, T. C. (2011). Discovery of second gene for solid dark green versus light green rind pattern in watermelon. *J. Hered.* 102, 489–493. doi: 10.1093/jhered/esr025
- Legendre, R., Kuzy, J., and McGregor, C. (2020). Markers for selection of three alleles of *ClSun25-26-27a* (*Clao11257*) associated with fruit shape in watermelon. *Mol. Breed.* 40, 19. doi: 10.1007/s11032-020-1104-2
- Levi, A., Thomas, C., Trebitsh, T., Salman, A., King, J., Karalius, J., et al. (2006). An extended linkage map for watermelon based on SRAP, AFLP, SSR, ISSR, and RAPD markers. *J. Am. Soc. Hortic. Sci.* 131 (3), 393–402.
- Levi, A., Thomas, C. E., Zhang, X., Joobeur, T., Dean, R. A., Wehner, T. C., et al. (2001). A genetic linkage map for watermelon based on randomly amplified polymorphic DNA markers. *J. Am. Soc. Hortic. Sci.* 126, 730–737.
- Liang, X., Gao, M., Amanullah, S., Guo, Y., Liu, X., Xu, H., et al. (2022). Identification of QTLs linked with watermelon fruit and seed traits using GBS-based high-resolution genetic mapping. *Sci. Hortic.* 303, 111237. doi: 10.1016/j.scienta.2022.111237
- Liao, N., Hu, Z., Li, Y., Hao, J., Chen, S., Xue, Q., et al. (2019). Ethylene-responsive factor 4 is associated with the desirable rind hardness trait conferring cracking resistance in fresh fruits of watermelon. *Plant Biotechnol.* 18 (4), 1066–1077. doi: 10.1111/pbi.13276
- 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, N., Kong, S., Zhou, D., Shang, J., Wang, J., Li, N., et al. (2021). A 13.96-kb chromosomal deletion of two genes is responsible for the tomato seed size in watermelon (*Citrullus lanatus*). *Plant Breed.* 140 (40), 1–9. doi: 10.1111/pbr.12954
- Li, K., Li, S., Luo, Y. B., Tian, H. Q., and Zhu, B. Z. (2015). Research progress on tomato fruit ripening regulator MADS-RIN. *Chin. Agric. Sci. Bull.* 31, 126–131.
- Li, S., Pan, Y. P., Wen, C. L., Li, Y. H., Liu, X. F., Zhang, X. L., et al. (2016). Integrated analysis in bi-parental and natural populations reveals *CsCLAVATA3* (*CsCLV3*) underlying carpel number variations in cucumber. *Theor. Appl. Genet.* 129, 1007–1022. doi: 10.1007/s00122-016-2679-1
- Lippman, Z., and Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme size in tomato using a cross between the small fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. giant heirloom. *Genetics* 158, 413–422. doi: 10.1093/genetics/158.1.413
- Li, N., Shang, J. L., Wang, J. M., Zhou, D., Li, N. N., and Ma, S. W. (2018). Fine mapping and discovery of candidate genes for seed size in watermelon by genome survey sequencing. *Sci. Rep.* 8, 843–854. doi: 10.1038/s41598-018-36104-w
- Li, N., Shang, J., Wang, J., Zhou, D., Li, N., and Ma, S. (2020). Discovery of the genomic region and candidate genes of the scarlet red flesh color (*Y^{sc}*) locus in watermelon (*Citrullus lanatus* L.). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00116
- Liu, C. Q., Gao, P., and Luan, F. S. (2014). A linkage map for watermelon and QTL analysis for fruit related traits. *Sci. Agric. Sin.* 47, 2814–2829.
- Liu, S., Gao, P., Wang, X., Davis, A., Baloch, A., and Luan, F. (2015). Mapping of quantitative trait loci for lycopene content and fruit traits in *Citrullus lanatus*. *Euphytica* 202, 411–426. doi: 10.1007/s10681-014-1308-9
- Liu, S., Gao, P., Zhu, Q., Luan, F., Davis, A. R., and Wang, X. (2016). Development of cleaved amplified polymorphic sequence markers and a CAPS based genetic linkage map in watermelon (*Citrullus lanatus* [Thunb.] matsum. and nakai) constructed using whole genome re-sequencing data. *Breed. Sci.* 66, 244–259. doi: 10.1270/jsbbs.66.244
- Liu, J. A., He, H. J., Guo, S. G., Zhang, H. Y., Ren, Y., Gong, G., et al. (2013). Physiological and biochemical mechanism for watermelon fruit ripening and softening. *J. Fruit Sci.* 30, 813–818.
- Liu, L., Sun, T., Liu, X., Guo, Y., Huang, X., Gao, P., et al. (2019). Genetic analysis and mapping of a striped rind gene (*st3*) in melon (*Cucumis melo* L.). *Euphytica* 215, 20. doi: 10.1007/s10681-019-2353-1
- Liu, Z., Xue, J., Jin, F., and Bai, Y. (2007). Relationship between crack and peel structure of eggplant and its heterosis. *North China Agric. J.* 03, 141–147.
- Liu, J. H., Xu, B. Y., Zhang, J., and Jin, Z. Q. (2010). The interaction of MADS box transcription factors and manipulating fruit development and ripening. *Hereditas* 32, 893–902.
- Liu, D., Yang, H., Yuan, Y., Zhu, H., Zhang, M., Wei, X., et al. (2020). Comparative transcriptome analysis provides insights into yellow rind formation and preliminary mapping of the *Clyr* (yellow rind) gene in watermelon. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00192
- Li, B. B., Zhao, S. J., Dou, J. L., Aslam, A., Gebremeskel, H., Gao, L., et al. (2019). Genetic mapping and development of molecular markers for a candidate gene locus controlling rind color in watermelon. *Theor. Appl. Genet.* 132, 2741–2753. doi: 10.1007/s00122-019-03384-3
- Li, B., Zhao, Y., Zhu, Q., Zhang, Z., Fan, C., Amanullah, S., et al. (2017). Mapping of powdery mildew resistance genes in melon (*Cucumis melo* L.) by bulked segregant analysis. *Sci. Hortic.* 220, 160–167. doi: 10.1016/j.scienta.2017.04.001
- Luan, F., Fan, C., Sun, L., Cui, H., Amanullah, S., Tang, L., et al. (2019). Genetic mapping reveals a candidate gene (*eg*) for egusi seed in watermelon. *Euphytica* 215, 182. doi: 10.1007/s10681-019-2504-4
- Lu, F., Xu, Y., Zhao, Y., Cao, D., Feng, J. M., Guo, S. G., et al. (2009). Construction of permanent genetic map and comparative analysis of xijiang hami melon [*Cucumis melo* L. ssp. *melo* convar. *ameri* (Pang.) greb.]. *Acta Hortic. Sin.* 36, 1767–1774.
- Lu, B. Y., Zhou, H. W., Chen, X., Luan, F. S., Wang, X. Z., and Jiang, Y. (2016). QTL analysis of fruit traits in watermelon. *J. Fruit Sci.* 33, 1206–1218.
- Lv, Y., Gao, P., Liu, S., Fang, X., Zhang, T., Liu, T., et al. (2022). Genetic mapping and QTL analysis of stigma color in melon (*Cucumis melo* L.). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.865082
- Lv, P., Li, N., Liu, H., Gu, H. H., and Zhao, W. E. (2015). Changes in carotenoid profiles and in the expression pattern of the genes in carotenoid metabolisms during fruit development and ripening in four watermelon cultivars. *Food. Chem.* 174, 52–59. doi: 10.1016/j.foodchem.2014.11.022
- Lou, L. L., and Wehner, T. C. (2016). Qualitative inheritance of external fruit traits in watermelon. *Hortic. Sci.* 51, 487–496. doi: 10.21273/HORTSCI.51.5.487
- Maragal, S., Rao, E. S., and Lakshmana Reddy, D. C. (2019). Genetic analysis of fruit quality traits in prebred lines of watermelon derived from a wild accession of *Citrullus amarus*. *Euphytica* 215, 199. doi: 10.1007/s10681-019-2527-x
- Maragal, S., Nagesh, G. C., Reddy, D. C. L., and Rao, E. S. (2022). QTL mapping identifies novel loci and putative candidate genes for rind traits in watermelon. *3 Biotech.* 12, 46. doi: 10.1007/s13205-022-03112-7
- McKay, J. W. (1936). Factor interaction in *Citrullus*. *J. Hered.* 27, 110–112. doi: 10.1093/oxfordjournals.jhered.a10482
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi: 10.1101/gr.107524.110
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Meru, G., Fu, Y., Leyva, D., Sarnoski, P., and Yagiz, Y. (2018). Phenotypic relationships among oil, protein, fatty acid composition and seed size traits in cucurbita pepo. *Sci. Hortic.* 233, 47–53. doi: 10.1016/j.scienta.2018.01.030
- Meru, G., and McGregor, C. (2013). Genetic mapping of seed traits correlated with seed oil percentage in watermelon. *Hortic. Sci.* 48, 955–959.
- Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B., and Westoby, M. (2005). A brief history of seed size. *Science* 307, 576–580. doi: 10.1126/science.1104863
- Moreno, E., Obando, J. M., Dos-Santos, N., Fernandez-Trujillo, J. P., Monforte, A. J., and Garcia-Mas, J. (2008). Candidate genes and QTLs for fruit ripening and softening in melon. *Theor. Appl. Genet.* 116, 589–602. doi: 10.1007/s00122-007-0694-y
- Nakkanong, K., Yang, J. H., and Zhang, M. F. (2012). Carotenoid accumulation and carotenogenic gene expression during fruit development in novel interspecific inbred squash lines and their parents. *Agric. Food. Chem.* 60, 5936–5944. doi: 10.1021/jf3007135
- Nimmakayala, P., Tomason, Y. R., Abburi, V. L., Alvarado, A., Saminathan, T., Vajja, V. J., et al. (2016). Genome-wide differentiation of various melon horticultural groups for use in GWAS for fruit firmness and construction of a high resolution genetic map. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01437
- Olsen, M. K., and Wendel, F. J. (2013). Crop plants as models for understanding plant adaptation and diversification. *Plant Evol. Dev.* 4. doi: 10.3389/fpls.2013.00290
- Osae, B. A., Amanullah, S., Liu, H., Liu, S., Saroj, A., Zhang, C., et al. (2022). CAPS marker-based genetic linkage mapping and QTL analysis for watermelon ovary, fruit and seed-related traits. *Euphytica* 218, 39. doi: 10.1007/s10681-022-02990-5
- Osae, B. A., Liu, S., Amanullah, S., Gao, P., Fan, C., and Luan, F. (2021). QTL mapping of significant seed traits of watermelon (*Citrullus lanatus* L.). *Pak. J. Bot.* 53 (3), 923–931. doi: 10.30848/PJB2021-3(24)
- Pan, Y. P., Wang, Y. H., McGregor, C., Liu, S., Luan, F. S., Gao, M. L., et al. (2020). Genetic architecture of fruit size and shape variation in cucurbits: a comparative perspective. *Theor. Appl. Genet.* 133, 1–21. doi: 10.1007/s00122-019-03481-3
- Paris, H. S. (2015). Origin and emergence of the sweet dessert watermelon, *Citrullus lanatus*. *Ann. Bot.* 116, 133–148. doi: 10.1093/aob/mcv077
- Park, S. W., Kim, K. T., Kang, S. C., and Yang, H. B. (2016). Rapid and practical molecular marker development for rind traits in watermelon. *Hortic. Environ. Biotech.* 57, 385–391.

- Pei, S., Liu, Z., Wang, X., Luan, F., Dai, Z., Yang, Z., et al. (2021). Quantitative trait loci and candidate genes responsible for pale green flesh colour in watermelon (*Citrullus lanatus*). *Plant Breed.* 140, 349–359. doi: 10.1111/pbr.12908
- Pereira, L., Ruggieri, V., Perez, S., Alexiou, K. G., Fernandez, M., Jahrmann, T., et al. (2018). QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. *BMC Plant Biol.* 8, 324. doi: 10.1186/s12870-018-1537-5
- Perkins-Veazie, P., Collins, J. K., Davis, A. R., and Roberts, W. (2006). Carotenoid content of 50 watermelon cultivars. *J. Agric. Food Chem.* 54, 2593–2597. doi: 10.1021/jf052066p
- Poole, C. F. (1944). Genetics of cultivated cucurbits. *J. Hered.* 35, 122–128. doi: 10.1093/oxfordjournals.jhered.a105364
- Poole, C., and Grimball, P. (1938). Inheritance of new sex forms in cucumis melo L. *J. Heredity* 30, 21–25. doi: 10.1093/oxfordjournals.jhered.a104626
- Poole, C. F., Grimball, P. C., and Porter, D. R. (1941). Inheritance of seed characters in watermelon. *J. Agric. Res.* 63, 433–456.
- Poole, C. F., and Grimball, P. C. (1945). Interaction of sex, shape, and weight genes in watermelon. *J. Agric. Res.* 71, 533–552. doi: 10.2503/jjshs.64.543
- Porter, D. R. (1937). Inheritance of certain fruit and seed characters in watermelons. *Hilgardia* 10, 489–509. doi: 10.3733/hilg.v10n12p489
- Prothro, J., Sandlin, K., Abdel-Haleem, H., Bachlava, E., and McGregor, C. (2012a). Main and epistatic quantitative trait loci associated with seed size in watermelon. *J. Am. Soc. Hortic. Sci.* 137, 452–457. doi: 10.21273/JASHS.137.6.452
- Prothro, J., Sandlin, K., Gill, R., Bachlava, E., White, V., and Knapp, S. (2012b). Mapping of the egusi seed trait locus (eg) and quantitative trait loci associated with seed oil percentage in watermelon. *J. Am. Soc. Hortic. Sci.* 137, 311–315. doi: 10.21273/JASHS.137.5.311
- Ramamurthy, K. R., and Waters, B. M. (2015). Identification of fruit quality and morphology QTLs in melon (*Cucumis melo*) using a population derived from *flexuosus* and *cantalupensis* botanical groups. *Euphytica* 204 (1), 163–177. doi: 10.1007/s10681-015-1361-z
- R Core Team (2018). *R: a language and environment for statistical computing* (Vienna: R Foundation for Statistical Computing). Available at: <https://www.R-project.org/>.
- Ren, Y., Guo, S., Zhang, J., He, H., Sun, H., Tian, S., et al. (2018). A tonoplast sugar transporter underlies a sugar accumulation QTL in watermelon. *Plant Physiol.* 176, 836–850. doi: 10.1104/pp.17.01290
- Ren, Y., McGregor, C., Zhang, Y., Gong, G., Zhang, H., Guo, S., et al. (2014). An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). *BMC Plant Biol.* 14, 33. doi: 10.1186/1471-2229-14-33
- Renner, S. S., Chomicki, G., and Greuter, W. (2014). (2313) proposal to conserve the name *Momordica lanata* (*Citrullus lanatus*) (watermelon, cucurbitaceae), with a conserved type, against *Citrullus battich*. *Taxon* 63, 941–942. doi: 10.12705/634.29
- Ren, Y., Zhao, H., Kou, Q., Jiang, J., Guo, S., Zhang, H., et al. (2012). A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. *PLoS One* 7, e29453. doi: 10.1371/journal.pone.0029453
- Rhodes, B., and Dane, F. (1999). Gene list for watermelon. *Cucurbit Genet. Cooperative Rep.* 22, 61–77.
- Rhodes, B., and Zhang, X. (1995). Gene list for watermelon. *Cucurbit Genet. Cooperative Rep.* 18, 69–84.
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., and Mesirov, J. P. (2017). Variant review with the integrative genomics viewer. *Cancer Res.* 1, 77 (21), e31–e34. doi: 10.1158/0008-5472.CAN-17-0337
- Ruffalo, M., LaFramboise, T., and Koyutürk, M. (2011). Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics* 27 (20), 2790–2796. doi: 10.1093/bioinformatics/btr477
- Sandlin, K., Prothro, J., Heesacker, A., Khalilian, N., Okashah, R., Xiang, W., et al. (2012). Comparative mapping in watermelon [*Citrullus lanatus* (Thunb.) matsum. et nakai]. *Theor. Appl. Genet.* 125, 1603–1618. doi: 10.1007/s00122-012-1938-z
- Shi, J. X., Adato, A., Alkan, N., He, Y., Lashbrooke, J., Matas, A. J., et al. (2013). The tomato *sl SHINE 3* transcription factor regulates fruit cuticle formation and epidermal patterning. *New Phytol.* 197 (2), 468–480. doi: 10.1111/nph.12032
- Subburaj, S., Lee, K., Jeon, Y., Tu, L., Son, G., Choi, S., et al. (2019). Whole genome resequencing of watermelons to identify single nucleotide polymorphisms related to flesh color and lycopene content. *PLoS One* 14 (10), e0223441. doi: 10.1371/journal.pone.0223441
- Sun, L., Zhang, Y., Cui, H., Zhang, L., Sha, T., Wang, C., et al. (2020). Linkage mapping and comparative transcriptome analysis of firmness in watermelon (*Citrullus lanatus*). *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00831
- Tadmor, Y., King, S., Levi, A., Davis, A., Meir, A., Wasserman, B., et al. (2005). Comparative fruit colouration in watermelon and tomato. *Food Res. Int.* 38, 837–841.
- Tanaka, T., Wimol, S., and Mizutani, T. (1995). Inheritance of fruit shape and seed size of watermelon. *J. Jpn. Soc. Hortic. Sci.* 64, 543–548.
- Thiel, T., Kota, R., Grosse, I., Stein, N., and Graner, A. (2004). SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. *Nucleic Acids Res.* 32 (1), e5. doi: 10.1093/nar/gnh006
- Tian, Y. F., Wei, Y. Y., Chang, X. K., Yang, M. K., Wu, W. J., and Wei, X. (2012). Effects of seed size on germinating period and seedling quality of cucumber, watermelon and squash. *Shandong Agric. Sci.* 44, 40–42. doi: 10.14083/j.issn.1001-4942.2012.08.039
- van der Graaff, E., Laux, T., and Rensing, S. A. (2009). The WUS homeobox containing (WOX) protein family. *Genome Biol.* 10, 248. doi: 10.1186/gb-2009-10-12-248
- van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., et al. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00227
- van der Knaap, E., and Ostergaard, L. (2018). Shaping a fruit: developmental pathways that impact growth patterns. *Semin. Cell Dev. Biol.* 79, 27–36. doi: 10.1016/j.semcdb.2017.10.028
- Vegas, J., Garcia-Mas, J., and Monforte, A. J. (2013). Interaction between QTLs induces an advance in ethylene biosynthesis during melon fruit ripening. *Theor. Appl. Genet.* 126 (6), 1531–1544. doi: 10.1007/s00122-013-2071-3
- Walters, S. A., Abdelaziz, M., and Bouharrou, R. (2021). Local melon and watermelon crop populations to moderate yield responses to climate change in north Africa. *Climates* 9 (8), 129. doi: 10.3390/cl9080129
- Wang, X., Li, H., Gao, Z., Wang, L., and Ren, Z. (2020). Localization of quantitative trait loci for cucumber fruit shape by a population of chromosome segment substitution lines. *Sci. Rep.* 10, 11030. doi: 10.1038/s41598-020-68312-8
- Wang, C., Qiao, A., Fang, X., Sun, L., Gao, P., Davis, A. R., et al. (2019). Fine mapping of lycopene content and flesh color related gene and development of molecular marker-assisted selection for flesh color in watermelon (*Citrullus lanatus*). *Front. Plant Sci.* 10 1240. doi: 10.3389/fpls.2019.01240
- Wang, D., Zhang, M., Xu, N., Yang, S., Dou, J., Liu, D., et al. (2022). Fine mapping a *CIGS* gene controlling dark-green stripe rind in watermelon. *Sci. Hortic.* 291, 110583. doi: 10.1016/j.scienta.2021.110583
- Weetman, L. M. (1937). Inheritance and correlation of shape, size and color in the watermelon, *Citrullus vulgaris* schrad. *Res. Bull.* 20, 224–256.
- Wehner, T. C. (2007). Gene list for watermelon. *Cucurbit. Genet. Coop. Rep.* 30, 96–120.
- Wehner, T. C., Shetty, N. V., and Elmstrom, G. W. (2001). “Breeding and seed production,” in *Watermelons: characteristics, production, and marketing*. Ed. D. N. Maynard (Alexandria: ASHS Press), 27–73.
- Wei, Q., Fu, W. Y., Wang, Y. Z., Qin, X. D., Wang, J., Li, J., et al. (2016). Rapid identification of fruit length loci in cucumber (*Cucumis sativus* L.) using next-generation sequencing (NGS)-based QTL analysis. *Sci. Rep.* 6, 27496. doi: 10.1038/srep27496
- Weng, Y., Colle, M., Wang, Y., Yang, L., Rubinstein, M., Sherman, A., et al. (2015). QTL mapping in multiple populations and development stages reveals dynamic quantitative trait loci for fruit size in cucumbers of different market classes. *Theor. Appl. Genet.* 128, 1747–1763. doi: 10.1007/s00122-015-2544-7
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis* (New York: Springer).
- Wu, S., Wang, X., Reddy, U., Sun, H., Bao, K., Gao, L., et al. (2019). Genome of ‘Charleston gray’, the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the U.S. national plant germplasm system watermelon collection. *Plant Biotechnol. J.* 17, 2246–2258. doi: 10.1111/pbi.13136
- Wu, S., Zhang, B., Keyhaninejad, N., Rodriguez, G. R., Kim, H. J., Chakrabarti, M., et al. (2018). A common genetic mechanism underlies morphological diversity in fruits and other plant organs. *Nat. Commun.* 9 4734. doi: 10.1038/s41467-018-07216-8
- Xiang, C. P., Shi, X. M., and Zhang, Y. (2002). Effect of seed size on seed quality and yield of bitter melon. *J. Changjiang Vegetables* 35–36.
- Xia, D., Zhou, H., Liu, R., Dan, W., Li, P., Wu, B., et al. (2018). GL3.3, a novel QTL encoding a GSK3/SHAGGY-like kinase, epistatically interacts with GS3, to form extra-long grains in rice. *Mol. Plant* 11, 754–756. doi: 10.1016/j.molp.2018.03.006
- Yamaguchi, H., Sato, S., Isobe, S., Tabata, S., and Fukuoka, H. (2014). Draft genome sequence of eggplant (*Solanum melongena* L.): the representative solanum species indigenous to the old world. *DNA Res.* 21, 649–660. doi: 10.1093/dnares/dsu027

- Yang, T., Amanullah, S., Pan, J., Chen, G., Liu, S., Ma, S., et al. (2021). Identification of putative genetic regions for watermelon rind hardness and related traits by BSA-seq and QTL mapping. *Euphytica* 217, 1–18. doi: 10.1007/s10681-020-02758-9
- Yang, T., Liu, J., Li, X., Amanullah, S., Lu, X., Zhang, M., et al. (2022). Transcriptomic analysis of *Fusarium oxysporum* stress-induced pathosystem and screening of *Fom-2* interaction factors in contrasted melon plants. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.961586
- Yang, H. B., Sungwoo, P., Younghoon, P., Gungpyo, L., Kang, S. C., and Yongkwon, K. (2015). Linkage analysis of the three loci determining rind color and stripe pattern in watermelon. wonye kwahak kisulchi =. *Korean J. Hortic. Sci. Technol.* 33 (4), 559–565. doi: 10.7235/hort.2015.14070
- Yoko, S., Masaki, F., Mamiko, K., Toshitsugu, N., Junji, K., Nobutaka, S. N. K., et al. (2014). Tomato FRUITFULL homologs regulate fruit ripening via ethylene biosynthesis. *Biosci. Biotechnol. Biochem.* 78, 231–237. doi: 10.1080/09168451.2014.878221
- Yue, Z., Ma, R., Cheng, D., Yan, X., He, Y., Wang, C., et al. (2021). Candidate gene analysis of watermelon stripe pattern locus *CLSP* ongoing recombination suppression. *Theor. Appl. Genet.* 134, 3263–3277. doi: 10.1007/s00122-021-03891-2
- Zhang, J. (1996). Inheritance of seed size from diverse crosses in watermelon. *Cucurbit Genet. Coop. Rep.* 19, 67–69.
- Zhang, F., Gong, G., Wang, Q., He, H., and Xu, Y. (2006). Analysis of watermelon quality structure. *J. Fruit Sci.* 23, 266–269.
- Zhang, J., Gong, G. Y., Guo, S. G., Ren, Y., Zhang, H. Y., and Xu, Y. (2014). Fine mapping of the flesh color controlling genes in watermelon (*Citrullus lanatus*). *Cucurbitaceae 2014 Proc.*, 111–116.
- Zhang, J., Guo, S., Ren, Y., Zhang, H., Gong, G., Zhou, M., et al. (2017). High-level expression of a novel chromoplast phosphate transporter CIPHT4;2 is required for flesh color development in watermelon. *New Phytol.* 213, 1208–1221. doi: 10.1111/nph.14257
- Zhang, T., Liu, J., Amanullah, S., Ding, Z., Cui, H., Luan, F., et al. (2021). Fine mapping of *Clao15407* controlling plant height in watermelon. *J. Amer. Soc. Hortic. Sci.* 146, 196–205. doi: 10.21273/JASHS04934-20
- Zhang, X., Rhodes, B., and Wang, M. (1994). “Genes controlling watermelon seed size,” in *Cucurbitaceae '94: Evaluation and enhancement of cucurbit germplasm*. Eds. G. Lester and J. Dunlap (Alexandria, VA: ASHS Press), 144–147.
- Zhang, J., Sun, H., Guo, S., Ren, Y., Li, M., Wang, J., et al. (2020). Decreased protein abundance of lycopene β -cyclase contributes to red flesh in domesticated watermelon. *Plant Physiol.* 183, 1171–1183. doi: 10.1104/pp.19.01409
- Zhang, G. F., and Zhang, J. N. (2011). Inheritance of seed size in watermelon. *Jiangsu Agric. Sci.* 39, 216–217. doi: 10.15889/j.issn.1002-1302.2011.04.170
- Zhang, Z., Zhang, Y., Sun, L., Qiu, G., Sun, Y., Zhu, Z., et al. (2018). Construction of a genetic map for *Citrullus lanatus* based on CAPS markers and mapping of three qualitative traits. *Sci. Hortic.* 233, 532–538. doi: 10.1016/j.scienta.2017.10.029
- Zhou, H. W., Lu, B. Y., Ma, H. Y., Gao, P., Luan, F. S., Gao, Q. F., et al. (2016). QTL mapping of watermelon seed traits. *Acta Hortic. Sin.* 43, 715–723. doi: 10.16420/j.issn.0513-353x.2015-0973
- Zhu, Q., Gao, P., Liu, S., Zhu, Z., Amanullah, S., Davis, A., et al. (2017). Comparative transcriptome analysis of two contrasting watermelon genotypes during fruit development and ripening. *BMC Genom.* 18, 3. doi: 10.1186/s12864-016-3442-3



OPEN ACCESS

EDITED BY

Md. Anowar Hossain,
University of Rajshahi, Bangladesh

REVIEWED BY

Sikandar Amanullah,
Northeast Agricultural University, China
Pranav Pankaj Sahu,
Czech Academy of Sciences, Czechia

*CORRESPONDENCE

Yogesh Vikal
✉ yvikal-soab@pau.edu

SPECIALTY SECTION

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

RECEIVED 06 December 2022

ACCEPTED 02 February 2023

PUBLISHED 15 February 2023

CITATION

Jagtap AB, Yadav IS, Vikal Y, Praba UP,
Kaur N, Gill AS and Johal GS (2023)
Transcriptional dynamics of maize leaves,
pollens and ovules to gain insights into
heat stress-related responses.
Front. Plant Sci. 14:1117136.
doi: 10.3389/fpls.2023.1117136

COPYRIGHT

© 2023 Jagtap, Yadav, Vikal, Praba, Kaur, Gill
and Johal. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Transcriptional dynamics of maize leaves, pollens and ovules to gain insights into heat stress-related responses

Ashok Babadev Jagtap¹, Inderjit Singh Yadav¹, Yogesh Vikal ^{1*},
Umesh Preethi Praba¹, Navneet Kaur¹, Adeshpal Singh Gill¹
and Gurmukh S. Johal²

¹School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India, ²Department of Botany and Pathology, Purdue University, West Lafayette, IN., United States

Heat stress (HS) is one of the alarming issues today due to global warming and is the foremost detrimental to crop production. Maize is one of the versatile crops grown over different agro-climatic conditions. However, it is significantly sensitive to heat stress, especially during the reproductive phase. The heat stress tolerance mechanism is yet to be elucidated at the reproductive stage. Thus, the present study focused on identifying transcriptional changes in two inbreds, LM 11 (sensitive to HS) and CML 25 (tolerant to HS), under intense heat stress at 42°C during the reproductive stage from three tissues viz. flag leaf, tassel, and ovule. Samples from each inbred were collected after 5 days of pollinations for RNA isolation. Six cDNA libraries were constructed from three separate tissues of LM 11 and CML 25 and sequenced using an Illumina HiSeq2500 platform. A total of 2,164 (1127 up-regulated and 1037 down-regulated) differentially expressed genes (DEGs) were identified with 1151, 451, and 562 DEGs in comparisons of LM 11 and CML 25, corresponding to a leaf, pollen, and ovule, respectively. Functional annotated DEGs associated with transcription factors (TFs) viz. AP2, MYB, WRKY, PsbP, bZIP, and NAM, heat shock proteins (HSP20, HSP70, and HSP101/ClpB), as well as genes related to photosynthesis (PsaD & PsaN), antioxidation (APX and CAT) and polyamines (Spd and Spm). KEGG pathways analyses showed that the metabolic overview pathway and secondary metabolites biosynthesis pathway, with the involvement of 264 and 146 genes, respectively, were highly enriched in response to heat stress. Notably, the expression changes of the most common HS-responsive genes were typically much more significant in CML 25, which might explain why CML 25 is more heat tolerant. Seven DEGs were common in leaf, pollen, and ovule; and involved in the polyamines biosynthesis pathway. Their exact role in maize heat stress response would warrant further studies. These results enhanced our understanding to heat stress responses in maize.

KEYWORDS

heat stress, maize, RNA-seq, comparative transcriptomics, differentially expressed genes (DEGs)

Introduction

Diverse environmental challenges pose a severe threat and renewed concern to the world's food security for the burgeoning human population (Tiwari and Yadav, 2019). Climate change might result in a wide variety of impacts on agricultural production (Porter, 2005). The development of genotypes with enhanced abiotic stresses is paramount to reinforce crop productivity and produce enough food to meet the demands of the predicted global population in 2050 (Gilliam et al., 2017). Drastic temperature fluctuations due to climate change frequently occur during plant growth and development (Bita and Gerats, 2013). High temperature stresses exclusively during reproductive phenophase are becoming the main concern for plant scientists under the fast-changing climatic scenario, affecting crop production and productivity worldwide (Tiwari and Yadav, 2019). Therefore, it is imperative to build up heat-resilience crop plants to cope with high temperatures due to climate change.

To gain a comprehensive understanding of the molecular mechanisms involved in heat stress (HS) tolerance, next-generation sequencing (NGS) approaches like transcriptomics or RNA-sequencing (RNA-seq) is a powerful tool for whole-genome gene expression profiling and is especially useful for studying complex gene regulatory networks (McGettigan, 2013; Frey et al., 2015). Transcriptomics has been broadly studied in several crops species like maize (Fernandes et al., 2008; Frey et al., 2015; Shi et al., 2017; Qian et al., 2019; Li and Ye, 2022; Xuhui et al., 2022), rice (Zhang et al., 2013), wheat (Qin et al., 2008; Nandha et al., 2019; Azameti et al., 2022), pepper (Li et al., 2015) and barley (Mangelsen et al., 2011). Various genes and metabolites get activated under heat stress, such as transcriptional factors, hormones, and Heat Shock Proteins (HSPs) that play a crucial role in heat stress tolerance. However, very few studies reported the comparative transcriptome analysis between heat-tolerant and heat-sensitive cultivars of crop plants- rice, maize, and pepper (Shi et al., 2017).

Maize (*Zea mays* L.), the 'Queen of Cereals', is the third most important cereal crop globally (Rouf Shah et al., 2016). It is one of the most versatile emerging C4 crops having high plasticity under diverse agro-climatic conditions across the globe. Forthwith, maize is grown in regions with prevailing 18–27°C habitually optimum temperatures. Nevertheless, it can also be raised at 33–38°C with optimum yield (Tiwari and Yadav, 2019). Temperatures beyond 38°C will drastically impact the heat stress on the maize crop, and consequently, the economic productivity of maize will be less (Koirala et al., 2017). More clearly, several studies have identified high temperature (heat stress) as the main threat to future maize cultivation in distinct relevant production regions like India (Gourdji et al., 2013). Heat stress during the flowering stage in maize decreases chlorophyll content, reduced membrane-thermostability, increases anthesis-silking interval (ASI), causes leaf firing and tassel blast, and reduces pollen viability and yield (Hussain et al., 2019; Inghelandt et al., 2019; Martins et al., 2019; Noor et al., 2019). The morphological and physiological effect of maize under heat stress is discussed in detail by El-Sappah et al. (2022) and Tas (2022).

It has been elucidated that transcription factors (TFs), heat shock proteins (HSPs) response pathways, response to reactive oxygen species (ROS), increasing production of antioxidant and osmoprotectants, and network of hormones participate in plant heat tolerance (Qian et al., 2019). The recent findings revealed that TFs belonging to AP2/EREBP,

MYB, WRKY, bHLH, NAC, and bZIP families play a vital role in regulating heat stress-related responses at the molecular level (Li et al., 2017; Li and Ye, 2022). HSPs function as molecular chaperones and regulate the folding, localization, accumulation, and degradation of protein molecules, and induce the endoplasmic reticulum-localized unfolded protein response (ER-UPR) (Kotak et al., 2007; Hu et al., 2009; Ul haq et al., 2019; Li and Howell, 2021). Also, ascorbate peroxidase (APX) and catalase (CAT) detoxify the ROS produced during heat stress (Asthir, 2015). Moreover, several plant hormones like auxins, abscisic acid (ABA), jasmonic acid (JA), cytokinins (CKs), ethylene, gibberellin, and brassinosteroids are involved in heat stress tolerance (Li et al., 2015; Li et al., 2017).

Presently, little is known about the molecular mechanism of heat resilience in maize (Qian et al., 2019). The transcriptomic responses of maize to heat stress have been reported in a few studies and mainly focused on gene expression changes at the seedling stage (Frey et al., 2015; Li et al., 2017; Shi et al., 2017; Qian et al., 2019; Zhao et al., 2019; Li and Ye, 2022; Xuhui et al., 2022). To our best knowledge, heat stress responses in crop plants at the reproductive stage have received less attention. Moreover, transcriptomic responses in different plant parts of maize during the reproductive phase have not been elucidated yet. Therefore, it is necessary to determine the molecular mechanisms involved in heat stress in maize to understand how maize plants respond and adapt to heat stress at the reproductive stage and breed heat-resilient ready maize crops. In the present study, LM 11 (heat-sensitive) and CML 25 (heat-tolerant) maize inbreds were exposed to high-temperature stress at 42°C during the flowering stage. Transcriptional dynamics among leaf, pollen, and ovule were studied to detect the differential gene expression. It has led to the identification of potential candidate genes that could be deployed for heat-resilient maize breeding.

Materials and methods

Plant materials and sampling

The experimental material comprised two parental maize inbred lines viz. LM 11, heat stress susceptible (HS), and CML 25, heat stress-tolerant (HT). The seedlings of two inbreds were raised during the second week of March in glasshouse conditions at 28°C/23°C and 16 h light (Day)/8 h dark (Night) photoperiod at the School of Agricultural Biotechnology, Punjab Agricultural University, India, during *spring* 2016. The plants were grown till they reached the reproductive stage (Figure 1). The reproductive phase of maize is categorized into six stages, with the emergence of tasselling and silking; followed by a blister where kernels with clear liquid get secreted and filled with milky fluid; accompanied by doughy consistency and extended kernels and milk line progression towards the kernel tip and finally, a black layer formed at the base of grains. At the reproductive growth stage, from tassel emergence to early grain-filling (lag-phase), maize plants were exposed to natural heat stress and experienced 42°C during the daytime and 35°C during the nighttime. Drought stress is confounded naturally during heat stress. To maintain the microclimate conditions with low RH (<40%) and to avoid the compound effect of drought and heat stress, regular irrigation was applied for at least two weeks during tassel emergence until one week after pollination, which increases the probability of irreversible damage due to heat stress. Both inbreds showed differential



FIGURE 1
The maize inbred lines used in the present study. (A) LM 11 (Heat susceptible). (B) CML25 (Heat tolerant).

responses to heat stress for phenological attributes like top leaf firing, tassel blast, pollen viability and shedding duration, kernel number and weight, and yield (Jagtap, 2020). Three different tissue samples, viz. flag leaf, pollen, and ovule from the inbreds, LM 11 and CML 25, were collected in 5 replicates after five days of pollination. Tissue purity was maintained by bagging the tassel and cob. Ovules were isolated from ear florets with a silk length of ~10 cm by removing the silk and ovary wall with forceps and cutting the ovule at its base from the floret under the microscope. Each tissue was pooled (pool of five plants) for each inbred to reduce biological sampling error. A total of six samples (3 tissues x 2 inbreds) were immediately frozen in liquid nitrogen and stored at -80°C until processed for RNA isolations.

Total RNA isolation, library construction, and Illumina sequencing

Total RNA was isolated from each tissue sample (flag leaf, pollen, and ovule) of LM 11 and CML 25 under heat stress conditions using the NucleoSpin RNA Plant kit (Macherey-Nagel, Duren, Germany) according to the manufacturer's protocol. The RNA integrity number (RIN) and concentration were checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). The RNA (1 µg) having RIN values >6 was used for further analysis. 5 µg of total RNA was used for mRNA enrichment, and complementary DNA (cDNA) library was constructed using the manufacturer's protocol of Illumina HiSeq 2500 RNA library preparation kit (Illumina, San Diego, CA). The quality and quantity of cDNA libraries were checked using the Qubit 2.0 Fluorometer

(Thermo Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies, Singapore). The libraries were then sequenced using HiSeq Illumina 2500 sequencing platform (Illumina, San Diego, CA) outsourced from the Nucleome Informatics Pvt. Ltd, Hyderabad, India. High throughput sequencing of transcriptome libraries generated an aggregate of 2.93 billion raw reads accounting for 45.19 Gb of data.

Pre-processing and *de novo* assembly

The raw reads were processed with FastQC (Andrews, 2010) to check the quality of the sequences. Low-quality regions and adapter fragments were removed from the raw reads based on all known Illumina adapter sequences with the options 2:30:10 via Trimmomatic 0.36 (Bolger et al., 2014). The trimmed data were also checked for quality of sequencing before the start of further analysis using fast QC with an average PHRED score of 20. Reads below the length cut-off of 100 nucleotides were discarded. Read pairs with only one surviving read were dropped after trimming. The Q20, Q30, and GC contents were estimated to perform all downstream analyses. Transcriptome assembly was done based on the left. fq and right. fq files using Trinity v2.4.0 (Haas et al., 2013). The transcript abundance was determined at 0.1 dispersion.

Enrichment of differentially expressed genes (DEGs)

The differential gene expression was studied using EdgeR (Robinson et al., 2010). Genes with a false discovery rate (FDR) of <0.05 and a fold change of >2 were considered as differentially expressed. The number of DEGs among and within conditions was plotted as a Venn diagram using Venny tools (<http://bioinfogp.cnb.csic.es/tools/venny/>) (Oliveros, 2007). Volcano plots were prepared to identify the number of transcripts regulated under heat stress conditions in different samples. The volcano center represents the fold change of zero, and either side of the center indicates the down (negative values) and up-regulation (positive values) of transcripts, respectively. Significant DEGs are represented by red and green dots with $|\log_2(\text{fold change})| \geq 2$ and FDR value less than 0.05.

Gene ontology and pathway enrichment analysis

DEGs and consensus sequences of isoforms were mapped to GO classifications using Blast2GO (Conesa and Götz, 2008). Gene Ontology (GO) enrichment was performed for DEGs identified in the leaf, pollen, and ovule to gain insights into their involvement in various functional annotations under heat stress conditions. Around 2,164 DEGs were subjected to GO analysis by the WEGO application (Ashburner et al., 2000). KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis was performed by the KOBAS2.0 packages (Xie et al., 2011). Also, the gene function annotation was accomplished by BLASTX against the databases Nr and Pfam. Pathway analysis of differentially expressed transcripts involved in specific pathways was done using MapMan version 3.6.0 RC1 with a P-value of ≤ 0.05 (Thimm et al., 2004).

Validation of DEGs by quantitative real-time PCR (qRT-PCR)

An aliquot of total RNA isolated from heat-stressed leaf samples of LM 11 (HS) and CML 25 (HT) was used for cDNA synthesis by PrimeScript™ first strand cDNA synthesis kit (Takara, Japan) as per the manufacturer's instructions. To validate the reliability of gene expression obtained by RNA seq, a set of six DEGs in the leaf were randomly selected for qRT-PCR. Gene-specific primers were designed using GenScript Real-time PCR (TaqMan) Primer Design tool (<https://www.genscript.com/tools/real-time-pcr-tagman-primer-design-tool>).

The qRT-PCR reactions were carried out in triplicate using the SYBR Premix ExTaq™ II (Takara, Japan) and run-on Light Cycler 96 Real-Time PCR system (Roche, USA). Each reaction contains 5 µl of SYBR Green Master, 0.8 µl of template cDNA, 0.4 µl of each of the primers (10 µM), and 3.4 µl of nuclease-free water with a total volume of 10 µl. The qRT-PCR profile was as follows: 2 minutes at 95°C followed by 40 cycles of 10 seconds at 95°C, 30 seconds at 60°C with fluorescent amplification signal detection, and 30 seconds at 72°C. The melting curve was obtained by PCR following the last cycle: 15 seconds at 95°C followed by constant heating between 65°C for 15 seconds and 95°C for 2 seconds. The reference gene *18S ribosomal RNA (rRNA)* from maize was used as an endogenous (internal) control for normalization in all the qRT-PCR analyses. Three biological replicates were performed for each sample, and data were indicated as mean ± SE (n = 3). Cycle threshold (CT) difference between the reference *18S* gene and the target gene product was used to calculate the relative expression levels of the genes using the $2^{-\Delta\Delta Ct}$ method (Schmittgen and Livak, 2008).

Results

Transcript profiling under heat stress

A total of 141.03 and 152.86 million raw reads and 125.86 and 135.61 clean reads were obtained from the cDNA sample of LM 11 and CML 25, respectively. LM 11 leaf, pollen, and ovule had 56.88, 38.71, and 45.43 million raw reads, while 52.92, 44.85, and 55.09 million raw reads from leaf, pollen, and ovule, respectively, were obtained from CML 25. The reads with adapter contamination and low base quality ($\leq Q20$) were removed, and high-quality (HQ) clean reads were retained. The maximum number of raw and clean base reads were obtained in the LM 11 leaf, followed by the CML 25 ovule, CML 25 Leaf, LM 11 ovule, and other samples (Table 1). The overall

GC content ranged from 54.71 to 61.2%. A higher number of transcripts were observed in class 200–1000 bp. The principal component analysis (PCA) of the above-mentioned data (Supplementary Figure S1) suggests that the RNA-seq results meet the requirement of DEG identification. It is the first transcriptome library reported for heat stress in maize from three different tissues, which affects the synthesis of kernels as high temperature affects the reproductive stage of the plant (Jagtap et al., 2020).

Differentially expressed genes (DEGs) under heat stress

A total of 2,164 transcripts were differentially expressed between LM 11 and CML 25. 1151, 451, and 562 DEGs were specific in LM 11 versus CML 25 leaf, pollen, and ovule, respectively. A total of 1,095 (52.5%), 419 (20.1%), and 503 (24.1%) DEGs were uniquely expressed in LM 11 leaf versus CML 25 leaf, LM 11 pollen versus CML 25 pollen and LM 11 ovule versus CML 25 ovule, respectively (Figure 2A). Eleven (0.5%) and thirty-eight (1.8%) DEGs were common between leaves with pollen and ovule, respectively. 14 (0.7%) and 7 (0.3%) DEGs were similar between pollen and ovule and between all three samples, respectively (Figure 2A). Likewise, in LM11_leaf versus CML25_leaf, 578 DEGs were up-regulated, and 573 DEGs were down-regulated. Similarly, in LM11_pollen versus CML25_pollen, 231 DEGs were up-regulated, and 220 DEGs were down-regulated. While in LM11_ovule versus CML25_ovule, 318 and 244 DEGs were up-regulated and down-regulated, respectively (Figures 2B–D). In addition, a heat map of the overall expression pattern of the DEGs revealed that many unique DEGs were highly expressed in CML 25 leaf and LM 11 leaf samples compared to pollen and ovule samples (Supplementary Figure S2).

The highest up-regulated and down-regulated DEGs in the leaf, pollen, and ovule are summarized in Tables 2, 3. Nine pollen DEGs showed high expression encoding for a beta-expansion protein, endoglucanase, serine/threonine protein kinase, pectin esterase, beta-amylase, and acid phosphatase family proteins. One DEG with the highest expression level in the leaf is involved in cytochrome b6-f complex iron-sulfur subunit (Table 2). Likewise, out of the top 10 down-regulated DEGs, six were found in pollen and four in the leaf. The down-regulated DEGs in pollen encode endoglucanase, pectin esterase inhibitor, beta-amylase, glucan endo-1,3 beta-glucosidase, and pollen-specificity family proteins. Four highly down-regulated DEGs in the leaf are involved in dehydrin, calcium-dependent protein kinase, glycerate dehydrogenase, and eukaryotic translation initiation factor 3 subunit proteins (Table 3).

TABLE 1 Statistics of RNA-seq data obtained from three tissues of LM 11 and CML 25 inbreds under heat stress at reproductive phase.

| Inbred | Tissue | Raw reads | Clean reads | Raw bases (Gb) | Clean bases (Gb) | GC content (%) |
|--------|--------|-----------|-------------|----------------|------------------|----------------|
| LM 11 | leaf | 56883812 | 50600778 | 8.72 | 7.28 | 60.27 |
| | pollen | 38714402 | 34675722 | 5.97 | 4.99 | 56.82 |
| | ovule | 45436160 | 40586048 | 6.99 | 5.85 | 55.02 |
| CML 25 | leaf | 52920330 | 46307576 | 8.12 | 6.63 | 61.20 |
| | pollen | 44852338 | 39999254 | 6.89 | 5.76 | 57.08 |
| | ovule | 55090774 | 49309984 | 8.5 | 7.06 | 54.71 |

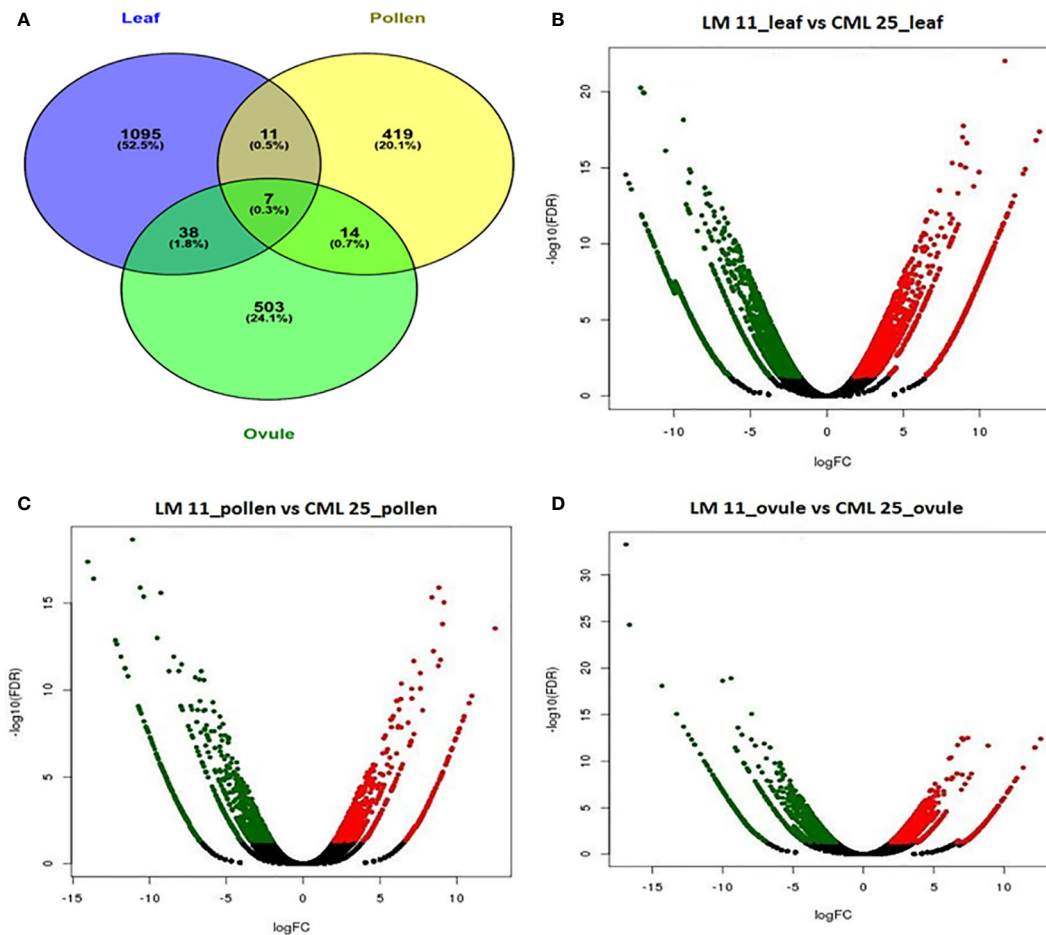


FIGURE 2

Differentially expressed genes (DEGs) involved in response to heat stress in maize at reproductive stage from three different tissues. **(A)** Venn diagram of DEGs in response to changes in leaf, pollen, and ovule. **(B)** Volcanic plot displaying DEGs in leaf of LM11 (S) versus CML 25 (T). **(C)** Volcanic plot displaying DEGs in pollen of LM 11 (S) versus CML 25 (T). **(D)** Volcanic plot displaying DEGs in ovule of LM 11 (S) versus CML 25 (T). Red dots represent up-regulated DEGs; green dots represent down-regulated DEGs. The x-axis values correspond to the \log_2 (fold change) value; the y-axis corresponds to the mean expression value of the $-\log_{10}$ (p-value) between LM 11 and CML 25. The Volcano plots for differentially expressed transcripts represent $FDR \leq 0.05$ and $|\log_2$ (fold change) ≥ 2 .

Gene ontology (GO) classification of DEGs

A total of 238 enriched GO terms were assigned to DEGs based on stringent p-value (0.001) and q-value (0.001) (Figure 3A) in LM 11 versus CML 25 leaf under heat stress conditions. These GO terms were further categorized into 132 (55.47%) biological process (BP), followed by 60 (25.21%) cellular component (CC) and 46 (19.32%) molecular function (MF) (Supplementary Table S1). While in LM 11 versus CML 25, pollen GO terms were categorized into 89 (48.90%) BP, followed by 47 (25.82%) CC and 46 (25.27%) MF (Supplementary Table S2). Correspondingly, a total of 216 enriched GO terms in ovule were assigned to DEGs and were classified into 114 (52.77%) BP, followed by 53 (24.53%) CC and 49 (22.68%) MF (Supplementary Table S3).

Exclusively in leaf, pollen, and ovule under the BP category, the term “metabolic process” (GO:0008152) and “cellular process” (GO:0009987) were commonly highly enriched components. The top two enriched terms under CC were “cell” (GO:0005623) and “cell part” (GO:0044464). Catalytic (GO:0003824) and binding activity (GO:0005488) were highly heightened under the MF category. Collectively, molecular function regulator (GO:0098772),

signal transducer activity (GO:0004871), and metabolic process (GO:0008152) were highly enriched in leaf, pollen, and ovule. Exclusively, the metabolic process (GO:0008152) was prevalent in leaf, pollen, and ovule (Figure 3B).

Metabolic pathways enrichment analysis

KEGG pathway analysis of 2164 DEGs from all the tissues revealed that 264 and 146 genes were involved in the metabolic overview pathway and biosynthesis of secondary metabolites, respectively. Apart from them, carbon metabolism, starch and sucrose metabolism, biosynthesis of amino acids, protein processing in the endoplasmic reticulum, plant hormone signal transduction, plant-pathogen interactions, the amino sugar, and nucleotide sugar metabolism, and arginine & proline metabolism were significantly involved in response to heat stress (Table 4). MapMan displayed the involvement of nearly all the pathways concerning DEGs in the leaf. Among the key pathways, proteolysis had the highest number of BINs (65), followed by signaling (57), secondary metabolite (55), and abiotic stresses (34), which

TABLE 2 Top 10 up-regulated DEGs in leaf, pollen, and ovule in comparisons of LM 11 (S) and CML 25 (T) inbreds under heat stress conditions.

| Transcript Id | Tissue | Log ₂ Fold Change | P-value | FDR | Transcript description | Total GO term | GO IDs |
|-------------------|--------|------------------------------|----------|----------|--|---------------|--|
| comp84863_c0_seq1 | leaf | 9.66 | 5.14E-24 | 8.42E-20 | cytochrome b6-f complex iron-sulfur subunit | 8 | F:GO:0009496; C:GO:0009535; P:GO:0015979; C:GO:0016021; P:GO:0022900; F:GO:0045158; F:GO:0046872; F:GO:0051537 |
| comp80999_c0_seq5 | pollen | 9.43 | 5.78E-23 | 9.04E-20 | beta-expansin 1a precursor | 5 | C:GO:0005576; C:GO:0005618; C:GO:0016020; P:GO:0019953; P:GO:0071555 |
| comp74578_c0_seq2 | pollen | 9.20 | 1.15E-22 | 1.61E-19 | endoglucanase 8 | 4 | C:GO:0005576; F:GO:0008810; P:GO:0030245; P:GO:0071555 |
| comp80312_c0_seq1 | pollen | 9.03 | 9.48E-22 | 7.85E-19 | serine/threonine-protein kinase At5g01020-like | 5 | F:GO:0004675; F:GO:0005524; C:GO:0005886; P:GO:0006468; P:GO:0007178 |
| comp73316_c0_seq1 | pollen | 9.00 | 4.97E-22 | 5.00E-19 | pollen allergen Phl p 11 | 1 | C:GO:0005615 |
| comp66485_c0_seq2 | pollen | 8.91 | 3.76E-22 | 4.08E-19 | Pectinesterase family protein, expressed | 5 | C:GO:0005618; F:GO:0030599; P:GO:0042545; F:GO:0045330; P:GO:0045490 |
| comp72137_c0_seq1 | pollen | 8.90 | 7.23E-22 | 6.37E-19 | pectinesterase inhibitor domain containing protein | 4 | F:GO:0030599; P:GO:0043086; F:GO:0046910; C:GO:0071944 |
| comp80939_c0_seq2 | pollen | 8.59 | 4.80E-21 | 3.38E-18 | beta-amylase | 4 | P:GO:0000272; C:GO:0009507; F:GO:0016161; F:GO:0102229 |
| comp78238_c0_seq1 | pollen | 8.45 | 3.51E-20 | 2.35E-17 | acid phosphatase 1 precursor | 2 | F:GO:0003993; P:GO:0016311 |
| comp83327_c1_seq5 | pollen | 8.34 | 2.35E-19 | 1.33E-16 | predicted protein | 1 | F:GO:0005509 |

TABLE 3 Top 10 down-regulated DEGs in leaf, pollen, and ovule in comparisons of LM 11 and CML 25 inbreds under heat stress conditions.

| Transcript Id | Tissue | Log ₂ Fold Change | P-value | FDR | Transcript description | Total GO term | GO IDs |
|--------------------|--------|------------------------------|----------|----------|---|---------------|--|
| comp74586_c0_seq2 | pollen | -10.53 | 6.32E-26 | 2.22E-22 | endoglucanase 8 | 4 | C:GO:0005576; F:GO:0008810; P:GO:0030245; P:GO:0071555 |
| comp72137_c0_seq3 | pollen | -10.33 | 1.09E-25 | 3.07E-22 | pectinesterase inhibitor domain containing protein | 4 | F:GO:0030599; P:GO:0043086; F:GO:0046910; C:GO:0071944 |
| comp77346_c1_seq10 | leaf | -9.95 | 1.64E-24 | 5.37E-20 | dehydrin COR410 | 3 | P:GO:0006950; P:GO:0009415; F:GO:0046872 |
| comp80939_c4_seq2 | pollen | -9.62 | 6.87E-24 | 1.38E-20 | beta-amylase 1, chloroplastic-like | 4 | P:GO:0000272; C:GO:0009507; F:GO:0016161; F:GO:0102229 |
| comp63912_c0_seq2 | leaf | -8.38 | 6.50E-20 | 3.55E-16 | calcium-dependent protein kinase, isoform AK1 | 9 | F:GO:0004683; F:GO:0005516; F:GO:0005524; C:GO:0005634; C:GO:0005737; F:GO:0009931; P:GO:0018105; P:GO:0035556; P:GO:0046777 |
| comp79197_c3_seq1 | pollen | -7.77 | 3.20E-18 | 1.45E-15 | glucan endo-1,3-beta-glucosidase 8-like | 5 | F:GO:0004553; P:GO:0005975; C:GO:0016021; F:GO:0030247; C:GO:0046658 |
| comp78238_c0_seq3 | pollen | -7.75 | 3.78E-18 | 1.66E-15 | acid phosphatase 1 precursor | 2 | F:GO:0003993; P:GO:0016311 |
| comp72119_c1_seq2 | leaf | -7.37 | 5.24E-17 | 1.15E-13 | glycerate dehydrogenase | 5 | C:GO:0005829; F:GO:0016618; F:GO:0030267; F:GO:0051287; P:GO:0055114 |
| comp84091_c0_seq1 | leaf | -7.34 | 4.67E-17 | 1.09E-13 | eukaryotic translation initiation factor 3 subunit I-like | 5 | P:GO:0001732; F:GO:0003743; C:GO:0005852; C:GO:0016282; C:GO:0033290 |
| comp63821_c1_seq1 | pollen | -7.32 | 4.22E-17 | 1.49E-14 | pollen-specific protein C13-like | 1 | C:GO:0005615 |

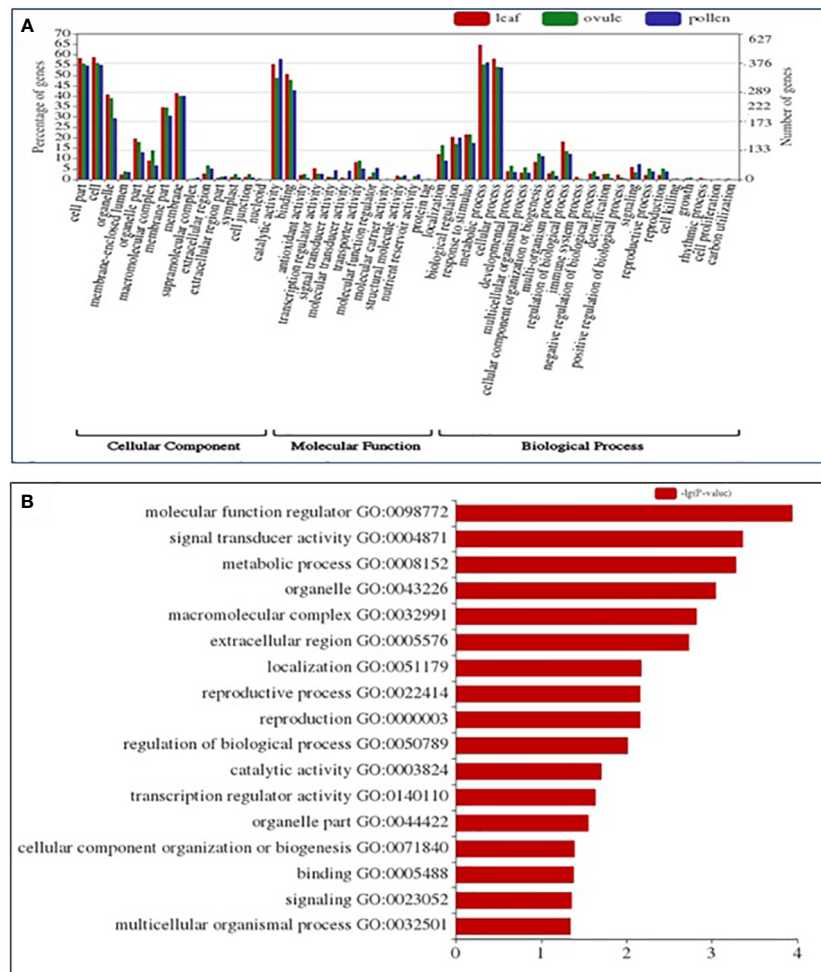


FIGURE 3

GO and KEGG enrichment analysis of DEGs identified in the comparison of leaf, pollen, and ovule of LM 11 (S) and CML 25 (T) under heat stress. **(A)** GO analysis categorized DEGs into Cellular Component (CC), Molecular Function (MF), and Biological Process (BP). The right Y-axis indicates the number of genes in GO categories. The left Y-axis indicates the percentage of a specific category of genes in that main category. The X-axis defines GO terms. **(B)** Abundant KEGG enrichment assigned to DEGs in leaf, pollen, and ovule, collectively under heat stress.

represented the functional categories to which genes were assigned (Figure 4A). Similarly, in pollen, the cell wall had the highest number of BINs (34), followed by signaling (28), proteolysis (27), and abiotic stresses (11) (Figure 4B). Furthermore, in the ovule, proteolysis had the highest number of BINs (47), followed by abiotic stresses (20), secondary metabolite (19), redox state (11), and cell wall (Figure 4C).

Alterations in the expression of transcription factors (TFs)

A total of 24 classes of TFs were identified from the 55 differentially expressed transcripts of the leaf. Most TFs were attributed to MYB (Myb DNA-binding, Myb_CC_LHEQLE, Myb_DNA-bind_4, Myb_DNA-bind_6, and bZIP followed by WRKY, AP2, and PsbP. Most of the transcripts for Apetala 2 (AP2) were up-regulated in the leaf, and among all, one transcript was highly expressed with fold change of 5.22 (up-regulated) followed by WRKY (4.74) (Figure 5A). A total of four transcripts for PsBP were up-regulated. Most of the MYB transcripts and nine transcripts for bZIP

were down-regulated (Supplementary Table S4). Similarly, eight classes of TF families viz. NAM, WRKY, zf-C2H2_6, ArfGap-C2, Myb_DNA-bind_6, zf-C3HC4_3, Exo70, and AP2 were attributed in comparisons of LM 11 versus CML 25 pollen. All differentially expressed TFs in pollen were down-regulated except WRKY, Exo70, and AP2 (Supplementary Table S5). Likewise, 11 classes of TF families viz. Asp, SapB_2, AP2, DELLA, GRAS, YABBY, bZIP, Cpn60_TCP1, WRKY, Myb_DNA-binding, and Arf were ascribed comparative to LM 11 versus CML 25 ovule. AP2, DELLA, GRAS, YABBY, Myb_DNA-binding, and Arf were up-regulated (fold changes ranging from 3.37 to 4.46), whereas, Asp, SapB_2, bZIP_1, and WRKY were down-regulated (fold changes varying from -3.31 to -5.44) (Figure 5B; Supplementary Table S6).

Induction of metabolic processes genes

A total of 21 and 34 DEGs were up-regulated and down-regulated, respectively, in comparisons of LM 11 versus CML 25 leaf. The up-regulated DEGs are involved in sugar transmembrane

TABLE 4 Top twenty KEGG pathway enrichment of the DEGs of leaf, pollen and ovule under heat stress conditions.

| Term | ID | Input number | Background number | P-Value | Corrected P-Value |
|---|----------|--------------|-------------------|-------------|-------------------|
| Metabolic pathways | zma01100 | 264 | 2609 | 4.68E-39 | 5.20E-37 |
| Biosynthesis of secondary metabolites | zma01110 | 146 | 1481 | 1.04E-20 | 5.75E-19 |
| Carbon metabolism | zma01200 | 33 | 372 | 6.61E-05 | 0.000458578 |
| Starch and sucrose metabolism | zma00500 | 31 | 251 | 2.56E-07 | 7.10E-06 |
| Biosynthesis of amino acids | zma01230 | 28 | 371 | 0.002236704 | 0.008866932 |
| Protein processing in endoplasmic reticulum | zma04141 | 26 | 316 | 0.001038062 | 0.004608996 |
| Plant hormone signal transduction | zma04075 | 26 | 355 | 0.004522581 | 0.015687704 |
| Plant-pathogen interaction | zma04626 | 25 | 209 | 5.90E-06 | 7.27E-05 |
| Amino sugar and nucleotide sugar metabolism | zma00520 | 24 | 186 | 2.80E-06 | 4.10E-05 |
| Arginine and proline metabolism | zma00330 | 22 | 78 | 3.56E-11 | 1.32E-09 |
| Oxidative phosphorylation | zma00190 | 22 | 170 | 6.85E-06 | 7.60E-05 |
| Ribosome | zma03010 | 22 | 504 | 0.399824335 | 0.558433653 |
| Phenylpropanoid biosynthesis | zma00940 | 21 | 242 | 0.001661749 | 0.006831636 |
| Glycolysis/Gluconeogenesis | zma00010 | 20 | 193 | 0.000282769 | 0.001426698 |
| Spliceosome | zma03040 | 19 | 269 | 0.018878607 | 0.053731419 |
| Glutathione metabolism | zma00480 | 18 | 139 | 4.50E-05 | 0.000393884 |
| Endocytosis | zma04144 | 18 | 270 | 0.0347085 | 0.087560079 |
| Glyoxylate and dicarboxylate metabolism | zma00630 | 15 | 110 | 0.00011265 | 0.000735541 |
| Cysteine and methionine metabolism | zma00270 | 15 | 137 | 0.000936044 | 0.004329204 |
| Pentose and glucuronate interconversions | zma00040 | 14 | 76 | 1.01E-05 | 0.000102308 |

transporter activity, response to reactive oxygen species (ROS), abscisic acid (ABA), oxidative stress, photosynthesis (photosystem I and chloroplast thylakoid membrane), and cytochrome-c peroxidase activity ([Supplementary Table S7](#)). The down-regulated DEGs have roles in the oxidation-reduction process, hydrolase and glyoxylate reductase (NADP) activity, transmembrane transport, and response to light stimulus ([Supplementary Table S8](#)). Twenty-three and 19 DEGs were up-regulated and down-regulated in comparisons of LM 11 versus CML 25 pollen. Up-regulated DEGs are associated with the catabolic process, cellular response to nitrogen and phosphate starvation, mitochondrial functions, and cellular response to oxidative stress (ROS) ([Supplementary Table S9](#)). One of the up-regulated DEGs (comp86232_c0_seq4) regulates inflorescence development, photoperiodism, and flowering. Pollen-specific NTP303 precursor protein (comp87307_c0_seq7) is involved in the oxidation-reduction process. Down-regulated DEGs belong to the zinc finger domain family protein, beta-glucosidase, alkaline/neutral invertase, vacuolar membrane, and microtubule-associated protein ([Supplementary Table S10](#)). One of the down-regulated DEG (comp70405_c0_seq2) was the ZIM motif family protein involved in flower development.

A total of 50 and 45 DEGs were up-regulated and down-regulated, respectively, in comparisons of LM 11 versus CML 25 ovule. Among the up-regulated transcripts, comp81007_c0_seq9, comp67757_c0_seq2, comp44452_c0_seq1, comp61084_c0_seq2, and comp13987_c0_seq1 were ovule specific in expressions as well as related to abiotic stresses.

One of the DEG (comp81007_c0_seq9) belongs to lipoxygenase involved in the seed germination processes. Also, corresponding four DEGs *viz.*, comp67757_c0_seq2, comp44452_c0_seq1, comp61084_c0_seq2, and comp13987_c0_seq are related to oleosin, dehydrin, membrane protein At3g27390, and putative late embryogenesis abundant protein, respectively. These four up-regulated DEGs are involved in lipid storage, response to water deprivation, vegetative to the reproductive phase transition of the meristem, and positive regulation of response to water deprivation ([Supplementary Table S11](#)). Two down-regulated transcripts, comp86012_c3_seq2 and comp83060_c3_seq3, are associated with glyceraldehyde-3-phosphate dehydrogenase (NAD⁺) (phosphorylating) activity. Another DEG, comp86139_c1_seq7, related to beta-expansin (EXPB7), has roles in the sexual reproduction process. Similarly, DEG comp83923_c2_seq9 was histone H4 specific and is regulated in response to water stress ([Supplementary Table S12](#)).

Identification of expressed HSPs

A total of fourteen, seven, and four DEGs related to HSPs were identified in leaf, pollen, and ovule, respectively. Six DEGs were highly expressed in the leaf, of which only one was up-regulated (HSP20), and five were down-regulated. In pollen, three HSP genes were up-regulated, two encoding for HSP20 and the other encoding 101 kDa heat shock protein (Hsp101/ClpB). One down-regulated DEG in pollen related to Rieske Iron-Sulfur Protein was associated with

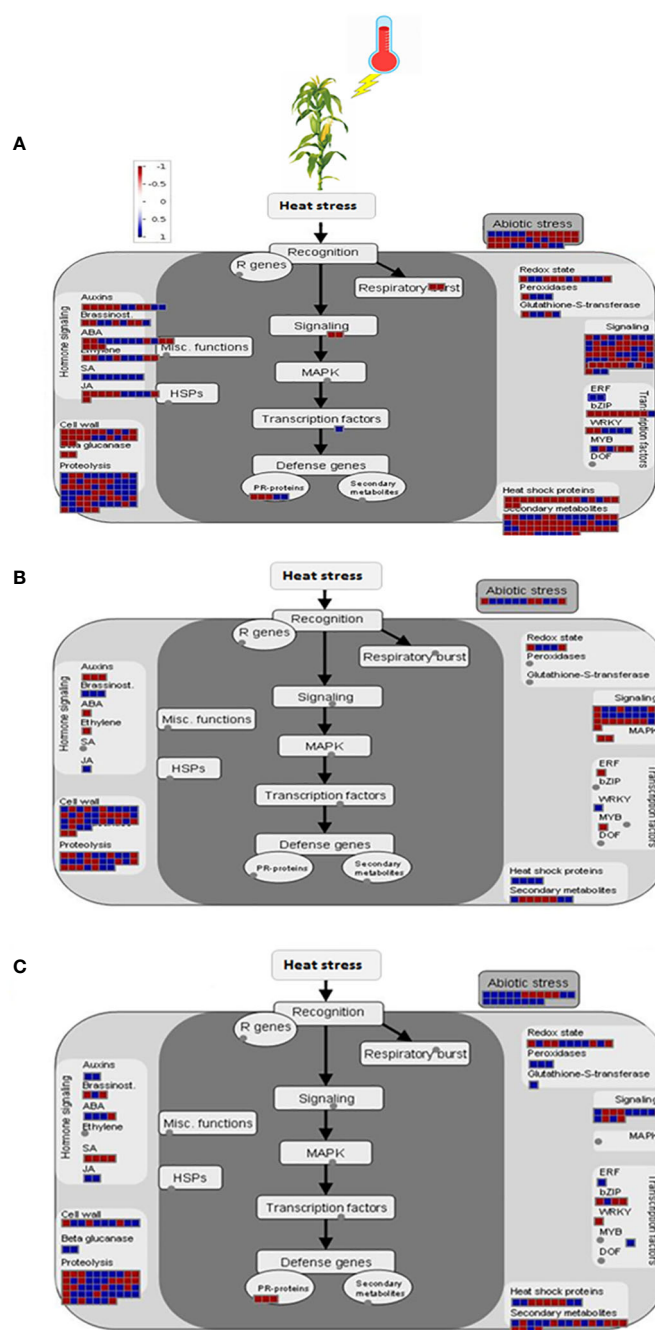


FIGURE 4
MapMan display for heat responsive genes in leaf (A), pollen (B), and ovule (C). Biological functions of overall heat responsive genes at FDR < 0.05 and $|\log_2(\text{fold change})| \geq 2$ under heat stress conditions. Colors represent \log_2 fold changes higher (blue) for up-regulated and lower (red) for down-regulated than 0.

HSP70 (Figure 5C). Likewise, four genes for HSP20 (2 genes) and HSP70 (2 genes) were up-regulated, whereas three genes encoding for HSP20 were down-regulated in the ovule.

Hormone biosynthesis and signal transduction-related genes

A total of 12 and 20 DEGs related to hormone biosynthesis were up-regulated and down-regulated, respectively, in comparisons of LM

11 versus CML 25 leaf. The highest up-regulated DEGs are dehydrin COR410, followed by terpene synthase 7, pyrophosphate synthase, ZIM motif family protein, and catalase. These are actively involved in the cellular response to a salicylic acid stimulus, nitric oxide and abscisic acid (ABA), regulation of jasmonic acid (JA) mediated signaling pathway and response to auxin. In addition, several other up-regulated genes belong to DNA binding, transcription factor activity, polyprenol biosynthetic process, terpene synthase activity, and response to water deprivation (Supplementary Table S13). Likewise, down-regulated DEGs in the leaf are related to jasmonic

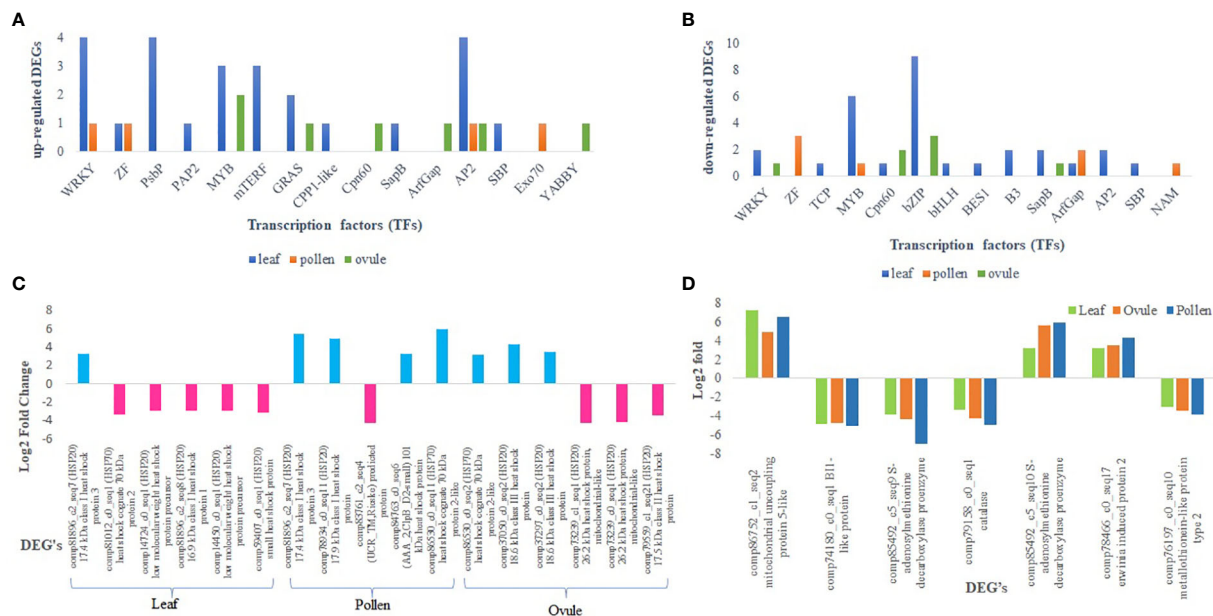


FIGURE 5

Differential expression of representative genes involved in molecular function in response to heat stress. (A) Transcription factors those are commonly up-regulated in leaf and ovule, and leaf and pollen. (B) Transcription factors those are commonly down-regulated in leaf and ovule, and leaf and pollen. (C) Relative expression of heat shock proteins (HSPs) in leaf, pollen, and ovule. The blue color indicates the DEGs are up-regulated and the pink color indicates DEGs are down-regulated. (D) Graphical representation of commonly expressed DEGs in leaf, pollen and ovule that are involved in the polyamine biosynthesis pathway.

acid (JA), abscisic acid (ABA) synthesis, auxin-activated signaling pathway, response to water deprivation, calcium-mediated signaling, tricarboxylic acid cycle (TCA), and salicylic acid-mediated signaling pathways. Three calcium-dependent protein kinases were up-regulated in LM 11 compared to CML 25 pollen and are involved in DNA binding transcription factor activity, calmodulin-dependent protein kinase activity, calcium ion binding, calmodulin binding, ATP binding, abscisic acid-activated signaling pathway (Supplementary Table S14). One DEG encodes for late embryogenesis abundant protein that was highly up-regulated (fold change; 8.26) and involved in stress-related responses. Furthermore, seven DEGs were down-regulated in LM 11 pollen associated with zeaxanthin epoxidase, probable indole-3-acetic acid (IAA)-amido synthetase GH3.8, calcium-dependent protein kinase 34-like, ZIM motif family protein, gibberellin 2-oxidase, phospho-2-dihydro-3-deoxyheptonate aldolase, and HlyIII predicted protein (Supplementary Table S14). Similarly, 19 up-regulated DEGs in the ovule of LM 11 versus CML 25 are associated with seed-specific expressions, and one down-regulated DEG belongs to catalase and is related to the signaling pathway (Supplementary Table S15).

Differential expression of polyamines biosynthesis pathway related genes

Seven polyamines biosynthesis pathway-related transcripts were differentially expressed in leaf, pollen, and ovule. Three of the genes were up-regulated in all three tissues that have roles in mitochondrial transport, adenosylmethionine decarboxylase activity, and spermine biosynthetic process. Four down-regulated genes are involved in cell-

organelle-related responses under heat stress and metal ion binding (Figure 5D).

qRT-PCR validation of DEGs

The RNAseq results were validated by selecting six genes randomly and primers were designed (Supplementary Table S16). The ratio of comparative expression level found between LM 11 and CML 25 was calculated and the log2 fold changes were compared with the result of RNA-seq data (Figure 6). The qRT-PCR data showed a significant correlation ($R^2 = 0.7961$) with RNA-seq data, which supported the authenticity of expression patterns revealed by RNA-seq (Supplementary Table S17).

Discussion

Transcriptome analysis

Heat stress (HS) due to an increase in global mean temperature as a result of climate change can remarkably suppress plant growth and development. It has been reported as one of the most critical causes of yield reduction and dry matter production in many crops, including maize (Lobell et al., 2008; Hasanuzzaman et al., 2013; Qian et al., 2019). It is negatively associated with anthesis, silking, and grain-filling reproductive stages (Noor et al., 2019; Longmei et al., 2021). LM 11 had a higher canopy temperature (40.13°C) than CML 25 (38.38°C) as the heat susceptible have high canopy temperature as observed in wheat and cotton genotypes (Carmo-Silva et al., 2012).

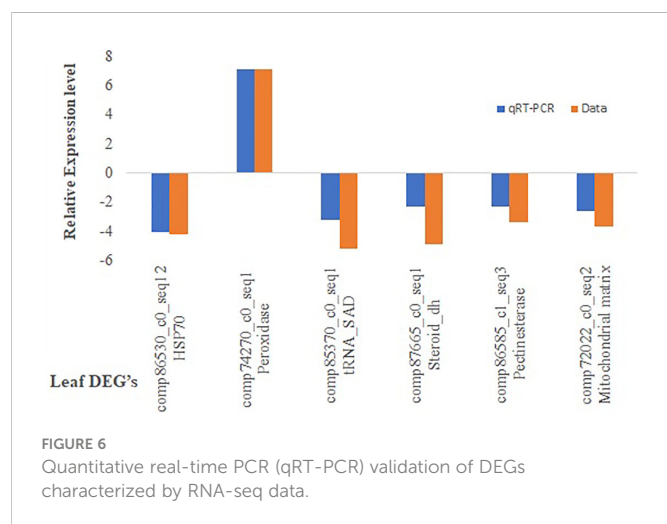


FIGURE 6
Quantitative real-time PCR (qRT-PCR) validation of DEGs characterized by RNA-seq data.

The reproductive stage in plants is sensitive to heat stress, and pollen viability is directly affected by heat stress. Reduced pollen viability was observed in LM 11 (40%) under heat stress compared to CML 25 (84.04%). The impact of heat stress was more significant in LM 11 *per se* to kernel number per ear (Jagtap, 2020). Thus, there is a need to dissect physiological and molecular mechanisms underlying heat stress responses and adaptation at the reproductive stage in maize. In the past decade, the RNA-seq approach has been widely used for revealing molecular mechanisms of heat stress responses at the seedling stage (Frey et al., 2015; Shi et al., 2017; Kim et al., 2019; Li and Ye, 2022; Xuhui et al., 2022).

In the present study, comparative transcriptomics was performed for leaf, pollen, and an ovule of heat-tolerant inbred CML 25, and susceptible inbred LM 11 was analyzed under prolonged heat stress at 42°C. A total of 1127 up-regulated and 1037 down-regulated DEGs were identified in response to heat stress. The higher number of up-regulated genes indicates the cumulative activation of defense responses upon heat stress. Similarly, in heat transcriptomic studies conducted in wheat, rice, and maize, the number of induced genes was 3–6 times more than the number of repressed genes (Fernandes et al., 2008; Qin et al., 2008; Zhang et al., 2013). In the present study, leaf tissue under heat stress showed maximum DEGs (1151) followed by ovule (562) and pollen (451). The higher number of DEGs in the flag leaf tissue may be attributed to its higher metabolic activity under heat stress, as it is responsible for assimilating synthesis and translocation to the developing organs (Li et al., 2017; Shi et al., 2017; Qian et al., 2019).

In our study, the most enriched GO are biological processes comprising metabolic processes, cellular processes, and regulation. Among the molecular function category, catalytic activity and binding were mainly dominant. Also, Wang et al. (2016) categorized the enriched term “metabolic process” as a top GO term in comparative transcriptomics of Chinese cabbage to reveal the heat-responsive genes. Rahmati Ishka et al. (2018) showed that under the biological process, more than two-thirds of the over-represented genes belong to just two categories metabolic and cellular processes. It can be envisaged that crosstalk might exist among different pathways involved in various abiotic stresses (Singh et al., 2019). Thus, transcriptional reprogramming might have a role in heat stress tolerance.

Transcription factors triggers regulatory networks

Transcription factors play fundamental roles in biotic and abiotic stresses and are considered frontline defenders. Other studies suggested that most of the identified DEGs under heat stress in crop plants encode members of the ERF, MYB, bZIP, bHLH, WRKY, NAC, and MYB-related TF families (Wang et al., 2016; Li et al., 2017; Zhao et al., 2017; Qian et al., 2019; Li and Ye, 2022) (Table 5). In the present study, we found several TFs expressed under heat stress in both cultivars like MYB, bZIP, WRKY, AP2, NAM, DELLA, etc., and might have a role in overcoming the heat stress in CML 25. A fraction of TFs, including 9 MYB and bHLH, 6 WRKY, 3 Asp, and 1 bZIP, were expressed in the leaf, while 3 bHLH and each one of MYB, WRKY, and Asp were regulated in the ovule. Above all, DELLA (LOC100280169) and Aspartic peroxidase (LOC100127531) genes are commonly up and down-regulated in the leaves and ovules of maize.

Contrastingly, we observed tissue-specific differential expression of a few TFs, SBP, NAM, Exo70, and YABBY, which are mainly expressed at reproductive stages. These TFs were specific to leaf, pollen, or ovule. In our study, SBP was exclusively expressed in the leaf, which plays a critical role in many biological processes, notably flower development (Li et al., 2019). Also, we found that NAM and Exo70 TFs were expressed only in pollen. The NAM is associated with pollen and tapetum development in biotic and abiotic stress responses (Yang et al., 2018). The Exo70 plays a vital role in the exocyst complex function for pollen development, pollen grain germination, and pollen tube elongation (Synek et al., 2017). We also reported that the YABBY TF (a subfamily of ZF TF) was up-regulated in the ovule. It is basically involved in lateral organ development, dorsoventral polarity, and abiotic stress responses (Zhang et al., 2019).

Shifts in energy metabolism and signal transduction plays crucial role in heat stress tolerance

Heat stress prompts the formation of reactive nitrogen species (RNS like NO) and reactive oxygen species (ROS), such as OH[•], H₂O₂, and O₂^{•−}, resulting in increased electrolyte leakage and lipid peroxidation. It led to enhance activities of antioxidant enzymes. Among them, the superoxide radical (O₂^{•−}) is dismuted by superoxide dismutase (SOD) into H₂O₂ and further scavenged by catalase (CAT) and peroxidases (such as POD) through converting into H₂O (Alam et al., 2017). Various studies reported that the cytochromes P450-related genes were expressed differentially under heat stress conditions in response to increased concentrations of ROS (Frey et al., 2015; Wang et al., 2016). In the present study, many oxidative metabolism-related genes like cytochromes P450 were up-regulated, confirming that ROS are generated during heat stress.

Heat stress affects plant photosynthesis negatively by inactivating photosystems, PSII, and PSI (Mathur et al., 2014). Photochemical reactions in thylakoid lamellae and carbon metabolism in chloroplast stroma have been noticed as the injury sites under heat stress (Wahid et al., 2007). DEGs related to PSII and PSI were up-regulated during heat stress in the present study. Several studies reported differential expression of genes encoding photosynthetic electron transfer-

TABLE 5 Candidate genes conferring heat stress tolerance identified in previous studies in maize.

| Stage | Tissue | Candidate genes | Reference |
|--|---|---|--------------------|
| 21 days old seedlings | Leaves | 607 heat responsive genes, 39 heat tolerance genes | Frey et al. (2015) |
| 6-day old seedlings | Leaves | Transcription factors (ERFs,NACs,ARF,MYB, andHD-ZIP) | Li et al. (2017) |
| Three-week-old seedlings | Leaves | Secondary metabolite biosynthetic pathway genes | Shi et al. (2017) |
| V3 and R1 | Leaves, stalks, roots, tassels, ears, and silks | Pentatricopeptide repeat (PPR) proteins, miR168 and miR528 | He et al. (2019) |
| Three-week-old seedlings (five-leaf stage) | Leaves | Heat shock proteins (Hsp40, Hsp70, Hsp90, Hsp100, and small Hsps) and Transcription factors (AP2-EREBP, MYB, bHLH, b-ZIP, and WRKY) | Qian et al. (2019) |
| V3 | Leaves | Spliceosome metabolic pathways | Zhao et al. (2019) |
| V3 | Leaves | Protein Kinases (35 CDPK, 9 MAPK, 53 serine/threonine-protein kinase), 6 aquaporin, 3 anion channel protein, 41 MFS transporter, 168 TF, 94 HSPs, 1 prolinerich protein, 25 POX, 28 E3 ubiquitin-protein ligase, 24 ubiquitin-conjugating enzyme E2, 12 auxin-responsive protein IAA, and 3 abscisic acid receptor PYR/PYL family | Wu et al. (2020) |
| 2-5-day-old seedlings | Etiolated coleoptiles | Protein renaturation, Biomembrane repair, Osmotic adjustment, and Redox balance | Li and Ye (2022) |
| R1 | Flag leaf, pollen, ovule | Transcription Factors (AP2, MYB, WRKY, PsbP, bZIP, and NAM), heat shock proteins (HSP20, HSP70, and HSP101/ClpB), photosynthesis genes(PsaD&PsaN), antioxidation (APX and CAT) and polyamines (Spd and Spm). | Present study |

Cytb6/f, and PSII subunits, PsaD and PsaN, in response to heat stress (Wang et al., 2016; Wang et al., 2017). Similarly, dehydrins (DHNs) are a family of plant proteins induced in response to abiotic stresses or during later stages of embryogenesis (Close, 1997). The DHNs are highly hydrophilic and thermostable and act as chaperones that bind to calcium and help in water storage to impede cells from excessive dehydration (Hara et al., 2001; Yu et al., 2018). In the present investigation, DHNs were up-regulated in the ovule in response to heat stress. Tissue-specific expression and functional role of dehydrins per heat tolerance of C4 sugarcane (*Saccharum officinarum*) were confirmed by Galani et al. (2013). Several studies showed the role of dehydrins in abiotic stress tolerance in plants by stabilizing membranes, enzymes, and nucleotides in cells (Yu et al., 2018; Liu et al., 2019). The activation of genes involved in photosystem, antioxidants, and lipid peroxidation could be related to the maintenance of membrane integrity to confer heat tolerance.

Role of hormones in signaling cascades of heat stress

Under heat stress, hormone homeostasis is altered, including hormone stability, biosynthesis, total contents, and compartmentalization (Davies, 2004). Although the involvement of hormones in plant heat resilience is complex, the signal pathway of hormones is not yet elucidated under heat

stress. Many studies have delineated that optimizing certain hormones can enhance heat resilience in plants (Kotak et al., 2007; Wahid et al., 2007). Li et al. (2015) revealed that hormones such as ABA, auxin, jasmonic acid (JA), cytokinins (CKs), ethylene, gibberellin, and brassinosteroid are likely to be involved in heat stress tolerance. Several hormones, including ABA, brassinosteroids (BRs), and ethylene, possibly interacted through complex networks to regulate heat stress responses (Qu et al., 2013). In the present study, we found that DEGs related to P450 were up-regulated that involved in brassinosteroid biosynthesis. Contrastingly, DEG encoding remorin was down-regulated and negatively regulated the brassinosteroid mediated signaling pathway. Likewise, the ZIM motif family protein encoding was differentially expressed for pollen development under heat stress. JASMONATE ZIM-domain (JAZ) subfamily proteins have a role in biological processes such as development, stress-related, and hormone responses in *Arabidopsis*, rice, chickpea, and grape (Saha et al., 2016). Interestingly, both up-and down-regulated genes involved in hormone response pathways were identified, indicating that genes might help these pathways to keep the homeostasis under heat stress.

Induction of HSPs under heat stress-a natural phenomenon

HSP genes such as HSP70, HSP101/ClpB, and HSP20 were induced during heat stress. HSP20 was up-regulated in all three

tissues, whereas HSP101/ClpB and HSP70 were up-regulated in pollen and ovule, respectively. Extensive studies have demonstrated the protection of HSP70, HSP101, HSP20, and sHSPs family proteins from heat stress (Wang et al., 2016; Wang et al., 2017; Guo et al., 2019; Raju et al., 2020). Clp proteins belong to the large AAA+ (ATPases associated with diverse cellular activities) superfamily proteins. Furthermore, several studies illustrated that hydrogen peroxide (H₂O₂) generated during heat stress could enhance the ABA-dependent expression of HSP70 and sHSPs to tolerate heat stress (Liu et al., 2013; Li et al., 2014; Zhang et al., 2015). Therefore, the higher levels of expression are to acclimatize heat stress. It indicates that HSPs act as a molecular chaperone that prevents protein aggregation and denaturation resulting in maintaining protein structure. Further, heat stress induces membrane fluidity which activates lipid signaling and affects the Ca²⁺ channels and antioxidants. Therefore, increased lipid saturation is an important aspect of maintaining membrane fluidity. These cascades of reactions act as a primary signal for the activation of heat stress tolerance.

Synthesis of polyamines

S-adenosylmethionine decarboxylase (SAMDC) is a key enzyme controlling the rate of polyamines formation that plays a pivotal role in plant growth, development, and adaptation to abiotic stresses (Mellidou et al., 2016). In the present investigation, the gene encoding for SAMDC showed up and down-regulation in all three tissues. Heat stress at 38°C and above suppressed the SAMDC activity, resulting in impaired polyamine biosynthesis and inhibition of pollen germination (Chen et al., 2015). It may be likely the cause of reduced pollen viability in heat-susceptible inbred LM 11 under high temperatures in the present study. The overexpression of the SAMDC gene resulted in elevated levels of Spd and/or Spm and enhanced the plant tolerance to abiotic stresses (Chen et al., 2015). Song et al. (2002) reported that suppression of SAMDC activity is a significant cause of inhibition of pollen germination and tube growth in tomatoes at high-temperature. Further, the essential role of Spd and Spm in pollen viability and seed development has been studied by Chen et al. (2015).

The present study ascertained that the metabolic overview pathway and secondary metabolites biosynthesis pathway, with the involvement of 264 and 146 genes, respectively, were the top two metabolic pathways in response to heat stress. Also, carbon metabolism, starch, sucrose metabolism, biosynthesis of amino acids, and proteins processing in the endoplasmic reticulum and plant hormone signal transduction are the most enriched pathways under heat stress. Shi et al. (2017) reported the expression of proteins under heat stress in sweet maize involved in a series of biological processes from translation to metabolic pathways and secondary metabolite synthesis. During heat stress, the seed response recorded in *Brassica napus* stated that heat treatment specifically affects the pathways, including ribosome, biosynthesis of amino acids, starch and sucrose metabolism, and protein

processing in the endoplasmic reticulum and carbon metabolism (Guo et al., 2019). The upregulation of the genes involved in these pathways indicates that a set of genes are regulated to minimize protein modification during heat stress and impart tolerance in CML 25. The detected candidate genes could be exploited in maize heat-resilience breeding programs.

Conclusion

The comparative transcriptomic studies in CML 25 and LM 11 maize inbreds under heat stress at the reproductive stage from three tissues (leaf, pollen, and ovule) gave insights into tissue-specific stress-related genes involved in biological pathways like metabolic processes, secondary metabolite synthesis, starch, and sucrose metabolism, carbon metabolism, protein synthesis, etc. Most DEGs were transcription factors, heat shock proteins, antioxidants, hormone biosynthesis, and polyamine biosynthesis-related genes. Seven DEGs were common in leaf, pollen, and ovule; and involved in the polyamines biosynthesis pathway that could be further explored in understanding heat tolerance mechanism. The up-regulated genes identified in heat-tolerant inbred CML 25 would be the potential candidate genes that could be utilized for the development of heat-resilient maize using marker-assisted backcross breeding.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRJNA656908.

Author contributions

Conceptualization: YV and GJ; Project administration and resources: YV; Methodology: AJ, YV, and IY; Bioinformatics analysis: AJ and IY; Methodology: IY and AJ; Validation: NK and AS; Original draft: AJ, UP, and YV; Finalized the manuscript: AJ, YV, IY, and GJ. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are thankful to the Science and Engineering Research Board (SERB), India, Department of Science and Technology (DST), India and Confederation of Indian Industry (CII), India for granting the Prime Minister's Fellowship for Doctoral Research to AJ. We thank the Science and Engineering Research Board (SERB), India for granting one year overseas visiting doctoral fellowship (OVDF) to AJ.

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.

References

- Alam, M. A., Seetharam, K., Zaidi, P. H., Dinesh, A., Vinayan, M. T., and Nath, U. K. (2017). Dissecting heat stress tolerance in tropical maize (*Zea mays* L.). *Field Crop Res.* 204, 110–119. doi: 10.1016/j.fcr.2017.05.016
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Asthir, B. (2015). Protective mechanisms of heat tolerance in crop plants. *J. Plant Interact.* 10, 202–210. doi: 10.1080/17429145.2015.1067726
- Azameti, M., Ranjan, A., Singh, P. K., Gaikwad, K., Singh, A. K., Dalal, M., et al. (2022). Transcriptome profiling reveals the genes and pathways involved in thermo-tolerance in wheat (*Triticum aestivum* L.) genotype Raj 3765. *Sci. Rep.* 12, 14831. doi: 10.1038/s41598-022-18625-7
- Bitra, C. E., and Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front. Plant Sci.* 4, doi: 10.3389/fpls.2013.00273
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Carmo-Silva, A. E., Gore, M. A., Andrade-Sanchez, P., French, A. N., Hunsaker, D. J., and Salvucci, M. E. (2012). Decreased CO₂ availability and inactivation of rubisco limit photosynthesis in cotton plants under heat and drought stress in the field. *Environ. Exp. Bot.* 83, 1–11. doi: 10.1016/j.envexpbot.2012.04.001
- Chen, X., Lu, S. C., Wang, Y. F., Zhang, X., Lv, B., Luo, L. Q., et al. (2015). OsNAC2 encoding a NAC transcription factor that affects plant height through mediating the gibberellic acid pathway in rice. *Plant J.* 82, 302–314. doi: 10.1111/tpj.12819
- Close, T. J. (1997). Dehydrins: a commonality in the response of plants to dehydration and low temperature. *Physiol. Plant* 100, 291–296. doi: 10.1111/j.1365-3054.1997.tb04785.x
- Conesa, A., and Götz, S. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 619832. doi: 10.1155/2008/619832
- Davies, P. J. (2004). *Plant hormones: Biosynthesis, signal transduction, action!*. 3rd ed. (Netherlands: Springer) 2004.
- El-Sappah, E. H., Rather, S. A., Wani, S. H., Elrys, A. S., Bilal, M., Huang, Q., et al. (2022). Heat stress-mediated constraints in maize (*Zea mays*) production: Challenges and solutions. *Front. Plant Sci.* 13, 879366. doi: 10.3389/fpls.2022.879366
- Fernandes, J., Morrow, D., Casati, P., and Walbot, V. (2008). Distinctive transcriptome responses to adverse environmental conditions in *Zea mays* L. *Plant Biotechnol. J.* 6, 782–798. doi: 10.1111/j.1467-7652.2008.00360.x
- Frey, F. P., Urbany, C., Hüttel, B., Reinhardt, R., and Stich, B. (2015). Genome-wide expression profiling and phenotypic evaluation of European maize inbreds at seedling stage in response to heat stress. *BMC Genom.* 16, 123. doi: 10.1186/s12864-015-1282-1
- Galani, S., Wahid, A., and Arshad, M. (2013). Tissue-specific expression and functional role of dehydrins in heat tolerance of sugarcane (*Saccharum officinarum*). *Protoplasma* 250, 577–583. doi: 10.1007/s00709-012-0443-1
- Gilliam, M., Able, J. A., and Roy, S. J. (2017). Translating knowledge in abiotic stress tolerance to breeding programs. *Plant J.* 90, 898–917. doi: 10.1111/tpj.13456
- Gourdji, S. M., Sibley, A. M., and Lobell, D. B. (2013). Global crop exposure to critical high temperatures in the reproductive period: Historical trends and future projections. *Environ. Res. Lett.* 8, 024041. doi: 10.1088/1748-9326/8/2/024041
- Guo, R., Wang, X., Han, X., Li, W., Liu, T., Chen, B., et al. (2019). Comparative transcriptome analyses revealed different heat stress responses in high- and low-GS *Brassica alboblabra* sprouts. *BMC Genom.* 20, 269. doi: 10.1186/s12864-019-5652-y
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., and Bowden, J. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hara, M., Terashima, S., and Kuboi, T. (2001). Characterization and cryoprotective activity of cold-responsive dehydrin from *Citrus unshiu*. *J. Plant Physiol.* 58, 1333–1339. doi: 10.1078/0176-1617-00600
- Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *Int. J. Mol. Sci.* 14, 9643–9684. doi: 10.3390/ijms14059643
- He, J., Jiang, Z., Gao, L., You, C., Ma, X., Wang, X., et al. (2019). Genome-wide transcript and small RNA profiling reveals transcriptomic responses to heat stress. *Plant Physiol.* 181 (2), 609–629. doi: 10.1104/pp.19.00403
- Hu, W., Hu, G., and Han, B. (2009). Genome-wide survey and expression profiling of heat shock proteins and heat shock factors revealed overlapped and stress specific response under abiotic stresses in rice. *Plant Sci.* 176, 583–590. doi: 10.1016/j.plantsci.2009.01.016
- Hussain, H. A., Men, S., Hussain, S., Chen, Y., Ali, S., Zhang, S., et al. (2019). Interactive effects of drought and heat stresses on morpho-physiological attributes, yield, nutrient uptake and oxidative status in maize hybrids. *Sci. Rep.* 9, 3890. doi: 10.1038/s41598-019-40362-7
- Inghelant, D. V., Frey, F. P., Ries, D., and Stich, B. (2019). QTL mapping and genome-wide prediction of heat tolerance in multiple connected populations of temperate maize. *Sci. Rep.* 9, 14418. doi: 10.1038/s41598-019-50853-2
- Jagtap, A. B. (2020). *Identification and characterization of high temperature stress responsive genes in maize (Zea mays L.)*. PhD thesis (Ludhiana: Punjab Agricultural University).
- Jagtap, A. B., Vikal, Y., and Johal, G. S. (2020). Genome-wide development and validation of cost-effective KASP marker assays for genetic dissection of heat stress tolerance in maize. *Int. J. Mol. Sci.* 21, 7386. doi: 10.3390/ijms21197386
- Kim, J., Manivannan, A., Kim, D. S., Lee, E. S., and Lee, H. E. (2019). Transcriptome sequencing assisted discovery and computational analysis of novel SNPs associated with flowering in *Raphanus sativus* in-bred lines for marker-assisted backcross breeding. *Hortic. Res.* 6, 120. doi: 10.1038/s41438-019-0200-0
- Koirala, K. B., Giri, Y. P., Rijal, T. R., Zaidi, P. H., Sadananda, A. R., and Shrestha, J. (2017). Evaluation of grain yield of heat stress resilient maize hybrids in Nepal. *Int. J. Appl. Sci. Biotech.* 5, 511. doi: 10.3126/ijasbt.v5i4.18774
- Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., and Scharf, K. D. (2007). Complexity of the heat stress response in plants. *Curr. Opin. Plant Biol.* 10, 310–316. doi: 10.1016/j.pbi.2007.04.011
- Li, P., Cao, W., Fang, H., Xu, S., Yin, S., Zhang, Y., et al. (2017). Transcriptomic profiling of the maize (*Zea mays* L.) leaf response to abiotic stresses at the seedling stage. *Front. Plant Sci.* 8, doi: 10.3389/fpls.2017.00290
- Li, J., Gao, X., Sang, S., and Liu, C. (2019). Genome-wide identification, phylogeny, and expression analysis of the SBP-box gene family in euphorbiaceae. *BMC Genom.* 20, 912. doi: 10.1186/s12864-019-6319-4
- Li, Z., and Howell, S. H. (2021). Heat stress responses and thermotolerance in maize. *Int. J. Mol. Sci.* 22, 948. doi: 10.3390/ijms22020948
- Li, H., Liu, S. S., Yi, C. Y., Wang, F., Zhou, J., Xia, X. J., et al. (2014). Hydrogen peroxide mediates abscisic acid-induced HSP70 accumulation and heat tolerance in grafted cucumber plants. *Plant Cell Environ.* 37, 2768–2780. doi: 10.1111/pce.12360
- Li, T., Xu, X., Li, Y., Wang, H., Li, Z., and Li, Z. (2015). Comparative transcriptome analysis reveals differential transcription in heat-susceptible and heat-tolerant pepper (*Capsicum annuum* L.) cultivars under heat stress. *J. Plant Biol.* 58, 411–424. doi: 10.1007/s12374-015-0423-z

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1117136/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Principal component analysis (PCA plot) of all RNA-seq maize samples at reproductive stage under heat stress.

SUPPLEMENTARY FIGURE 2

Heatmap of differentially expressed genes (DEGs) in response to heat stress. DEGs have been identified with significant ($P < 0.05$) and $|\log_2(\text{fold change})| \geq 2$ in leaf, pollen and ovule of CML 25 (HT) and LM 11 (HS) inbred. Colors represent \log_2 up-regulation (yellow) and down-regulation (blue) of DEGs.

- Li, Z. G., and Ye, X. Y. (2022). Transcriptome response of maize (*Zea mays* L.) seedlings to heat stress. *Protoplasma* 259 (2), 357–369. doi: 10.1007/s00709-021-01680-8
- Liu, Y., Li, D., Song, Q., Zhang, T., Li, D., and Yang, X. (2019). The maize late embryogenesis abundant protein ZmDHN13 positively regulates copper tolerance in transgenic yeast and tobacco. *Crop J.* 7, 403–410. doi: 10.1016/j.cj.2018.09.001
- Liu, T., Zhang, L., Yuan, Z., Hu, X., Lu, M., Wang, W., et al. (2013). Identification of proteins regulated by ABA in response to combined drought and heat stress in maize roots. *Acta physiol. Plant* 35, 501–513. doi: 10.1007/s11738-012-1092-x
- Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., and Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science* 319, 607–610. doi: 10.1126/science.1152339
- Longmei, N., Gill, G. K., Zaidi, P. H., Kumar, R. K., Nair, S. K., Hindu, V., et al. (2021). Genome wide association mapping for heat tolerance in sub-tropical maize. *BMC Genom.* 22, 154. doi: 10.1186/s12864-021-07463-y
- Mangelsen, E., Kilian, J., Harter, K., Jansson, C., Wanke, D., and Sundberg, E. (2011). Transcriptome analysis of high-temperature stress in developing barley caryopses: early stress responses and effects on storage compound biosynthesis. *Mol. Plant* 4, 97–115. doi: 10.1093/mp/ssq058
- Martins, M. A., Tomasella, J., and Dias, C. G. (2019). Maize yield under a changing climate in the Brazilian northeast: impacts and adaptation. *Agric. Water Manage.* 216, 339–350. doi: 10.1016/j.agwat.2019.02.011
- Mathur, S., Agrawal, D., and Jajoo, A. (2014). Photosynthesis: response to high temperature stress. *J. Photochem. Photobiol. B* 137, 116–126. doi: 10.1016/j.jphotobiol.2014.01.010
- McGettigan, P. A. (2013). Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* 17, 4–11. doi: 10.1016/j.cbpa.2012.12.008
- Mellidou, I., Moschou, P. N., Ioannidis, N. E., Pankou, C., Gemes, K., Valassakis, C., et al. (2016). Silencing s-Adenosyl-L-Methionine decarboxylase (SAMDC) in *Nicotiana tabacum* points at a polyamine-dependent trade-off between growth and tolerance responses. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00379
- Nandha, A. K., Mehta, D. R., Tulsani, N. J., Umretiya, N., Delvadiya, N., and Kachhadiya, H. J. (2019). Transcriptome analysis of response to heat stress in heat tolerance and heat susceptible wheat (*Triticum aestivum* L.) genotypes. *J. Pharmacognosy Phytochemistry* 8 (2), 275–284. doi: 10.1186/1471-2164-9-432
- Noor, J. J., Vinayan, M. T., Umar, S., Devi, P., Iqbal, M., Seetharam, K., et al. (2019). Morpho-physiological traits associated with heat stress tolerance in tropical maize (*Zea mays* L.) at reproductive stage. *Aust. J. Crop Sci.* 13, 536–545. doi: 10.21475/AJCS.19.13.04.P1448
- Oliveros, J. C. (2007) Venny-an interactive tool for comparing lists with Venn diagrams. Available at: <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
- Porter, J. R. (2005). Rising temperatures are likely to reduce crop yields. *Nature* 436, 174. doi: 10.1038/436174b
- Qian, Y., Ren, Q., Zhang, J., and Chen, L. (2019). Transcriptomic analysis of the maize (*Zea mays* L.) inbred line B73 response to heat stress at the seedling stage. *Gene* 692, 68–78. doi: 10.1016/j.gene.2018.12.062
- Qin, D., Wu, H., Peng, H., Yao, Y., Ni, Z., Li, Z., et al. (2008). Heat stress responsive transcriptome analysis in heat susceptible and tolerant wheat (*Triticum aestivum* L.) by using wheat genome array. *BMC Genom.* 9, 432. doi: 10.1186/1471-2164-9-432
- Qu, A. L., Ding, Y. F., Jiang, Q., and Zhu, C. (2013). Molecular mechanisms of the plant heat stress response. *Biochem. Biophys. Res. Commun.* 432, 203–207. doi: 10.1016/j.bmolp.2020.02.004
- Rahmati Ishka, M., Brown, E., Weigand, C., Tillett, R. L., Schlauch, K. A., Miller, G., et al. (2018). A comparison of heat-stress transcriptome changes between wild-type *Arabidopsis* pollen and a heat-sensitive mutant harboring a knockout of cyclic nucleotide-gated cation channel 16 (cngc16). *BMC Genomics* 19, 549. doi: 10.1186/s12864-018-4930-4
- Raju, G., Shanmugam, K., and Kasirajan, L. (2020). High-throughput sequencing reveals genes associated with high-temperature stress tolerance in sugarcane. *3 Biotech.* 10, 198. doi: 10.1007/s13205-020-02170-z
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rouf Shah, T., Prasad, K., and Kumar, P. (2016). Maize - a potential source of human nutrition and health: A review, cogent. *Food Agric.* 2, 1166995. doi: 10.1080/23311932.2016.1166995
- Saha, G., Park, J. I., Kayum, M. A., and Nou, I. S. (2016). A genome-wide analysis reveals stress and hormone responsive patterns of TIFY family genes in *Brassica rapa*. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00936
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Shi, J., Yan, B., Lou, X., Ma, H., and Ruan, S. (2017). Comparative transcriptome analysis reveals the transcriptional alterations in heat-resistant and heat-sensitive sweet maize (*Zea mays* L.) varieties under heat stress. *BMC Plant Biol.* 17, 26. doi: 10.1186/s12870-017-0973-y
- Singh, B., Salaria, N., Thakur, K., Kukreja, S., Gautam, S., and Goutam, U. (2019). Functional genomic approaches to improve crop plant heat stress tolerance. *F1000Res* 8, 1721. doi: 10.12688/f1000research.19840.1
- Song, J., Nada, K., and Tachibana, S. (2002). Suppression of S-adenosylmethionine decarboxylase activities is a major cause for high temperature inhibition of pollen germination and tube growth in tomato (*Lycopersicon esculentum* mill.). *Plant Cell Physiol.* 43, 619–627. doi: 10.1093/pcp/pcf078
- Synek, L., Vukašinović, N., Kulich, I., Hála, M., Aldorfová, K., Fendrych, M., et al. (2017). EXO70C2 is a key regulatory factor for optimal tip growth of pollen. *Plant Physiol.* 174, 223–240. doi: 10.1104/pp.16.01282
- Tas, T. (2022). Physiological and biochemical responses of hybrid maize (*Zea mays* L.) varieties grown under heat stress conditions. *Peer J.* 21, 10:e14141. doi: 10.7717/peerj.14141
- Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets on to diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939. doi: 10.1111/j.1365-3113.2004.02016.x
- Tiwari, Y. K., and Yadav, S. K. (2019). High temperature stress tolerance in maize (*Zea mays* L.): Physiological and molecular mechanisms. *J. Plant Biol.* 62, 93–102. doi: 10.1007/s12374-018-0350-x
- Ul haq, A., Ali, M., Khattak, A. M., Gai, W. X., Zhang, H. X., Wei, A. M., et al. (2019). Heat shock proteins: Dynamic biomolecules to counter plant biotic and abiotic stresses. *Int. J. Mol. Sci.* 20, 5321. doi: 10.3390/ijms2015321
- Wahid, A., Gelani, S., Ashraf, M., and Foolad, M. R. (2007). Heat tolerance in plants: an overview. *Environ. Expl. Bot.* 61, 199–223. doi: 10.1016/j.envexpbot.2007.05.011
- Wang, A., Hu, J., Huang, X., Li, X., Zhou, G., and Yan, Z. (2016). Comparative transcriptome analysis reveals heat-responsive genes in Chinese cabbage (*Brassica rapa* ssp. *chinensis*). *Front. Plant Sci.* 7, 939. doi: 10.3389/fpls.2016.00939
- Wang, K., Liu, Y., Tian, J., Huang, K., Shi, T., Dai, X., et al. (2017). Transcriptional profiling and identification of heat-responsive genes in perennial ryegrass by RNA-sequencing. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01032
- Wu, D. C., Zhu, J. F., Shu, Z. Z., Wang, W., Yan, C., Xu, S. B., et al. (2020). Physiological and transcriptional response to heat stress in heat-resistant and heat-sensitive maize (*Zea mays* L.) inbred lines at seedling stage. *Protoplasma* 257 (6), 1615–1637. doi: 10.1007/s00709-020-01538-5
- 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
- Xuhui, L., Weiwei, C., Siqi, L., Jungteng, F., Hang, Z., Xiangbo, Z., et al. (2022). Full-length transcriptome analysis of maize root tips reveals the molecular mechanism of cold stress during the seedling stage. *BMC Plant Biol.* 22, 398. doi: 10.1186/s12870-022-03787-3
- Yang, Q., Zhang, H., Liu, C., Huang, L., Zhao, L., and Zhang, A. (2018). A NAC transcription factor ZmNAC84 affects pollen development through the repression of *ZmRbohH* expression in maize. *J. Plant Biol.* 61, 366–373. doi: 10.1007/s12374-018-0227-z
- Yu, Z., Wang, X., and Zhang, L. (2018). Structural and functional dynamics of dehydrins: A plant protector protein under abiotic stress. *Int. J. Mol. Sci.* 19, 3420. doi: 10.3390/ijms19113420
- Zhang, X., Rerksiri, W., Liu, A., Zhou, X., Xiong, H., Xiang, J., et al. (2013). Transcriptome profile reveals heat response mechanism at molecular and metabolic levels in rice flag leaf. *Gene* 530, 185–192. doi: 10.1016/j.gene.2013.08.048
- Zhang, S., Wang, L., Sun, X., Li, Y., Yao, J., van Nocker, S., et al. (2019). Genome-wide analysis of the YABBY gene family in grapevine and functional characterization of VvYABBY4. *Front. Plant Sci.* 10, 1207. doi: 10.3389/fpls.2019.01207
- Zhang, L., Zhao, H. K., Dong, Q. L., Zhang, Y. Y., Wang, Y. M., Li, H. Y., et al. (2015). Genome-wide analysis and expression profiling under heat and drought treatments of HSP70 gene family in soybean (*Glycine max* L.). *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00773
- Zhao, Y., Hu, F., Zhang, X., We, Q., Dong, J., Bo, C., et al. (2019). Comparative transcriptome analysis reveals important roles of nonadditive genes in maize hybrid an'ong 591 under heat stress. *BMC Plant Biol.* 19, 273. doi: 10.1186/s12870-019-1878-8
- Zhao, Y., Tian, X., Wang, F., Zhang, L., Xin, M., Hu, Z., et al. (2017). Characterization of wheat MYB genes responsive to high temperatures. *BMC Plant Biol.* 17, 208. doi: 10.1007/s10812-021-01192-6



OPEN ACCESS

EDITED BY

Md. Anwar Hossain,
University of Rajshahi, Bangladesh

REVIEWED BY

Golam Jalal Ahammed,
Henan University of Science and
Technology, China
Peng Shuai,
Fujian Agriculture and Forestry University,
China
Jinhuan Chen,
Beijing Forestry University, China

*CORRESPONDENCE

Chuyu Ye

✉ yecy@zju.edu.cn

[†]These authors have contributed equally to
this work

SPECIALTY SECTION

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

RECEIVED 22 November 2022

ACCEPTED 07 February 2023

PUBLISHED 17 February 2023

CITATION

Sultana MH, Alamin M, Qiu J, Fan L and
Ye C (2023) Transcriptomic profiling
reveals candidate allelopathic genes in
rice responsible for interactions
with barnyardgrass.
Front. Plant Sci. 14:1104951.
doi: 10.3389/fpls.2023.1104951

COPYRIGHT

© 2023 Sultana, Alamin, Qiu, Fan and Ye.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Transcriptomic profiling reveals candidate allelopathic genes in rice responsible for interactions with barnyardgrass

Most. Humaira Sultana^{1†}, Md. Alamin^{2†}, Jie Qiu¹, Longjiang Fan¹
and Chuyu Ye^{1*}

¹Institute of Crop Science and Institute of Bioinformatics, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China, ²Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, China

Echinochloa crus-galli (barnyardgrass) is one of the most damaging weeds in rice fields worldwide. Allelopathy has been considered a possible application for weed management. Thus understanding its molecular mechanisms is important for rice production. This study generated transcriptomes from rice under mono- and co-culture with barnyardgrass at two-time points to identify the candidate genes controlling allelopathic interactions between rice and barnyardgrass. A total of 5,684 differentially expressed genes (DEGs) were detected, amongst which 388 genes were transcription factors. These DEGs include genes associated with momilactone and phenolic acid biosynthesis, which play critical roles in allelopathy. Additionally, we found significantly more DEGs at 3 hours than at 3 days, suggesting a quick allelopathic response in rice. Up-regulated DEGs involve diverse biological processes, such as response to stimulus and pathways related to phenylpropanoid and secondary metabolites biosynthesis. Down-regulated DEGs were involved in developmental processes, indicating a balance between growth and stress response to allelopathy from barnyardgrass. Comparison of DEGs between rice and barnyardgrass shows few common genes, suggesting different mechanisms underlying allelopathic interaction in these two species. Our results offer an important basis for identifying of candidate genes responsible for rice and barnyardgrass interactions and contribute valuable resources for revealing its molecular mechanisms.

KEYWORDS

transcriptome, allelopathy, barnyardgrass, *Echinochloa crus-galli*, rice, rice and barnyardgrass interaction

Introduction

Rice (*Oryza sativa*) has played a vital role in human nutrition and culture for over 10,000 years. It is one of the major staple crops worldwide, and more than fifty percent of the world's population consumes it (Khush, 2005). Asian countries grow and eat more than 90% of the rice in the world (Khush, 2005). In the middle of this century, the global

population will reach 9 billion if it continues to grow at current projections (Godfray et al., 2010). To meet the demand caused by the increasing world population, income, and consumption, we will have to increase rice production by 70% by 2050 (Varshney et al., 2011; Alamin et al., 2018). We need rice varieties with higher potential and higher yield stability to meet the challenges of manufacturing more rice from suitable lands (Khush, 2005). Much promise for developing new varieties is due to recent advancements in rice genomics.

Weeds are unacceptable and redundant plants that adversely influence human benefit by using land and water resources (Rao, 2000). Rice yield is decreased by weeds that compete with crops for nutrients, moisture, and light. Various factors further mediate yields, such as crop cultivars, weed variety, and the relative density of the crop and weeds (Rezaeieh et al., 2015). Damage caused by weeds is greater than that caused by pests (Oerke, 2006). Research has shown that weeds reduced crop production equivalent to \$95 billion per annum worldwide (Lundkvist and Verwijst, 2011). Barnyardgrass (*Echinochloa crus-galli*) is a universally harmful weed within the grass family (Guo et al., 2017). It is a typical and devastating weed in paddy fields (Xuan et al., 2006). It has been reported that barnyardgrass made complexity in 61 countries and a minimum of 36 different crops (Xuan et al., 2006). Research showed that 35% of rice yield globally loses due to interactions with barnyardgrass (Oerke and Dehne, 2004).

Allelopathy is the biological incident in which one organism influences another organism's growth, survival, or reproduction by releasing biochemical-termed allelochemicals (Cheng and Cheng, 2015). Secondary metabolites are the main component of allelochemicals, which are discharged into the atmosphere throughout the usual pathways (Cheng and Cheng, 2015). The allelochemicals reduce plant development due to the changes in plant growth controllers or phytohormones (Cheng and Cheng, 2015). Plants' most general and universal secondary metabolites are flavonoids (Du et al., 2010). Plant growth, advancement and adjustment to stress are influenced by flavonoids, which are created by a wing of the phenylpropanoid pathway (Shah and Smith, 2020). Auxin transport is regulated by flavonoids, resulting in the development of a special tissue and, subsequently, the whole plant influenced (Singh et al., 2021; Ahammed and Li, 2022). Terpenoids are architecturally diverse and play a significant role in different defense mechanisms in the plant (Dudareva et al., 2004; Cheng et al., 2007). Phenylpropanoid is a key component of plant-particularized metabolism. It acts as the main biochemical modulator of plant communication with insects and germs, acting with opposing attractive and offensive functions in protective phytoalexin reactions to contagion and herbivory (Ferrer et al., 2008). Pigmentation, defense against UV photodamage, morphological strength by polymeric lignin and diverse antimicrobial phytoalexins supported by phenylpropanoid (Ferrer et al., 2008). Lignin is the collective term for a significant collection of aromatic polymers, and this polymerase is stored mainly in the secondary cell wall (Vanhholme et al., 2010). In lignin biosynthesis, CAD is one of the first enzymes, and many CAD homologs have been isolated from diverse plant species, including rice (Vanhholme et al., 2010).

Several secondary metabolites, including fatty acids, indoles, momilactones, phenolics and terpenes, were discovered as major allelochemicals in rice plants, which suppress the growth of Barnyardgrass (Khanh et al., 2005). Recently, much more interest has been given to allelopathy research owing to its possible application for weed controlling and crop production (Khanh et al., 2007). Rice allelopathy has been used for controlling barnyardgrass in many studies (Xuan et al., 2006). Moreover, flavonoids, diterpenoids, and other biochemicals have been determined as effective allelochemicals from different parts of rice plants (Kong, 2007). Thus, rice allelopathy would accelerate the defense mechanism by the chemically mediated interchange between rice and Barnyardgrass plants (Kato-Noguchi, 2011). Many studies have been conducted using the molecular approach for the biosynthesis of momilactones and phenolic acids and their allelopathic impacts on barnyardgrass. Momilactones A played a vital role in rice protection against mycological pathogens (Mennan et al., 2012) and momilactone B is an important allelochemical in the rice-barnyardgrass interaction (Kato-Noguchi and Peters, 2013). Two gene clusters have been identified in the synthesis of the momilactones (Xu et al., 2012; Kato-Noguchi and Peters, 2013). It was proposed that the biosynthesis path of phenolic synthesis might be part of the acute allelopathic apparatuses for rice. This is because the PAL (phenylalanine ammonia-lyase) was more susceptible to allelopathic rice, PI312777, than non-allelopathic rice (Fang et al., 2010; Fang et al., 2013; Zhang et al., 2018). Other study demonstrated that the improved expression level of several genes, including *COMT*, *C4H*, *PAL*, and *F5H*, were associated with phenolic compounds of rice, resulting in a reverse impact on barnyardgrass (He et al., 2012).

Biological research has been changing rapidly due to current advances in genomic sequencing technologies, which significantly affect crop improvement (Edwards and Batley, 2010; Mochida and Shinozaki, 2010). Transcriptomics research has been done using the next-generation sequencing (NGS) supported RNA-Seq technique is a robust tool for exploring the function of genes in various tissues and varied environments (Chen et al., 2011). Zhang et al. (2010) reported that transcriptome analysis is very important for discovering genotypic and environmental interactions. However, genomics and bioinformatics information is still limited for weeds affecting modern crops, in particular, the involvement of transcription factors remains largely unknown (Duke and Gressel, 2010). Much progress has been made in NGS technology in recent years, yet few studies have focused on the biological mechanisms of the rice and barnyardgrass interaction.

In this study, RNA-seq was utilized to discover the genes as well as their functional relationship for controlling rice and barnyardgrass interactions at different time points. To improve the genomic sources for weeds, we sequenced the transcriptomes of rice co-cultured with barnyardgrass using the Illumina platform. In the present study, clustering of putative functional categories was carried out with the Gene Ontology (GO) structure, alignment into route utilization with the Kyoto Encyclopedia of Genes and Genomes (KEGG) databank, and visualization of profiling data sets within the context of existing knowledge for the functional investigation of DEGs utilizing the MapMan software. This study

provides valuable genetic information regarding barnyardgrass and rice interaction and is beneficial for practical weed regulation.

Materials and methods

Plant materials and growth conditions

The allelopathic experiment between rice (PI312777) and barnyardgrass (STB08) was conducted according to methods previously described (Guo et al., 2017; Sultana et al., 2019). In detail, the adapted transmit seeding in the agar method (Navarez and Olofsdotter, 1996) was applied to examine the allelopathic connections among rice and barnyardgrass. Ten PI312777 sprouted seeds were initially transmitted to a pot (10 cm basis dimension) compacted with 50 mL of 0.5% agar medium, and STB08 sprouted seeds were relocated to a dish with ddH₂O. The PI312777 seeds were organized in three lines with a 3-4-3 formation, and ten STB08 sprouted seeds were then relocated to the identical dish after 5 days, and five seeds were grown up between rows of PI312777 spores (Supplementary Figure S1). The PI312777 and STB08 plants were co-cultivated for 3h and 3d in a SAFE chamber (Ningbo, China) at 30°C in light (14 h) environments, 20°C in darkness (10 h) environments, and 75% moisture. The controls comprised of PI312777 alone grown in a container. The first sample (whole plant) and second sample (whole plant) were collected after 3h and 3d of treatment, rinsed five times with ddH₂O, instantly frozen in liquid N₂, and kept in RNase-free tubes at −80°C for RNA separation. For each analysis, ten plants from one pot were pooled. Three replicates for both control and treatment were used in this study.

RNA extraction, cDNA library preparation, and illumina sequencing

Total RNAs were extracted employing the RNeasy Plant Mini Kit (Qiagen) as stated by the manufacturer's guidelines. The first-strand cDNA was produced by a First Strand cDNA Synthesis kit (TaKaRa, Dalian, China). Illumina RNA-Seq libraries for PI312777 at 3h and 3d time points were organized and sequenced on a HiSeq 4000 technique complying with the producer's guidelines.

Data filtering, read mapping, and gene quantification

Data were purified earlier in the downstream investigation to reduce data noise. Reads with adaptors, low-quality reads, and reads with more than 10% unknown bases were removed. NGS QC Toolkit version v2.3.3 was utilized to attribute and purify of the sequencing data with the standard setting (Patel and Jain, 2012). Clean reads, the remaining reads after filtering, were saved as a FASTQ data format (Cock et al., 2010). Bowtie (Langmead et al., 2009) and TopHat2

(Kim et al., 2013) were utilized to map clean reads to reference genes and to the reference genome of the Nipponbare subspecies of rice (MSU7.0, <http://rice.plantbiology.msu.edu/>), respectively. After alignment, the Cufflinks package was applied to calculate the number of reads that were mapped to each gene within the gene model annotation file. The subsequent alignment files were then provided to Cuffdiff in the Cufflinks package to determine every gene's FPKM (fragments per kilobase per million reads) value.

Differentially expressed genes analysis and annotation

The Cuffdiff method (Trapnell et al., 2012) was used for identifying the DEGs (differentially expressed genes). DEGs were selected with p -value < 0.05 regarded as DEGs. The Gene Ontology (GO) enrichment was executed utilizing agriGO v2.0 (Tian et al., 2017) with Singular Enrichment Analysis (SEA), *Oryza sativa* japonica was chosen as the species and MSU7.0 gene ID (TIGR) as a reference background. The Fisher test with Yekutieli (FDR under dependency) Multi-test adjustment approach was used based on the significance level 0.05. The KO-Based Annotation System (KOBAS) (Xie et al., 2011) was used for getting a summary of the gene pathway network for KEGG analysis. The MapMan (version 3.6.0RC1) and pathways were downloaded from the MapMan Site (<http://mapman.gabipd.org/mapmanstore>). The mapping info of rice genes (osa_MSU_v7_mapping.txt.tar.gz) was imported to the MapMan to get a more standard overview of deviations in gene expression. To measure common genes between all samples of a two-time points experiment, we used jvenn (<http://jvenn.toulouse.inra.fr/app/index.html>), an integrative tool for comparing lists with Venn Diagrams which offers statistic charts founded on input data. Circos (Krzywinski et al., 2009) version-0.69-6 was used to visualize details of a chromosomal view of DEGs. The PlantTFDB database (<http://planttfdb.cbi.pku.edu.cn/prediction.php>) was used to identify transcription factors of DEGs.

Time-series expression profile

STEM (Short Time-series Expression Miner) was used to investigate short time-series expression data (Ernst and Bar-Joseph, 2006). It uses an innovative clustering process to distinguish between actual and arbitrary shape and group genes by allocating them to expression profiles. A profile is considered important if the total of genes allocated to it exceeds the number of genes predicted to happen by chance. The number of significant genes allocated to each profile against the projected number was calculated and modified for incorrect detection rate at the arrangement of a gene. The limit for the STEM clustering procedure was fixed to 50 for the “maximum number of model profiles” and the “maximum unit variation in model profiles across time points” was set to two to accept clustering at a realistic number of probable model profiles.

Identification of homolog genes

To identify the rice DEGs homologs to barnyardgrass (Guo et al., 2017) and previously identified allelochemical genes (Amb and Ahluwalia, 2016), we created a protein database by using *O. sativa* MSU7 protein sequences. BLASTP was then utilized to scan homologous genes in the protein database considering thresholds: 10^{-10} (E-value).

Gene expression validation using qRT-PCR

The Trizol (Invitrogen, Carlsbad, CA, USA) method was used for RNA extraction based on the company's protocol. cDNA preparation and qRT-PCR (quantitative real-time PCR) was performed according to a previous study (Sultana et al., 2019). The protocol for the qRT-PCR was as follows: 95°C for 30 s, 45 cycles at 95°C for 10 s for denaturation, annealing at 60°C for 10 s, and extension at 72°C for 20 s. Three biological repeats were implemented for every treatment. Relative expression levels were estimated following the process of (Livak and Schmittgen, 2001). We have calculated the qRT-PCR log₂ (FC) values based on the

calculation of the log₂ (normalized ratio values) from the mean values of three replicates for each gene and compared them with RNA-Seq log₂ (FC) values. The rice *actin* gene was utilized as a control, and the primers for the gene used in qRT-PCR are provided in Supplementary Table S1.

Results

Summary of RNA-seq data

The rice cultivar PI312777 (known to have high allelopathic potential) was cultivated independently (mono-cultured) or co-cultured with barnyardgrass to find potential genes that participated in allelopathy of rice against barnyardgrass. Rice transcriptomic data were then generated at two-time points (3h and 3d) with three biological replicates in control (mono-cultured) and treatment (co-cultured with barnyardgrass) conditions using RNA-Seq. There were 25.97 million (3h) and 25.48 million (3d) raw reads in twelve samples (Table 1). The numbers of total clean reads were 256 million (3h) and 250 million (3d) in twelve samples (Table 1). The overall average mapping rates were 75.12% (Table 1).

TABLE 1 Mapping results of RNA sequencing reads.

| Time points | Sample ID | Raw bases | Clean bases (%) | Raw reads | Clean reads (%) | Mapped reads (%) | Multiple alignments (%) |
|-------------|-----------|------------|---------------------|-----------|-------------------|------------------|-------------------------|
| 3h | 3h_C1 | 5968809900 | 5796177274 (98.64%) | 39792066 | 39175940 (98.45%) | 28011714 (71.5%) | 878439 (3.1%) |
| | 3h_C2 | 6417289200 | 6262043338 (98.81%) | 42781928 | 42250364 (98.76%) | 31823432 (75.3%) | 1397216 (4.4%) |
| | 3h_C3 | 5768741400 | 5605836673 (98.7%) | 38458276 | 37863546 (98.45%) | 27263894 (72.2%) | 845754 (3.1%) |
| | 3h_T1 | 6209253300 | 6033434398 (98.64%) | 41395022 | 40777296 (98.51%) | 29353016 (71.9%) | 1202374 (4.1%) |
| | 3h_T2 | 8228647200 | 7929442588 (97.98%) | 54857648 | 53950196 (98.35%) | 37421744 (69.4%) | 760703 (2%) |
| | 3h_T3 | 6370019700 | 6113475248 (97.97%) | 42466798 | 41599722 (97.96%) | 28632334 (68.8%) | 978825 (3.5%) |
| 3d | 3d_C1 | 5942038800 | 5846652549 (98.39%) | 39792066 | 38977948 (98.4%) | 29779939 (76.4%) | 1044610 (3.5%) |
| | 3d_C2 | 5929839300 | 5840730517 (98.49%) | 42781928 | 39058816 (98.8%) | 30807236 (78.9%) | 708212 (2.3%) |
| | 3d_C3 | 6627531300 | 6521318373 (98.39%) | 38458276 | 43387320 (98.2%) | 33754145 (77.8%) | 1084281 (3.3%) |
| | 3d_T1 | 7238626200 | 7123865032 (98.42%) | 41395022 | 47472718 (98.37%) | 38364086 (80.8%) | 1098492 (2.9%) |
| | 3d_T2 | 6706004400 | 6597933758 (98.39%) | 54857648 | 44002982 (98.43%) | 34153651 (77.6%) | 812104 (2.4%) |
| | 3d_T3 | 5778936900 | 5620990114 (97.26%) | 42466798 | 37578398 (97.54%) | 30372971 (80.8%) | 1018915 (3.4%) |

N.B 3h C1 and 3h T1 mean control sample 1 and treatment sample 1, respectively, at a 3-hour time point. 3d C1 and 3d T1 mean control sample 1 and treatment sample 1, respectively, at a 3-day time point. Other sample names indicate a similar way.

Identification of differentially expressed genes

A total of 5,684 differentially expressed genes (DEGs) were identified. Among them, 3,749 (65.96%) and 1,935 (34.04%) DEGs

were identified at 3h and 3d time points, respectively. DEGs with up-regulation were more common than those with down-regulation (Figure 1A). Moreover, a total of 393 common DEGs were found at the two-time points. Among them, 83 up-regulated and 113 down-regulated genes were common, 92 genes were up-regulated at 3h

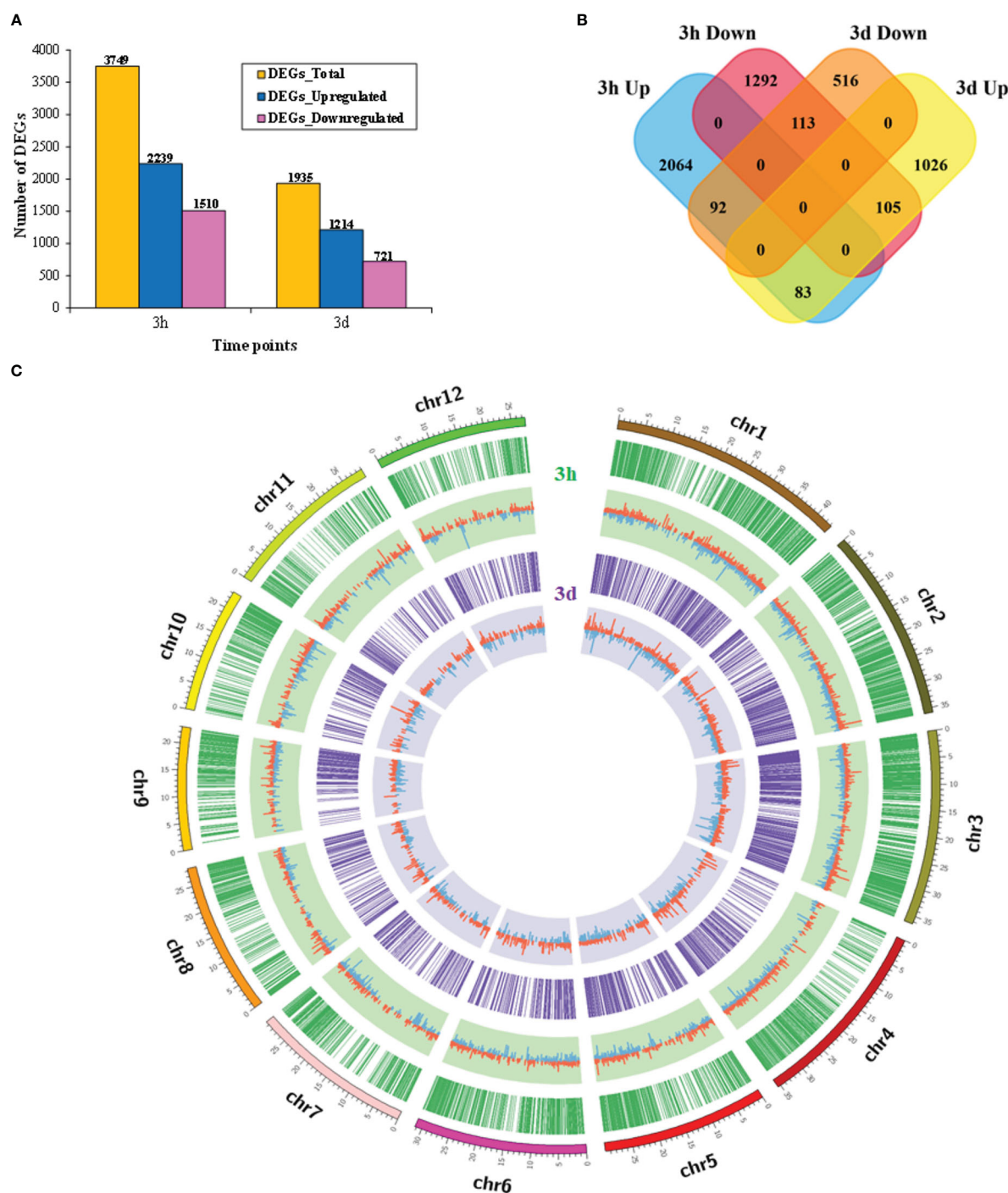


FIGURE 1

The expression profile of rice differentially expressed genes (DEGs) grown with barnyardgrass. (A) Bar plot represents the total number, up-regulated, and down-regulated DEGs at 3h and 3d time points, respectively. (B) Venn diagrams represent the numbers of DEGs and the overlaps of sets obtained across the two time points. (C) Distribution features of DEGs of the rice genome under rice and barnyardgrass interaction using a Circos plot. Different colors show the sizes of the 12 chromosomes in rice. The green circle and light green color represent the distribution and expression (log2 FC) value of DEGs, respectively, at 3h. The purple circle and light purple color represent the distribution and expression (log2 FC) value of DEGs, respectively, at 3d. Up- and down-regulated genes are shown by red and blue, respectively.

and down-regulated at 3d, and 105 genes were down-regulated at 3h and up-regulated at 3d (Figure 1B). Furthermore, a Circos plot was drawn to show up- and down-regulated gene distribution patterns in chromosomes at 3h and 3d. Results showed that the DEGs were evenly distributed over the 12 chromosomes in rice (Figure 1C).

Identification of allelochemical responsible transcription factors

The PlantTFDB database was used to identify the transcription factors (TFs) (Jin et al., 2017). A total of 388 TFs from 44 different TFs families were detected to be DEGs, such as bHLH, ERF, WRKY, NAC, MYB, bZIP, C2H2, GRAS, HD-ZIP, TALE, and GATA families (Figure 1A, Supplementary Table S2). These TFs

accounted for more than 50% of the total differentially expressed TFs. More differentially expressed TFs (278 TFs from 41 families) were found at 3h than those (110 TFs from 28 families) at the 3d time point (Figure 2B, Supplementary Table S2).

Gene ontology and pathway enrichment analysis of DEGs

Gene Ontology (GO) enrichment analysis for DEGs at 3h and 3d was performed ($P < 0.05$; Supplementary Table S3). The top 30 GO terms in each time point are given in Figures 3A, B. Twelve GO terms were common between the two-time points, which were “response to stimulus”, “metabolic process”, “response to abiotic stimulus”, “response to biotic stimulus”, “response to stress”, “DNA

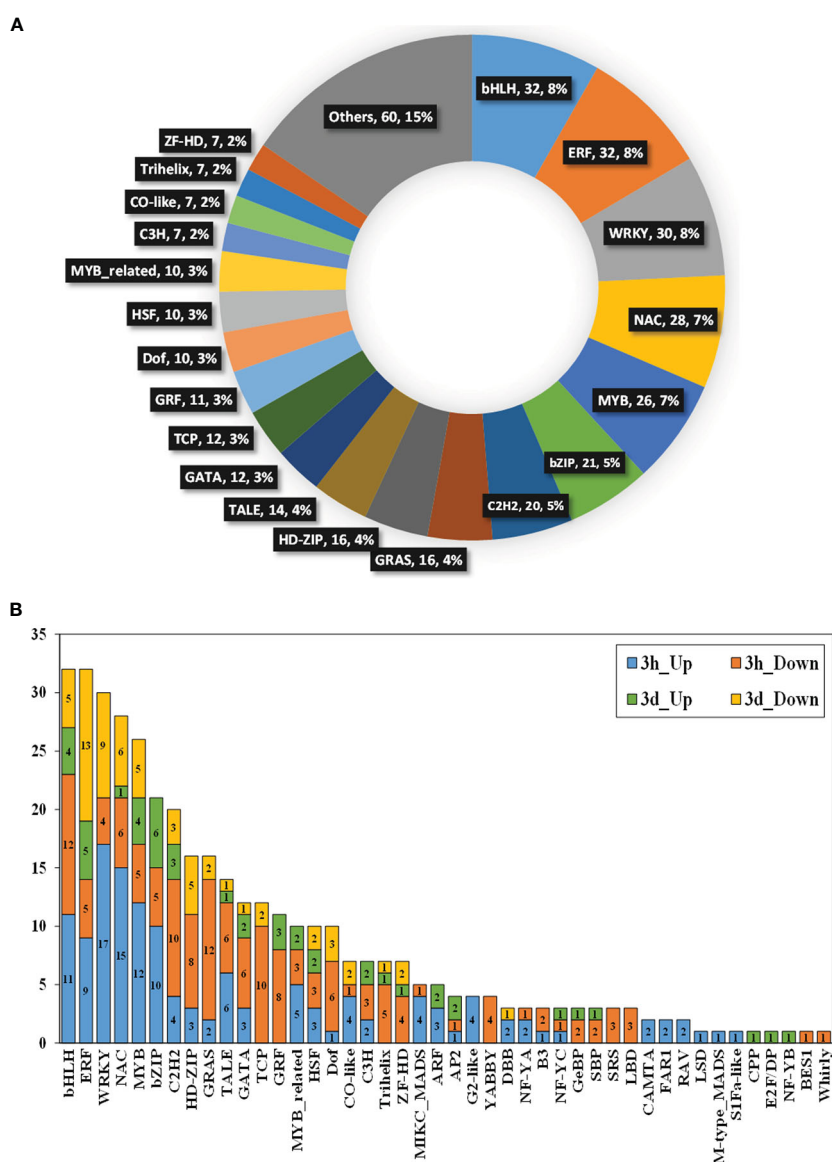


FIGURE 2

Transcription factors (TFs) analysis of DEGs identified for rice and barnyardgrass interaction. (A) Pie chart represent the TFs families with their number and percentage for all DEGs. (B) Number of genes associated with different transcription factors at 3h and 3d time points. Different colors represent different transcription factor gene families and the genes number for up-regulated and down-regulated genes for 3h and 3d time points.

binding”, “extracellular region”, “external encapsulating structure”, “cell wall”, “cell”, “cell part”, and “vacuole”.

Genes typically act together with one another to take part in specific biological activities. KEGG database (Kanehisa et al., 2016) dependent pathway analysis was carried out to investigate biological processes during the interaction of rice and barnyardgrass. The results showed that 1,098 DEGs were involved in the 36 pathways at 3h, and 404 DEGs were involved in 18 pathways at 3d time point

(Figures 3C, D, Supplementary Table S4). Most genes were enriched in “metabolic pathways”, followed by “biosynthesis of secondary metabolites”, “carbon metabolism”, “phenylpropanoid biosynthesis”, “starch and sucrose metabolism”, “plant hormone signal transduction”, “glutathione metabolism and cysteine”, and “methionine metabolism” at 3h (Figure 3C, Supplementary Table S4). “Ribosome,” “biosynthesis of secondary metabolites,” and “plant hormone signal transduction” were the most enriched

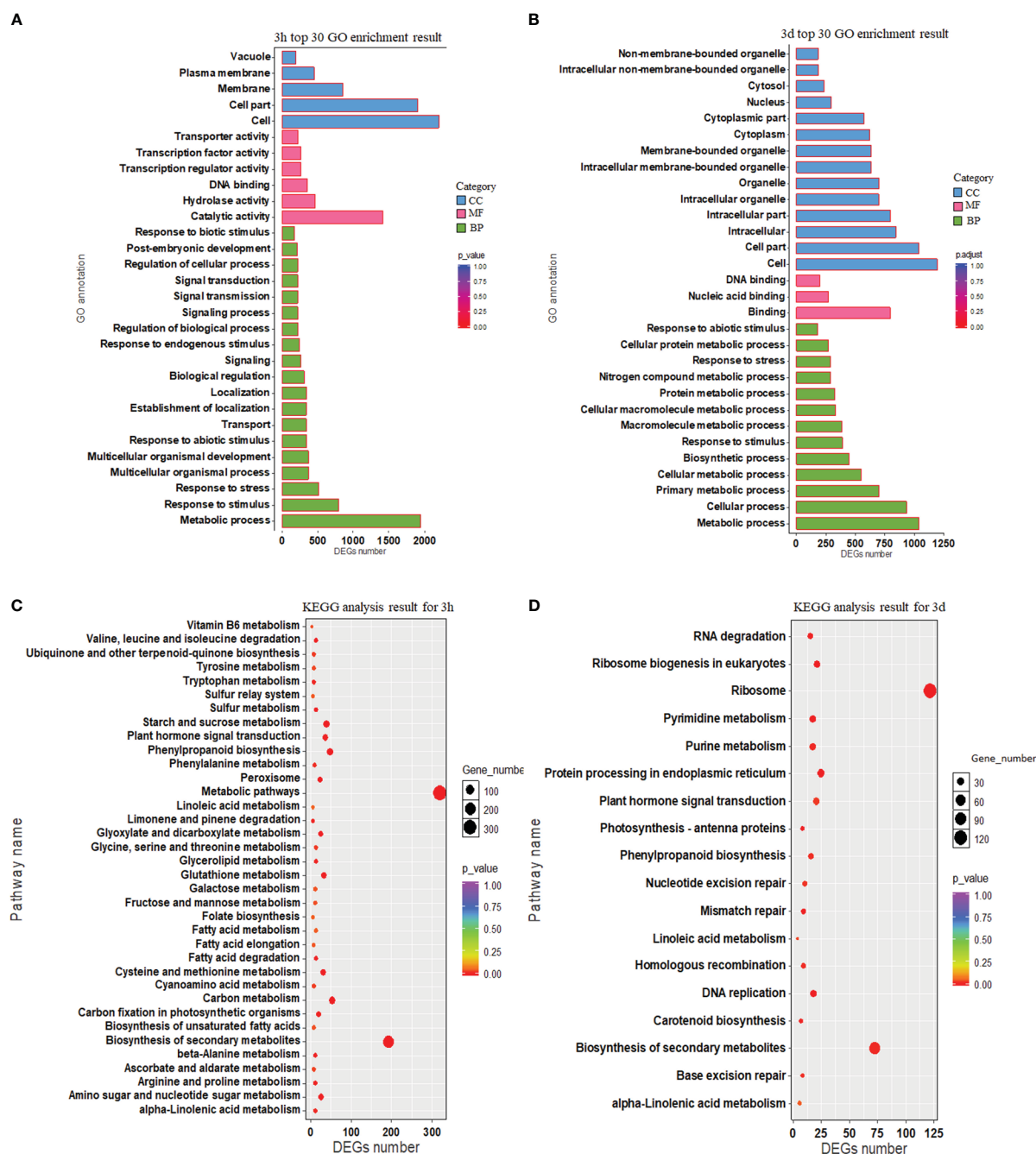


FIGURE 3

The functional annotation of rice DEGs. (A, B) Gene ontology (GO) annotation of DEGs for 3h and 3d. The vertical axis represents functional annotation information and the horizontal axis represents the number of differential genes annotated to the function. (C, D) KEGG pathway enrichment presented using a scatter plot for 3h and 3d. The vertical axis and horizontal axis represented by pathway name and DEGs number, respectively. The p -value ranges from 0 to 1 and the p -value is represented by the color of the dot; the smaller the p -value, the closer the color is to red. The relative number of DEGs contained under each function is denoted by the size of the dots.

KEGG at 3d (Figure 3D, Supplementary Table S4). The common pathways at the two-time points include “phenylpropanoid biosynthesis” (osa00940), “biosynthesis of secondary metabolites” (osa01110), and “plant hormone signal transduction” (osa04075).

Many genes have been identified in this study that are involved in metabolic and enzyme pathways in allelopathy interactions. Therefore, a secondary metabolic and large enzyme pathway of DEGs was investigated using MAPMAN. Results showed that the expression of genes associated with terpenoids, phenylpropanoids, simple phenols, lignin and lignans, and different flavonoid pathways were noticeably up-regulated at 3h (Figures 4A, B). Moreover, different enzymes, including Cytochrome P450, UDP

glycosyltransferase, glutathione-S-transferases, glucosidases, GDSL-lipases, Beta 1,3 glucan hydrolases, O-methyltransferases, peroxidases, and phosphatase were involved at the two-time points (Figures 5A, B).

Clustering of time-series expression profile

STEM was used to perform time-series expression profile clustering to exploration for shared temporal expression patterns based on our 5,684 DEGs. We identified four highly significant ($P < 0.05$) major temporal expression profiles that showed consistent

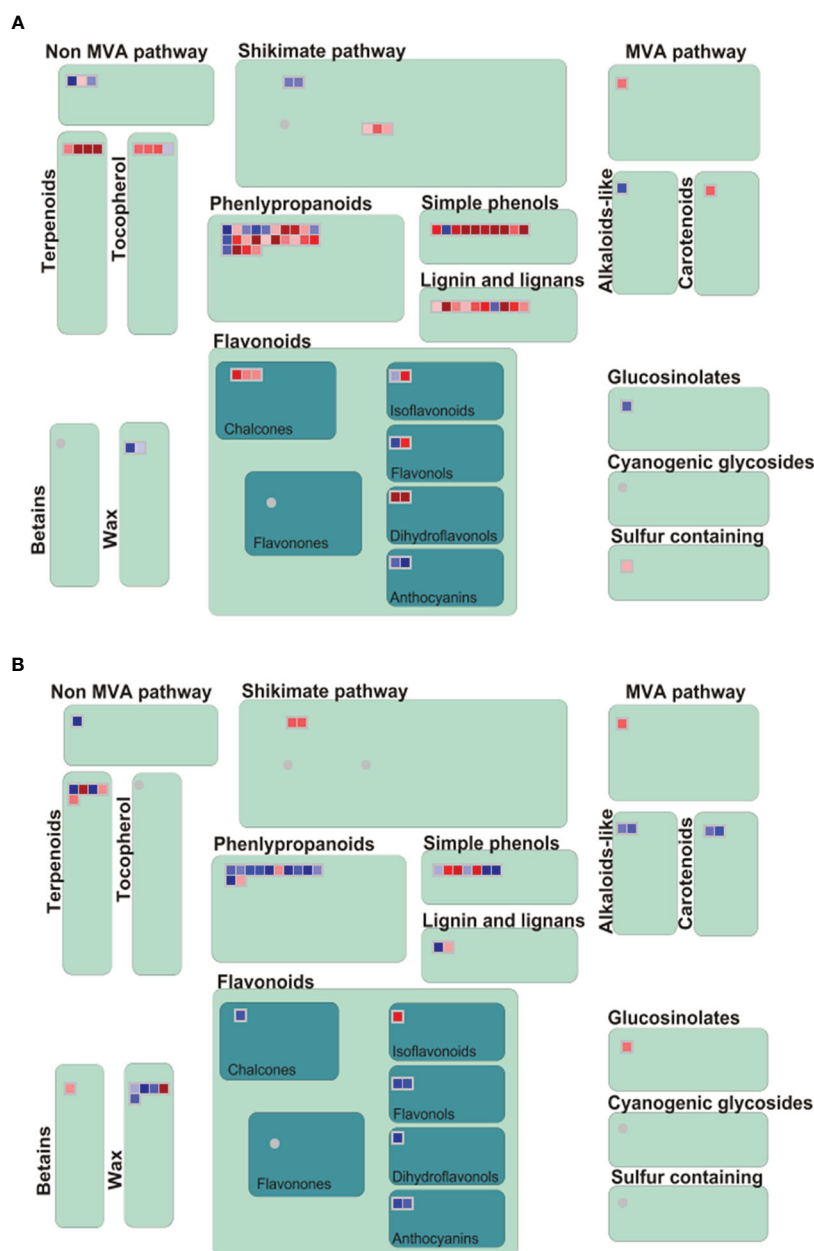


FIGURE 4

Analyses of secondary metabolism pathway of DEG using MAPMANs. (A, B) Secondary metabolism pathway analysis at 3h and 3d. Red and blue boxes indicate up-regulated and down-regulated genes.

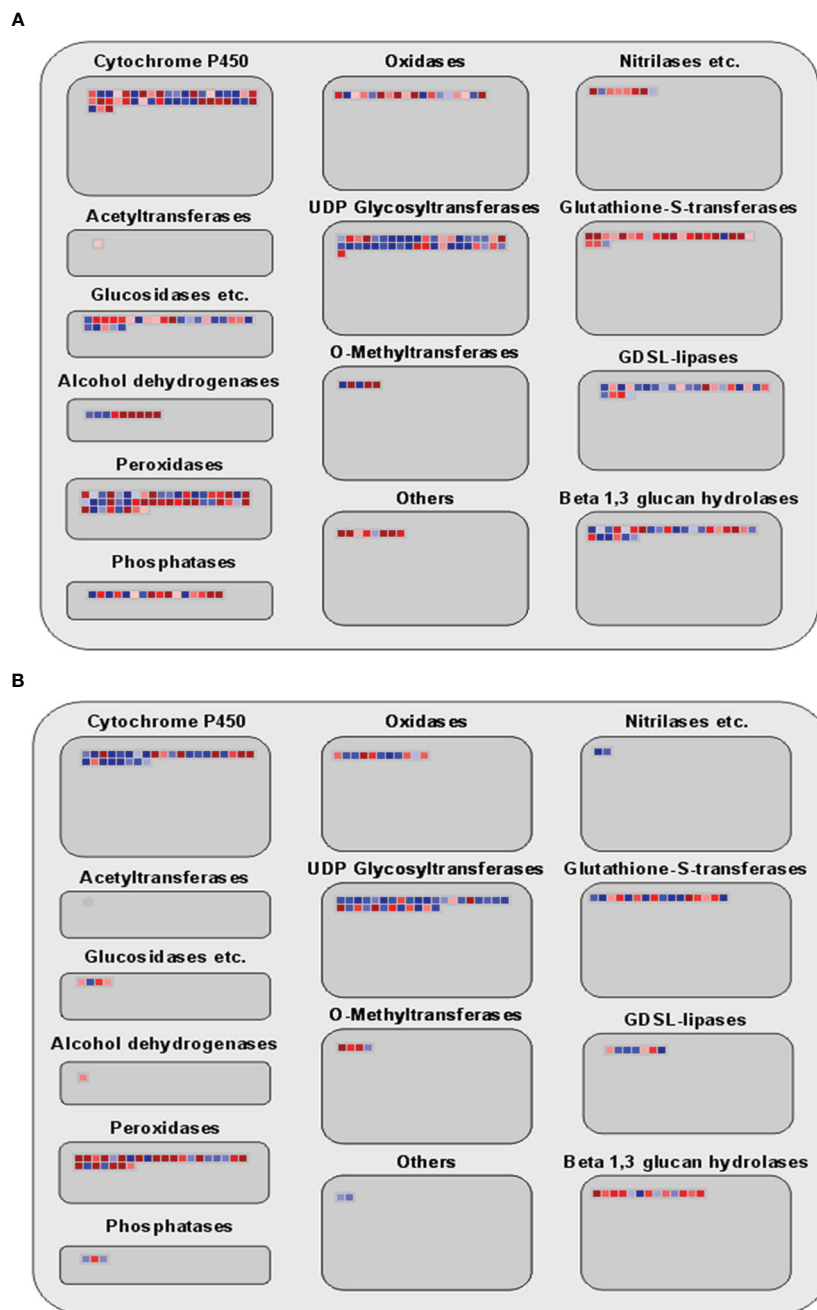


FIGURE 5

Analyses of enzyme family pathway of DEGs using MAPMAN. (A, B) Large enzyme family pathways of 3h and 3d time points. Red and blue boxes indicate up-regulated and down-regulated genes.

expression patterns in rice responsive to barnyardgrass co-cultivation (Figure 6A, Supplementary Table S5). Generally, four main expression cluster groups were identified.

In Profile 10 (404 genes), most of these genes' expression was up-regulated at 3h, but not changed at 3d, in rice responsive to barnyardgrass co-cultivation. In this pattern, genes were exclusively enriched in two pathways, namely glutathione metabolism (osa00480) and phenylpropanoid biosynthesis (osa00940) (Figure 6B, Supplementary Table S5). Profile 9 (360 genes) showed that the genes were up-regulated at 3h, followed by down-regulation at 3d, and the genes were enriched in two

pathways, including alpha-Linolenic acid metabolism (osa00592) and carotenoid biosynthesis (osa00906). Additionally, genes in both profiles were enriched in the biological process of "response to stimulus" (Figure 6B, Supplementary Table S5).

The other two profiles (256 genes in Profile 5 and 289 genes in Profile 6) consist of genes that are decreased at 3h. These genes were enriched in the "nitrogen compound metabolic process" and "development process" for GO terms and starch and sucrose metabolism (osa00500) for the KEGG pathway (Figure 6B, Supplementary Table S5).

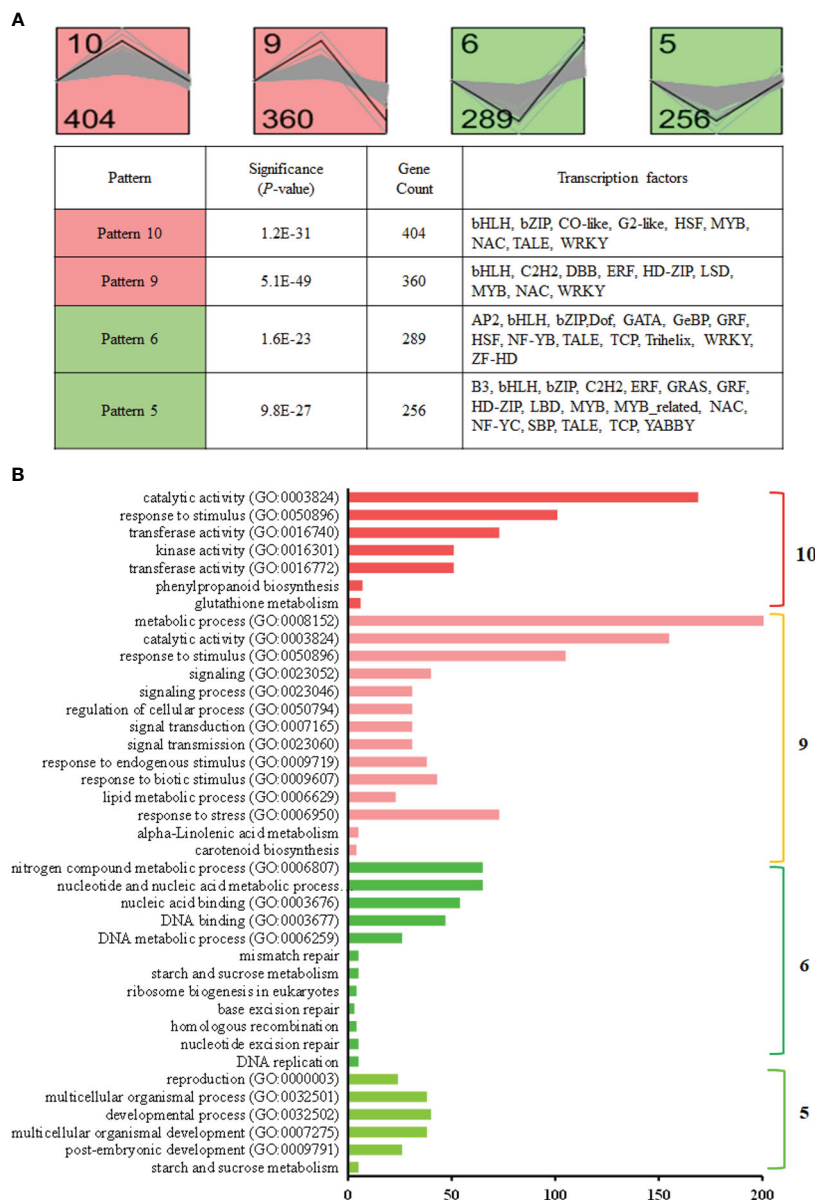


FIGURE 6

Dynamic change analysis of differentially co-expressed DEGs. **(A)** Protein clusters identified using the STEM algorithm that demonstrates coherent changes during 3h and 3d time points. Top panel: Model profiles based on fold change values over time for 3h and 3d. Significant profiles are highlighted in different colors. The black line indicates the fitted profile; gray lines are individual gene profiles. The top left number indicates the model profile number; the bottom left number indicates the number of genes contained in the profile. Bottom panel: the significance (*p*-value), gene counts, and transcription factors in the four representative gene expression patterns. **(B)** Functional enrichment based on gene patterns for significant profiles.

Comparison of DEGs in rice and barnyardgrass under co-cultivation

Previously, we have obtained transcriptomic data of barnyardgrass co-cultured with rice at 3h (Guo et al., 2017). A total of 4,945 DEGs, including 2,534 up-regulated and 2,411 down-regulated genes, were discovered in barnyardgrass. Here we have identified 396 homologous DEGs; 151 genes are unique (without duplicate) between rice and barnyardgrass at the 3h time point. Using a heat map, we have represented the homologous DEGs' expression values (Supplementary Figure S2). Results clearly showed a distinct expression pattern between rice and barnyardgrass.

Expression patterns of rice allelochemical genes

Two important pathways, namely momilactone- and phenolic acid-related genes, were involved during rice and barnyardgrass interaction. We have investigated five previously identified momilactone genes. Results showed that *OsCPS4* (*LOC_Os04g09900*), *OsKSL4* (*LOC_Os04g10060*), and *CYP99A3* (*LOC_Os04g10160*) are up-regulated in both 3h and 3d time points. *CYP99A2* (*LOC_Os04g09920*) is up-regulated at 3h but down-regulated at 3d, and *OsMAS* (*LOC_Os04g10010*) is down-regulated at both time points (Figure 7A). The expressions of these

genes in rice were proposed to respond to the allelopathic interaction between rice and barnyardgrass.

For the phenolic acid pathway, we found two PAL (phenylalanine ammonia-lyase) up-regulated DEGs, *LOC_Os05g35290* and *LOC_Os02g41670*; four CAD (cinnamyl-alcohol dehydrogenase) genes where *LOC_Os04g15920* is down-regulated and *LOC_Os11g40690*, *LOC_Os02g09490*, and *LOC_Os09g23530* up-regulated at DEGs at 3h time points in our analysis. Four up-regulated 4CL (4-coumaroyl CoA ligase) DEGs, *LOC_Os02g08100*, *LOC_Os06g44620*, *LOC_Os08g34790*, and *LOC_Os08g14760* were found at 3h which is an important branch point resulting to the group of flavonoids, lignins, and lignans. A total of seven DEGs were found in myricetin, kaempferol, quercetin, and catechin at both time points (Figure 7B). The gene expression results suggest that both momilactone and phenolic acid pathways are related to allelopathy, and it is difficult to pinpoint which is more important.

We have collected 55 genes or clones of previously identified allelopathy-associated genes from different rice varieties (Amb and

Ahluwalia, 2016) and used blast for these protein sequences with our identified DEGs' proteins to identify the rice allelochemical-related genes in this study. A total of 286 homolog DEGs of the putative allelopathy-associated genes were identified at 3h and 3d time points in this study. Among them, we have found significantly more genes at 3h than 3d, i.e., 201 (150 up- and 51 down-regulated) and 85 (62 up- and 23 down-regulated) genes at 3h and 3d, respectively, were identified (Supplementary Table S6).

qRT-PCR analysis results for RNA-Seq data validation

We have randomly used 17 genes to confirm the RNA-Seq data for 3h and 3d time points. The qRT-PCR results suggested that all studied genes displayed similar expression trends to the RNA-Seq results (Table 2). As the many common genes (393 genes) were involved in rice and barnyardgrass interactions, we investigated the

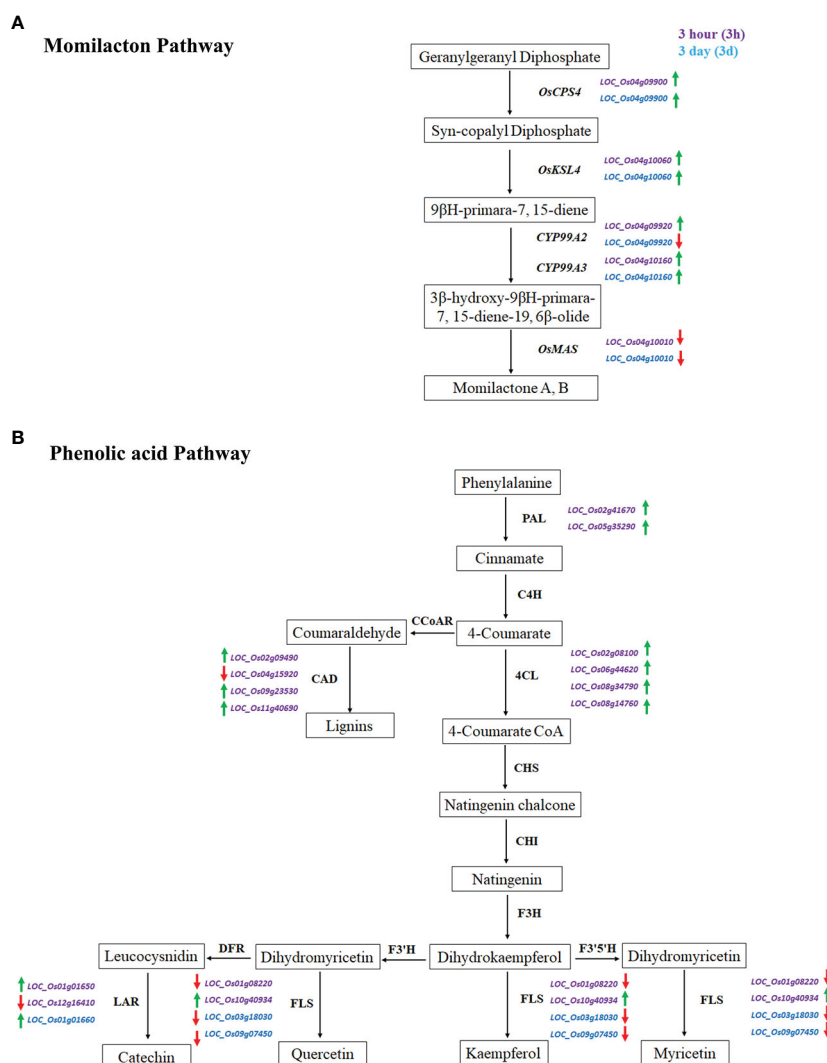


FIGURE 7

Expression of momilactone and phenolic acid pathways related genes during the rice and barnyardgrass interaction. (A) Momilactone pathway and (B) Phenolic acid pathway. Up-regulated and down-regulated genes denoted by green and red arrows, respectively.

eight common genes for 3h and 3d time points (Table 2). All the genes were up-regulated at 3h, but *LOC_Os12g25490* (O-methyltransferase) and *LOC_Os01g43750* (cytochrome P450) were down-regulated at 3d time point, which is consistent with RNA-Seq results. Three momilactone-related genes showed similar results, i.e., up-regulation at both 3h and 3d. Furthermore, four up-regulated genes involved in phenolic acid pathways at 3h were investigated by qRT-PCR. Similar patterns were observed between RNA-Seq expression results and qRT-PCR results. One gene (*LOC_Os01g01660*) was up-regulated and another gene was down-regulated (*LOC_Os03g18030*) for the two phenolic acid pathway-related genes at 3d (as compared with control) for the rice and barnyardgrass interaction (Table 2).

Discussion

Major challenges in maximizing crop yields involve weeds, and allelopathy is a favorable technique for weed regulation worldwide (Fang et al., 2013). Transcriptomics research has been done using the NGS supported RNA-Seq technique—a reliable tool for exploring the function of genes in various tissues and environments (Chen et al., 2011). In this study, we have performed RNA-Seq analyses to explore the changes in the transcriptome of rice co-cultured with barnyardgrass at the seedling stage. We have identified a total of 5684 DEGs at two-time points, with significantly more DEGs at 3 hours than 3 days, suggesting a quick allelopathic response in rice (Figure 1). This is also true for known allelopathy-related genes, for which

significantly more genes at 3h (201) than 3d (85) were DEGs (Supplementary Table S6). We also found 84 common up-regulated genes for the two time-periods, suggesting that these genes may be continuously leading the allelopathic response to barnyardgrass.

Research showed that the transcription factor was regulating the gene expression by controlling transcription beginning speed, where appropriate gene expression for each organism is very significant for their enlargement and maturity (Corrêa et al., 2008). TFs have a significant function for controlling plant propagation, maturation, and react to unfavorable situation conditions, including drought, chill, salinity, and high temperature (Zhai et al., 2013). Studies showed that different types of TFs play essential roles in various functional pathways such as bHLH involved in phytochrome signaling activity (Zhang et al., 2009), ERF involved diverse responses to environmental stimuli (Nakano et al., 2006), NAC involved in various mechanisms including developmental process (Nuruzzaman et al., 2010), and MYB participated in the control of anthocyanin biosynthesis (Ambawat et al., 2013). We have identified a total of 388 genes were related to 44 different TFs families, where most of them were allocated in the bHLH, ERF, WRKY, NAC, MYB, and bZIP (Figure 2A). Specifically, more TF-encoding genes were down-regulated (207 TFs) than up-regulated (181 TFs) in rice and barnyardgrass interaction (Figure 2B, Supplementary Table S2). From the TFs analysis results, we assumed that these TFs might be involved in rice and barnyardgrass interaction.

We found DEGs enriched in stress response-associated GO terms, such as “response to stimulus”, “response to biotic stimulus”

TABLE 2 RNA-Seq data validation using qRT-PCR.

| | 3h time point | | | 3d time point | | |
|---------------------------------------|----------------|-------------------------------|-------------------------------|----------------|-------------------------------|-------------------------------|
| | Gene Name | RNA_seq log ₂ (FC) | qRT-PCR log ₂ (FC) | Gene Name | RNA_seq log ₂ (FC) | qRT-PCR log ₂ (FC) |
| Common genes in 3h and 3d time points | LOC_Os03g13210 | 2.06 | 2.56 | LOC_Os03g13210 | 0.66 | 1.32 |
| | LOC_Os12g25490 | 1.47 | 1.65 | LOC_Os12g25490 | -0.56 | -0.24 |
| | LOC_Os03g56090 | 1.02 | 1.63 | LOC_Os03g56090 | 0.86 | 0.98 |
| | LOC_Os05g20930 | 0.9 | 1.3 | LOC_Os05g20930 | 1.95 | 1.75 |
| | LOC_Os09g35030 | 1.02 | 2.04 | LOC_Os09g35030 | 0.73 | 1.63 |
| | LOC_Os10g17260 | 1.14 | 1.57 | LOC_Os10g17260 | 0.59 | 0.25 |
| | LOC_Os06g37300 | 0.85 | 1.17 | LOC_Os06g37300 | 1.26 | 1.18 |
| | LOC_Os01g43750 | 0.92 | 2.85 | LOC_Os01g43750 | -0.93 | -2.81 |
| Momilactone pathway related genes | LOC_Os04g10160 | 1.06 | 2.71 | LOC_Os04g10160 | 0.62 | 0.4 |
| | LOC_Os04g10060 | 1.64 | 2.07 | LOC_Os04g10060 | 0.99 | 0.61 |
| | LOC_Os04g09900 | 1.27 | 1.67 | LOC_Os04g09900 | 0.12 | 0.19 |
| Phenolic acid pathway related genes | LOC_Os05g35290 | 0.39 | 0.45 | LOC_Os01g01660 | 0.87 | 0.48 |
| | LOC_Os02g41670 | 1.13 | 1.19 | LOC_Os03g18030 | -0.91 | -1.1 |
| | LOC_Os11g40690 | 1.72 | 1.55 | | | |
| | LOC_Os01g01650 | 0.86 | 0.76 | | | |

and “response to stress” which was expected as allelopathy interactions are a kind of stress to plants. Four significant temporal expression profiles showing consistent expression patterns in rice responsive to barnyardgrass co-cultivation have been identified, for which two profiles (Profiles 5 and 6) consist of genes that are decreased at 3h (Figure 6, Supplementary Table S5). Interestingly, these down-regulated genes were enriched in the “nitrogen compound metabolic process” and “development process” for GO terms and starch and sucrose metabolism (osa00500) for the KEGG pathway. This suggests a balance between growth and stress (allelopathy) response in rice.

Genes related to phenylpropanoid metabolism might increase the synthesis and discharge of allelochemicals and inhibit weeds (Fang et al., 2010). A recent study demonstrated that the DEGs were connected with phenylpropanoid biosynthesis, phenylalanine metabolism, and tyrosine biosynthesis in both allelopathic and non-allelopathic rice against control (Zhang et al., 2018). The KEGG pathway investigation exhibited that most genes were involved in “metabolic pathways”, followed by “biosynthesis of secondary metabolites”, “carbon metabolism”, “phenylpropanoid biosynthesis”, “starch and sucrose metabolism”, “plant hormone signal transduction”, and “glutathione metabolism” in this study (Figures 3C, D). KEGG results indicated that various pathways, including “phenylpropanoid biosynthesis”, might play a crucial function in controlling rice and barnyardgrass interactions.

Previous studies described the flavones, fatty acids, phenolic acids, steroids, and terpenoids are the rice allelochemicals (Kong et al., 2004; Seal et al., 2004; Macías et al., 2006; He et al., 2012; Kato-Noguchi and Peters, 2013; Zhang et al., 2018). Cytochrome P450 (P450s) are extensive in plant genomes, and the function of P450s are distributed in diverse biochemical pathway to create primary and secondary metabolites, including lipids, phenylpropanoids, terpenoids, and alkaloids (Mizutani and Ohta, 2010). Research showed that P450 CYP93G2 was the main enzyme for establishing C-glycosylflavones from flavanones in rice (Du et al., 2010). Glutathione S-transferase (GST), a broad category of main defense enzymes, is essential in protecting cells from various abiotic and biotic stress, including xenobiotic and heavy-metal perniciousness, oxidative stress, and pathogen assault in plants (Soranzo et al., 2004; Peng et al., 2023). A recent study showed that GST is essential for glutathione-bisphenol A mating and use of exogenous dopamine to plant enhanced glutathione levels resulting expands organic contaminant detoxification and stress resistance of cucumber seedlings (Ahmed and Li, 2023). Several studies demonstrated that the rice allelopathy is an inherited attribute connected with the molecular mediation of secondary metabolic pathways (Bi et al., 2007; Song et al., 2008; Fang et al., 2009). MapMan software was used along through secondary metabolism and enzyme pathway analysis in this study. Our results did show that genes involved in various secondary metabolic pathways, including terpenoids, phenylpropanoids, simple phenols, lignin and lignans, and different flavonoids pathways, were noticeably up-regulated in this study (Figure 4). Also, different enzymes, including cytochrome P450, glutathione-S-transferases, and glucosidases, were involved at 3h and 3d time points (Figure 5).

These findings indicated that allelopathic interaction is a complicated process between plants instead of a single critical pathway. Additionally, we compared the DEGs between rice and barnyardgrass and showed that very few genes were common. Particularly, expression patterns of common DEGs are not similar. These results demonstrated different mechanisms underlying the allelopathy interaction between the two species.

In summary, RNA-Seq data used to detect the crucial DEGs, gene functional analysis, KEGG pathways analysis, transcription factors related to rice allelopathy, and allelopathy related genes and their molecular mechanisms that impact of weeds on rice in this study. We supposed that the momilactone pathway might be more sensitive against barnyardgrass than other phenolic acid pathways from the expression pattern of rice allelochemical genes. The outcomes of the present study might provide more straight proof and evidence for the forthcoming study on the interaction between rice and barnyardgrass. We believe that this study could provide a valuable genetic resource for rice and barnyardgrass associated candidate allelopathy genes and should be useful for controlling weeds, which would outcome in the development of agriculture. However, the detailed roles of the identified genes related to allelopathy controlling weeds could be needed for further functional analysis by genetic approaches involved in rice and barnyardgrass interaction. Also, this study was inspected in the lab, which unable to scrutinize the interaction of rice and barnyardgrass in the natural field condition. Moreover, genome-wide association study (GWAS) is a well-know and widely used method for dissecting complex traits in plants (Yang et al., 2014; Alamin et al., 2022). GWAS could be utilized to identify the important variants associated with the interaction of rice and barnyardgrass. Furthermore, candidate genes associated with the interaction of rice and barnyardgrass could be identified by the expression of quantitative trait loci analysis.

Data availability statement

The data presented in the study are deposited in the BioProject repository, accession number PRJNA645506.

Author contributions

Conceptualization: CY and LF. Formal analysis: MS and MA. Validation: MS, MA, and CY. Writing-original draft: MS and MA. Writing-review and editing: CY, JQ, and LF. Funding acquisition: CY and LF. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation (32170621) and the Zhejiang Natural Science Foundation of China (Grant No. LZ17C130001).

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.2023.1104951/full#supplementary-material>

References

- Ahammed, G. J., and Li, X. (2022). Hormonal regulation of health-promoting compounds in tea (*Camellia sinensis* L.). *Plant Physiol. Biochem.* 185, 390–400. doi: 10.1016/j.plaphy.2022.06.021
- Ahammed, G. J., and Li, X. (2023). Dopamine-induced abiotic stress tolerance in horticultural plants. *Scientia Hort.* 307, 111506. doi: 10.1016/j.scienta.2022.111506
- Alamin, M., Sultana, M. H., Lou, X., Jin, W., and Xu, H. (2022). Dissecting complex traits using omics data: A review on the linear mixed models and their application in GWAS. *Plants* 11, 3277. doi: 10.3390/plants11233277
- Alamin, M., Zeng, D.-D., Sultana, M. H., Qin, R., Jin, X.-L., and Shi, C.-H. (2018). Photosynthesis, cellulose contents and ultrastructure changes of mutant rice leading to screw flag leaf. *Plant Growth Regul.* 85, 1–13. doi: 10.1007/s10725-018-0369-5
- Amb, M., and Ahluwalia, A. (2016). Allelopathy: potential role to achieve new milestones in rice cultivation. *Rice Sci.* 23, 165–183. doi: 10.1016/j.rsci.2016.06.001
- Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* 19, 307–321. doi: 10.1007/s12298-013-0179-1
- Bi, H. H., Zeng, R. S., Su, L. M., An, M., and Luo, S. M. (2007). Rice allelopathy induced by methyl jasmonate and methyl salicylate. *J. Chem. Ecol.* 33, 1089–1103. doi: 10.1007/s10886-007-9286-1
- Chen, G., Wang, C., and Shi, T. (2011). Overview of available methods for diverse RNA-seq data analyses. *Sci. China Life Sci.* 54, 1121–1128. doi: 10.1007/s11427-011-4255-x
- Cheng, F., and Cheng, Z. (2015). Research progress on the use of plant allelopathy in agriculture and the physiological and ecological mechanisms of allelopathy. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01020
- Cheng, A. X., Lou, Y. G., Mao, Y. B., Lu, S., Wang, L. J., and Chen, X. Y. (2007). Plant terpenoids: biosynthesis and ecological functions. *J. Integr. Plant Biol.* 49, 179–186. doi: 10.1111/j.1744-7909.2007.00395.x
- Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 38, 1767–1771. doi: 10.1093/nar/gkp1137
- Corrêa, L. G. G., Riaño-Pachón, D. M., Schrago, C. G., Dos Santos, R. V., Mueller-Roeber, B., and Vincent, M. (2008). The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PLoS One* 3(8): e2944. doi: 10.1371/journal.pone.0002944
- Du, Y., Chu, H., Chu, I. K., and Lo, C. (2010). CYP93G2 is a flavanone 2-hydroxylase required for c-glycosylflavone biosynthesis in rice. *Plant Physiol.* 154, 324–333. doi: 10.1104/pp.110.161042
- Dudareva, N., Pichersky, E., and Gershenzon, J. (2004). Biochemistry of plant volatiles. *Plant Physiol.* 135, 1893–1902. doi: 10.1104/pp.104.049981
- Duke, S. O., and Gressel, J. (2010). Weed genomics advance: a commentary. *Pest Manag. Sci.* 66, 1041–1041. doi: 10.1002/ps.1980
- Edwards, D., and Batley, J. (2010). Plant genome sequencing: applications for crop improvement. *Plant Biotechnol. J.* 8, 2–9. doi: 10.1111/j.1467-7652.2009.00459.x
- Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinf.* 7, 191. doi: 10.1186/1471-2105-7-191
- Fang, C.-X., He, H.-B., Wang, Q.-S., Qiu, L., Wang, H.-B., Zhuang, Y.-E., et al. (2010). Genomic analysis of allelopathic response to low nitrogen and barnyardgrass competition in rice (*Oryza sativa* L.). *Plant Growth Regul.* 61, 277–286. doi: 10.1007/s10725-010-9475-8
- Fang, C.-X., Xiong, J., Qiu, L., Wang, H.-B., Song, B.-Q., He, H.-B., et al. (2009). Analysis of gene expressions associated with increased allelopathy in rice (*Oryza sativa* L.) induced by exogenous salicylic acid. *Plant Growth Regul.* 57, 163. doi: 10.1007/s10725-008-9333-0
- Fang, C., Zhuang, Y., Xu, T., Li, Y., Li, Y., and Lin, W. (2013). Changes in rice allelopathy and rhizosphere microflora by inhibiting rice phenylalanine ammonia-lyase gene expression. *J. Chem. Ecol.* 39, 204–212. doi: 10.1007/s10886-013-0249-4
- Ferrer, J. L., Austin, M. B., Stewart, C. Jr., and Noel, J. P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.* 46, 356–370. doi: 10.1016/j.plaphy.2007.12.009
- Godfray, H. C., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383
- Guo, L., Qiu, J., Ye, C., Jin, G., Mao, L., Zhang, H., et al. (2017). Echinochloa crus-galli genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat. Commun.* 8, 1031. doi: 10.1038/s41467-017-01067-5
- He, H., Wang, H., Fang, C., Wu, H., Guo, X., Liu, C., et al. (2012). Barnyard grass stress up regulates the biosynthesis of phenolic compounds in allelopathic rice. *J. Plant Physiol.* 169, 1747–1753. doi: 10.1016/j.jplph.2012.06.018
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Kato-Noguchi, H. (2011). Barnyard grass-induced rice allelopathy and momilactone b. *J. Plant Physiol.* 168, 1016–1020. doi: 10.1016/j.jplph.2010.12.021
- Kato-Noguchi, H., and Peters, R. J. (2013). The role of momilactones in rice allelopathy. *J. Chem. Ecol.* 39, 175–185. doi: 10.1007/s10886-013-0236-9
- Khanh, T., Chung, M., Xuan, T., and Tawata, S. (2005). The exploitation of crop allelopathy in sustainable agricultural production. *J. Agron. Crop Sci.* 191, 172–184. doi: 10.1111/j.1439-037X.2005.00172.x
- Khanh, T., Xuan, T., and Chung, I. (2007). Rice allelopathy and the possibility for weed management. *Ann. Appl. Biol.* 151, 325–339. doi: 10.1111/j.1744-7348.2007.00183.x
- Khush, G. S. (2005). What it will take to feed 5.0 billion rice consumers in 2030. *Plant Mol. Biol.* 59, 1–6. doi: 10.1007/s11103-005-2159-5
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. doi: 10.1186/gb-2013-14-4-r36
- Kong, C. (2007). Allelochemicals involved in rice allelopathy. in *Allelopathy: New Concepts and Methodology*, 267–281.
- Kong, C., Xu, X., Liang, W., Zhou, Y., and Hu, F. (2004). Non-phenolic allelochemicals in root exudates of an allelopathic rice variety and their identification and weed-suppressive activity. *Acta Ecologica Sin.* 24, 1317–1322.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi: 10.1186/gb-2009-10-3-r25
- 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
- Lundkvist, A., and Verwijst, T. (2011). “Weed biology and weed management in organic farming,” in *Research in organic farming*, 10–41.
- Macías, F. A., Chinchilla, N., Varela, R. M., and Molinillo, J. M. (2006). Bioactive steroids from *Oryza sativa* L. *Steroids* 71, 603–608. doi: 10.1016/j.steroids.2006.03.001

- Mennan, H., Ngouajio, M., Sahin, M., Isik, D., and Kaya Altop, E. (2012). Quantification of momilactone b in rice hulls and the phytotoxic potential of rice extracts on the seed germination of alisma plantago-aquatica. *Weed Biol. Manage.* 12, 29–39. doi: 10.1111/j.1445-6664.2012.00433.x
- Mizutani, M., and Ohta, D. (2010). Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* 61, 291–315. doi: 10.1146/annurev-arplant-042809-112305
- Mochida, K., and Shinozaki, K. (2010). Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol.* 51, 497–523. doi: 10.1093/pcp/pcq027
- Nakano, T., Suzuki, K., Fujimura, T., and Shinshi, H. (2006). Genome-wide analysis of the ERF gene family in arabidopsis and rice. *Plant Physiol.* 140, 411–432. doi: 10.1104/pp.105.073783
- Navarez, D., and Olofsson, M. (1996). Relay seeding technique for screening allelopathic rice (*Oryza sativa*). *Proc. 2nd Int. Weed Control Congr., Copenhagen, Enfield, New Hampshire, United States.* 1285–1290.
- Nuruzzaman, M., Manimekalai, R., Sharoni, A. M., Satoh, K., Kondoh, H., Ooka, H., et al. (2010). Genome-wide analysis of NAC transcription factor family in rice. *Gene* 465, 30–44. doi: 10.1016/j.gene.2010.06.008
- Oerke, E.-C. (2006). Crop losses to pests. *J. Agric. Sci.* 144, 31–43. doi: 10.1017/S0021859605005708
- Oerke, E. C., and Dehne, H. W. (2004). Safeguarding production—losses in major crops and the role of crop protection. *Crop Prot.* 23, 275–285. doi: 10.1016/j.cropro.2003.10.001
- Patel, R. K., and Jain, M. (2012). NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PloS One* 7, e30619. doi: 10.1371/journal.pone.0030619
- Peng, X., Wang, N., Sun, S., Geng, L., Guo, N., Liu, A., et al. (2023). Reactive oxygen species signaling is involved in melatonin-induced reduction of chlorothalonil residue in tomato leaves. *J. Hazardous Materials* 443, 130212. doi: 10.1016/j.jhazmat.2022.130212
- Rao, V. S. (2000). *Principles of weed science* (Enfield, New Hampshire 03748, USA: CRC Press).
- Rezaeieh, A. D., Aminpanah, H., and Sadeghi, S. M. (2015). Competition between rice (*Oryza sativa* L.) and (barnyardgrass (*Echinochloa crus-galli* (L.) p. Beauv.) as affected by methanol foliar application. *Anais da Academia Bras. Ciências* 87, 879–890. doi: 10.1590/0001-3765201520140413
- Seal, A. N., Pratley, J. E., Haig, T., and An, M. (2004). Identification and quantitation of compounds in a series of allelopathic and non-allelopathic rice root exudates. *J. Chem. Ecol.* 30, 1647–1662. doi: 10.1023/B:JOEC.0000042074.96036.14
- Shah, A., and Smith, D. L. (2020). Flavonoids in agriculture: Chemistry and roles in, biotic and abiotic stress responses, and microbial associations. *Agronomy* 10, 1209. doi: 10.3390/agronomy10081209
- Singh, P., Arif, Y., Bajguz, A., and Hayat, S. (2021). The role of quercetin in plants. *Plant Physiol. Biochem.* 166, 10–19. doi: 10.1016/j.plaphy.2021.05.023
- Song, B., Xiong, J., Fang, C., Qiu, L., Lin, R., Liang, Y., et al. (2008). Allelopathic enhancement and differential gene expression in rice under low nitrogen treatment. *J. Chem. Ecol.* 34, 688–695. doi: 10.1007/s10886-008-9455-x
- Soranzo, N., Gorla, M. S., Mizzi, L., De Toma, G., and Frova, C. (2004). Organisation and structural evolution of the rice glutathione s-transferase gene family. *Mol. Genet. Genomics* 271, 511–521. doi: 10.1007/s00438-004-1006-8
- Sultana, M. H., Liu, F., Alamin, M., Mao, L., Jia, L., Chen, H., et al. (2019). Gene modules Co-regulated with biosynthetic gene clusters for allelopathy between rice and barnyardgrass. *Int. J. Mol. Sci.* 20, 3846. doi: 10.3390/ijms20163846
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community 2017 update. *Nucleic Acids Res.* 45, W122–W129.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiol.* 153, 895–905. doi: 10.1104/pp.110.155119
- Varshney, R. K., Bansal, K. C., Aggarwal, P. K., Datta, S. K., and Craufurd, P. Q. (2011). Agricultural biotechnology for crop improvement in a variable climate: hope or hype? *Trends Plant Sci.* 16, 363–371. doi: 10.1016/j.tplants.2011.03.004
- 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.
- Xu, M., Galhano, R., Wiemann, P., Bueno, E., Tiernan, M., Wu, W., et al. (2012). Genetic evidence for natural product-mediated plant-plant allelopathy in rice (*Oryza sativa*). *New Phytol.* 193, 570–575. doi: 10.1111/j.1469-8137.2011.04005.x
- Xuan, T. D., Chung, I. M., Khanh, T. D., and Tawata, S. (2006). Identification of phytotoxic substances from early growth of barnyard grass (*Echinochloa crus-galli*) root exudates. *J. Chem. Ecol.* 32, 895–906. doi: 10.1007/s10886-006-9035-x
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5, 5087. doi: 10.1038/ncomms6087
- Zhai, R., Feng, Y., Wang, H., Zhan, X., Shen, X., Wu, W., et al. (2013). Transcriptome analysis of rice root heterosis by RNA-seq. *BMC Genomics* 14, 19. doi: 10.1186/1471-2164-14-19
- Zhang, L.-Y., Bai, M.-Y., Wu, J., Zhu, J.-Y., Wang, H., Zhang, Z., et al. (2009). Antagonistic HLH/bHLH transcription factors mediate brassinosteroid regulation of cell elongation and plant development in rice and arabidopsis. *Plant Cell* 21, 3767–3780. doi: 10.1105/tpc.109.070441
- Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., et al. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20, 646–654. doi: 10.1101/gr.100677.109
- Zhang, Q., Li, L., Li, J., Wang, H., Fang, C., Yang, X., et al. (2018). Increasing rice allelopathy by induction of barnyard grass (*Echinochloa crus-galli*) root exudates. *J. Plant Growth Regul.* 37, 745–754. doi: 10.1007/s00344-017-9770-y



OPEN ACCESS

EDITED BY

Hairul Roslan,
Universiti Malaysia Sarawak, Malaysia

REVIEWED BY

Qiuming Yao,
University of Nebraska-Lincoln,
United States
Zhonglin Mou,
University of Florida, United States
Chonghui Ji,
Fujian Agriculture and Forestry University,
China

*CORRESPONDENCE

Bingyu Zhao
✉ bzha07@vt.edu
Bo Zhang
✉ zhang76@vt.edu

SPECIALTY SECTION

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

RECEIVED 29 November 2022

ACCEPTED 25 January 2023

PUBLISHED 20 February 2023

CITATION

Wang Z, Shea Z, Li Q, Wang K, Mills K,
Zhang B and Zhao B (2023) Evaluate the
guide RNA effectiveness via
Agrobacterium-mediated transient assays
in *Nicotiana benthamiana*.
Front. Plant Sci. 14:1111683.
doi: 10.3389/fpls.2023.1111683

COPYRIGHT

© 2023 Wang, Shea, Li, Wang, Mills, Zhang
and Zhao. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Evaluate the guide RNA effectiveness via *Agrobacterium*-mediated transient assays in *Nicotiana benthamiana*

Zhibo Wang, Zachary Shea, Qi Li, Kunru Wang, Kerri Mills,
Bo Zhang* and Bingyu Zhao*

School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States

CRISPR/Cas9-based genome editing system is a powerful tool for plant genetic improvement. However, the variable efficiency of guide RNA(s) (gRNA) represents a key limiting factor that hampers the broad application of the CRISPR/Cas9 system in crop improvement. Here, we employed the *Agrobacterium*-mediated transient assays to evaluate the effectiveness of gRNAs for editing genes in *Nicotiana benthamiana* and soybean. We designed a facile screening system based on indels that can be introduced by CRISPR/Cas9-mediated gene editing. A gRNA binding sequence (23 nucleotides) was inserted into the open reading frame of yellow fluorescent protein (YFP) gene (gRNA-YFP), which disrupted the YFP reading frame and results in no fluorescent signal when it was expressed in plant cells. Transiently co-expression of Cas9 and a gRNA targeting the gRNA-YFP gene in plant cells could restore the YFP reading frame and recover the YFP signals. We evaluated five gRNAs targeting *Nicotiana benthamiana* and soybean genes and confirmed the reliability of the gRNA screening system. The effective gRNAs targeting *NbEDS1*, *NbWRKY70*, *GmKTI1*, and *GmKTI3* had been used to generate transgenic plants and resulted in expected mutations on each gene. While a gRNA targeting *NbNDR1* was confirmed to be ineffective in transient assays. This gRNA indeed failed to trigger target gene mutations in stable transgenic plants. Thus, this new transient assay system can be used to validate the effectiveness of gRNAs before generating stable transgenic plants.

KEYWORDS

CRISPR/Cas9, guide RNA (gRNA), *Agrobacterium*-mediated transient assay, *N. benthamiana*, soybean

Introduction

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Cas system has emerged as a powerful tool for gene editing in different organisms (Liu and Fan, 2014; Chen et al., 2019). It has been applied to improve the agronomic traits of many crop plant species (Chen et al., 2019; Wang et al., 2022 (under review)). New gene-editing tools, including base

editing and prime editing, are continuously developed in both academic and private sectors (Zong et al., 2022).

CRISPR/Cas9 machinery employs short guiding RNA (gRNA) that has an 18–22 nucleotide (nt) region complementary to a short fragment in the gene of interest (Wu et al., 2014; Matson et al., 2019). With the navigation by gRNA, endonuclease Cas9 or its orthologues can bind and cleave target DNAs and introduce mutations at a desired location (Abudayyeh et al., 2016; Kim et al., 2017). The success of CRISPR/Cas9-mediated gene editing depends on recognizing the target DNA controlled by RNA–DNA interaction in the CRISPR/Cas9 system. Although the choice of gRNA would seem to be unlimited, studies in various organisms indicate that selection of gRNA significantly affects the efficiency of CRISPR/Cas9 targeting and/or DNA cleavage at target loci (Doench et al., 2014; Liang et al., 2016). Recent studies suggest that the nucleotide composition, length, and secondary structures of gRNAs can impact the specificity and efficiency of gene editing (Dang et al., 2015). Therefore, selection of an effective gRNA is critical for a successful gene editing experiment. In fact, the effectiveness of a gRNA is pivotal for the prime editing systems, which currently have low gene-editing efficiency (Chen et al., 2021a). For some plant species, such as apple, pepper, soybean, etc., the development of transgenics can take years to obtain stable transgenic plants. Therefore, selection of functionally validated Cas9/gRNA transformation constructs before the initiation of plant transformation process, will significantly increase the productivity of CRISPR/Cas9-mediated gene editing in crop plants.

The computational and experimental attempts have both been documented to evaluate the effectiveness of gRNAs in CRISPR/Cas9-mediated genome editing (Liang et al., 2016; Liu et al., 2020a; Xiang et al., 2021; Konstantakos et al., 2022). For example, Liang et al. (2016) proposed that the G/C nucleotide content of gRNA sequences should be around 30%–80%. The potential secondary structures in gRNAs may also interfere with their target recognition (Liang et al., 2016). Konstantakos et al. (2022) summarized the currently known computational tools that can be used to select effective gRNAs. Thus far, only two methods have been developed for the experimental evaluation of gRNAs in plant cells. The first one is utilizing a protoplast-based transient assay (Carlsen et al., 2021). However, these kind of protocols need DNA amplification and enzyme digestion and/or sequencing assistance to validate the effectiveness of gRNAs, which is inconvenient and cost-prohibitive (Liang et al., 2016; Liu et al., 2020a). The second method is based on small insertions and deletions (indels) that can be introduced by CRISPR/Cas9-mediated gene editing, which is also the principle we employed in this study. In brief, Liang et al. (2019) developed a reporter construct that carries the gene of interest (GOI) fused with a site-specific endonuclease Csy4. A green fluorescent protein gene (GFP) that carries the Cys4 recognition site was cloned adjacent to the GOI–Csy4 Fusion gene. When there is no gene editing event, the GFP transcripts will be destroyed by Csy4. When a gene-editing event occurs on the GOI, small indels generated by CRISPR/Cas9 system have a 2/3 chance of stopping the expression of Csy4, therefore, preventing the degradation of GFP transcripts and resulting in increased GFP signals (Liang et al., 2019). However, this sophisticated system is not straightforward, which can be further simplified for high-throughput screening assays.

Small insertions and deletions can be introduced by the nonhomologous end-joining (NHEJ) during the process of CRISPR/Cas9-mediated gene editing (Kumar et al., 2019). Some INDELS cause a frameshift of the targeting gene at the gRNA binding site (Kumar et al., 2019), resulting in the loss of function of a target gene. On the other hand, certain INDELS can also correct the pre-existed frameshift mutation of a targeting gene, therefore, restore the gene functionality. Based on this mechanism, we developed a more convenient protocol to evaluate the effectiveness of gRNAs via *Agrobacterium*-mediated transient assays in *Nicotiana benthamiana*. The newly developed system can aid in the assessment of gRNA efficiency before the initiation of stable transformation. We validated the functionality and reliability of the screening system by testing gRNAs targeting three genes from *N. benthamiana* and two genes from soybean. This new system does not require sequencing of a target gene, and it is expected to be broadly applicable for gRNA validation in genome editing experiments.

Results and discussions

Two gRNAs targeting NbEDS1 and NbNDR1, respectively, have different effectiveness in terms of triggering the Cas9-mediated gene editing

The enhanced disease susceptibility 1 (*EDS1*) and nonrace-specific disease resistance 1 (*NDR1*) genes encode two key components of plant immunity, respectively, which are required for plant disease resistance in diverse plant species (Aarts et al., 1998; Coll et al., 2011; Knepper et al., 2011; Lapin et al., 2020). In a previous research effort, we aimed to develop *N. benthamiana* mutant plant that is impaired in both *NbEDS1* and *NbNDR1* genes. We manually designed two gRNAs targeting *NbEDS1* and *NbNDR1*, respectively, based on the location of the PAM sequence (NGG). They were cloned as a tandem array in one CRISPR/Cas9 construct (pCas9-*NbEDS1*/*NbNDR1* gRNA) and transformed into *N. benthamiana* plant (Figure S1 and Figure S2A). In this construct, a U6 promoter is assigned to express each gRNA (Figure S2A), thus expressions of two gRNAs are ensured for the editing on both *NbEDS1* and *NbNDR1* genes. More than twenty T1 transgenic plants were recovered and genotyped. The *NbEDS1* genes were successfully edited in all genotyped transgenic plants (Figure S2B), while no edited *NbNDR1* genes were detected in any genotyped transgenic plants, even in the T2 and T3 generations (Figure S2C). Interestingly, we noticed that the *NbNDR1* gRNA sequence has 70% AT; the low GC content may suggest it has a low binding affinity to the target gene (Liang et al., 2016; Graf et al., 2019). We speculated that the gRNA targeting *NbEDS1* is more effective than the one targeting *NbNDR1*.

Previous reports suggest that the mutation of *NbEDS1* can suppress *N. benthamiana* disease resistance to *Xanthomonas euvesicatoria* (*Xe*) (Schultink et al., 2017; Martin et al., 2020). We inoculated the pCas9-*NbEDS1*/*NbNDR1* gRNA transgenic plants with *Xe* strain *Xe85-10*. All tested transgenic plants, but not the wild-type plants, were susceptible to the infection of *Xe85-10*, resulting in an apparent water-soaking phenotype at the inoculation sites (Figure

S2D). *NbNDR1* is required for the Rps2 (resistance to *P. syringae*)/AvrRpt2-mediated disease resistance (Century et al., 1995; Aarts et al., 1998; Knepper et al., 2011; Chen et al., 2021b). Therefore, we also inoculated pCas9-*NbEDS1*/*NbNDR1* gRNA transgenic plants, along with wild-type controls, with *Agrobacterium tumefaciens* strains carrying the 35S::*RPS2* and 35S::*avrRpt2*. As expected, transient co-expression of *Rps2* and *avrRpt2* genes triggered a programmed cell death phenotype on both transgenic and wild-type plants (Figure S2E). This result suggests that *NbNDR1* was not mutated, which is consistent with the DNA sequencing result (Figure S2C). Thus, our data suggest that even though two gRNAs targeting *NbEDS1* and *NbNDR1*, respectively, were assembled into one transformation construct, the effectiveness of two gRNA-triggered gene editing is different, where only *NbEDS1*-gRNA triggers the gene editing events.

To clarify our speculation, we performed *Agrobacterium*-mediated transient assays on *N. benthamiana*, and the co-inoculation diagram can be visualized in Figure 1A. *A. tumefaciens* strains carrying p35S::*NbEDS1*-yellow fluorescent protein (YFP) or p35S::*NbNDR1*-YFP with or without pCas9-*NbEDS1*/*NbNDR1* gRNA, were inoculated onto the *N. benthamiana* leaf. We previously also generated transgenic *N. benthamiana* plants expressing pCas9-*NbWRKY70* gRNA, where the *NbWRKY70* gene was successfully knocked out, suggesting that *NbWRKY70* gRNA is functional and efficient (Liu et al., 2016). Therefore, in this study, we also used pCas9-*NbWRKY70* gRNA and the p35S::*NbWRKY70*-YFP as controls.

The fluorescent signals of *NbEDS1*-YFP, *NbNDR1*-YFP, and *NbWRKY70*-YFP fusion proteins were detected under a fluorescent microscope (Figures 1B–J). The fluorescent signals of *NbEDS1*-YFP slightly reduced when it was co-expressed with pCas9-*NbEDS1*/*NbNDR1* gRNA, but not pCas9 or pCas9-*NbWRKY70* gRNA (Figures 1B–D). Similarly, the *NbWRKY70*-YFP signals were also reduced when it was co-expressed with pCas9-*NbWRKY70* gRNA (Figures 1E–G). On the contrary, the *NbNDR1*-YFP fluorescent signals were similar when it was co-expressed with pCas9-*NbEDS1*/*NbNDR1* gRNA, pCas9, or pCas9-*NbWRKY70* gRNA (Figures 1H–J).

This result again suggests that the gRNA targeting *NbNDR1* is ineffective.

Although the reduction in the YFP signal can be detected under a fluorescence microscope, we had never observed that the GFP signals could be suppressed entirely in our transient assays when the GOI-YFP was co-expressed with a cognate Cas9/gRNA. It is not ensured to address a conclusion on gRNA effectiveness based on the reduced YFP signals. To further validate the transient assay result, we performed a western blot analysis to check the accumulation of *NbEDS1*-YFP, *NbNDR1*-YFP, and *NbWRKY70*-YFP fusion proteins. The accumulation of *NbEDS1*-YFP (Figure 2A) and *NbWRKY70*-YFP (Figure 2B), but not *NbNDR1*-YFP (Figure 2C) fusion proteins, are dramatically reduced when they were co-expressed with the pCas9 construct carrying a corresponding gRNA. This result is consistent with the fluorescent signals detected under a fluorescent microscope (Figure 1).

Development of a pgRNA-YFP vector for functional validation of gRNAs

Although this transient assay method can assess the effectiveness of gRNAs used for CRISPR/Cas9-mediated genome editing in plant cells, it requires the cloning of the open reading frame (ORF) of a target gene, and further validation *via* western blotting analysis. We reasoned that, instead of cloning the ORF of a target gene, we can simply insert the gRNA binding sequence (23 nucleotides including the protospacer adjacent motif (PAM) sequence) into the YFP gene, the resulted gRNA-YFP fusion gene will have a frameshift in the YFP gene, which disrupts the YFP protein expression.

To this end, a binary vector, pXhoI-YFP, carrying a YFP gene, where it has a unique in-frame XhoI site adjacent to the start codon of YFP gene, was constructed (Figure 3A). A synthesized gRNA (23 nt) targeting sequence, including a PAM site, can be inserted into the XhoI site causing a frameshift of YFP (Figure 3A). The derived construct is denominated pgRNA-YFP. When it is co-expressed

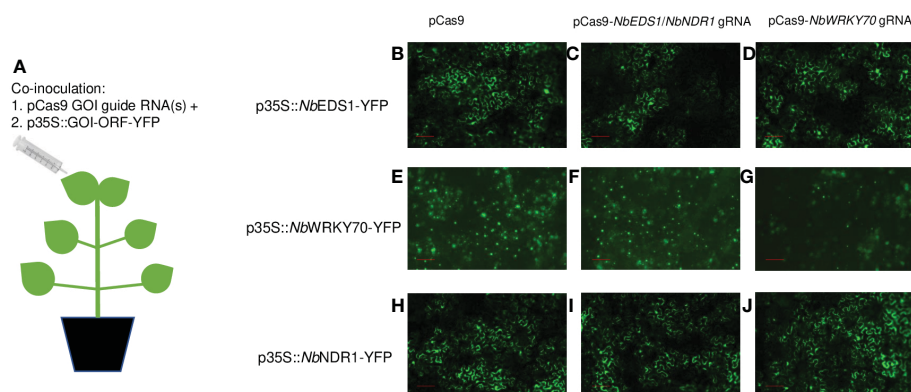


FIGURE 1

Evaluate the effectiveness of gRNAs targeting *NbEDS1*, *NbNDR1*, *NbWRKY70* via *Agrobacterium*-mediated transient assay. (A) The co-inoculation scheme of *A. tumefaciens* strains carrying pCas9-GOI gRNA and p35S::GOI-YFP on *N. benthamiana*. p35S::*NbEDS1*-YFP (B–D), p35S::*NbWRKY70*-YFP (E–G), and p35S::*NbNDR1*-YFP (H–J) were respectively co-inoculated with pCas9 empty vector (no gRNA), pCas9-*NbEDS1*/*NbNDR1* gRNAs, and pCas9-*NbWRKY70* gRNA. The YFP signals were detected by a fluorescent microscope at 48 hour post inoculation. *NbEDS1* gRNA and *NbWRKY70* gRNA, but not *NbNDR1* gRNA, can trigger gene editing when co-expressing with pCas9-GOI gRNA in *Agrobacterium*-mediated transient assays in *N. benthamiana*, which resulted in reduced fluorescent signals. Scale bar represents 100 μ m. GOI; gene of interest, YFP; yellow fluorescent protein, 35S; Cauliflower mosaic virus 35S promoter.

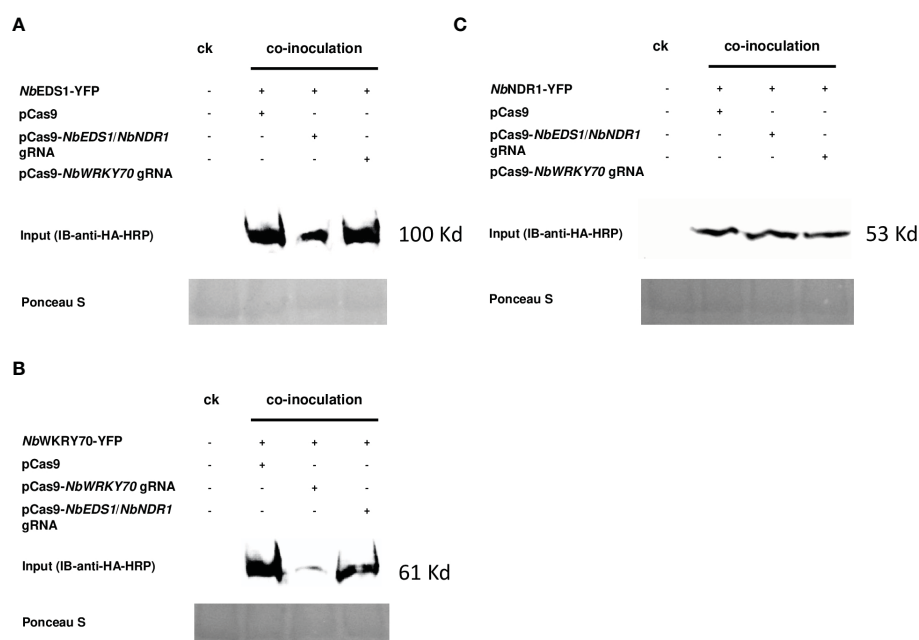


FIGURE 2

Image of the western blot used to detect the YFP fusion proteins. The protein accumulation of (A) *NbEDS1*-YFP, (B) *NbWRKY70*-YFP, and (C) *NbNDR1*-YFP were detected by probing the blot with anti-HA-HRP antibodies. (A) *tumefaciens* strains carrying p35S::*NbEDS1*-YFP, *NbWRKY70*-YFP, and p35S::*NbNDR1*-YFP, were co-inoculated with pCas9 empty vector, pCas9-*NbEDS1*/*NbNDR1* gRNA, and pCas9-*NbWRKY70* gRNA onto the *N. benthamiana* leaf, respectively. CK: the protein was extracted from *N. benthamiana* leaf infiltrated with 10 mM MgCl₂ that served as a negative control. The blot membrane was stained with Ponceau S to confirm equivalent loadings. +: present of the indicated construct, -: absence of the indicated construct.

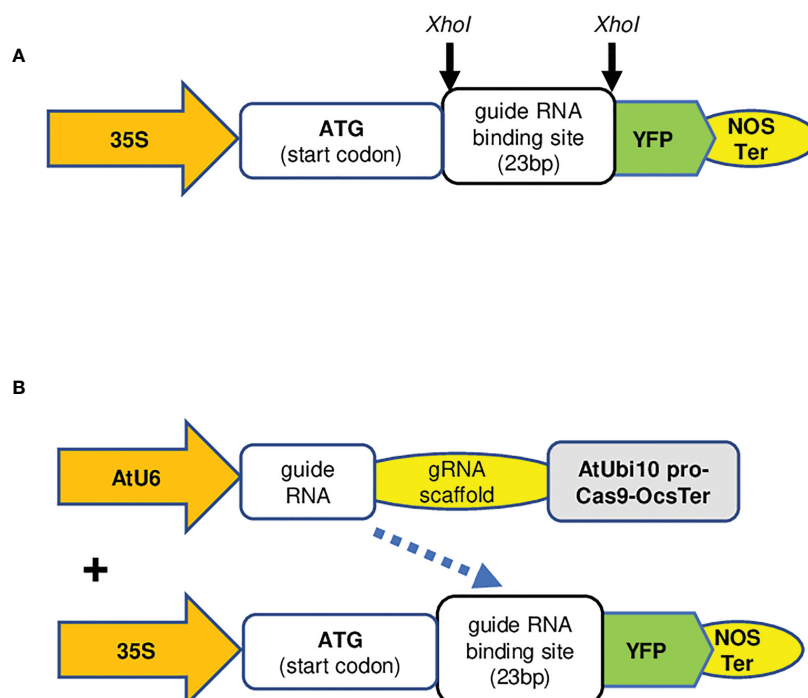


FIGURE 3

The key elements of the plasmid vectors used in this study. (A) The scheme of pgRNA-YFP vector. Two sequence complementary oligonucleotide primers of a given gRNA binding sequence (23 nt), including a PAM site, can be synthesized, annealed, and cloned to the *Xho*I sites of p*Xho*I-ccdB-YFP via Gibson cloning. 35S; Cauliflower mosaic virus 35S promoter, YFP; yellow fluorescent protein, Nos Ter; *Agrobacterium tumefaciens* nopaline synthase terminator. Other key elements are identical to the original pEALEYGATE 101 vector [14]. (B) The scheme of pCas9-GOI-gRNA and pgRNA-YFP, illustrates that pCas9-AtU6pro-gRNA expressing gRNA targets the YFP gene in the pgRNA-YFP construct. AtU6, Arabidopsis U6 promoter; AtUbi10 pro, Arabidopsis ubiquitin 10 promoter; OCS Ter, *Agrobacterium tumefaciens* octopine synthase gene terminator; PAM, protospacer adjacent motif.

with another binary vector, pCas9-GOI-gRNA, the YFP reading frame could be restored by the indels created *via* CRISPR/Cas9-mediated gene editing events and recover the YFP protein expression (Figure 3B). To improve the cloning efficiency, we further modified pXhoI-YFP by insertion of an XhoI DNA fragment carrying the *ccdB* gene (Liu et al., 2016). The bacterial clones carrying pXhoI-*ccdB*-YFP empty vector can be negatively selected because of the toxicity conditioned by the *ccdB* gene. The derived vector is named pXhoI-*ccdB*-YFP for cloning of any gRNA binding sequences. Cas9 is a large protein with more than 1,300 amino acids, and its expression level in transient assays might be a bottleneck that limits the efficiency of gene editing in a transient assay system. To simplify the protocol and increase gene editing efficiency, we generated a stable transgenic *N. benthamiana* line expressing the *Cas9* gene constitutively (Figure S3).

Functional test of the gRNA screening system *via* Agrobacterium-mediated transient assays

We validated the pgRNA-YFP/Cas9 system by re-testing the gRNAs targeting *NbEDS1*, *NbNDR1*, and *NbWRKY70* in *N. benthamiana*. Oligo primers based on the gRNA binding sequence of each gene were synthesized, annealed, and cloned into the XhoI site of pXhoI-*ccdB*-YFP. The resulting constructs were pNbEDS1 gRNA-YFP, pNbNDR1 gRNA-YFP, and pNbWRKY70 gRNA-YFP, respectively. The derived gene constructs were co-expressed with pCas9-*NbEDS1*/*NbNDR1* gRNA or pCas9-*NbWRKY70* gRNA by *Agrobacterium*-mediated transient assays in wild-type *N.*

benthamiana and transgenic *N. benthamiana* harboring the *Cas9* gene. YFP fluorescent signals can be detected when *Cas9* is co-expressed with pNbEDS1 gRNA-YFP and pNbWRKY70 gRNA-YFP, but not pNbNDR1 gRNA-YFP (Figures 4A–C). Therefore, we further confirmed that the gRNA targeting *NbNDR1* is ineffective for triggering gene editing, explaining why no *NbNDR1* gene editing events could be detected in stable transgenic plants. These results also imply the feasibility of the new gRNA evaluation system. In addition, we also performed tests on the pCas9 transgenic plants, which works as efficiently as the transient assay method. As previously discussed, the pCas9 transgenic plant makes the system more facile and straightforward. Therefore, we expect the pCas9 transgenic plant could be a valuable resource shared by the plant research community. For the first system described in this study, we have to compare the YFP signals in two samples, where the reduced fluorescent signals could be observed in plant cells with gene editing events. In the second system, the gain of the YFP signals is to compare one sample with YFP signals and another one without fluorescent signals. Therefore, the second method is much more convenient, and the data is more conclusive.

Evaluate the effectiveness of gRNAs targeting two soybean trypsin inhibitor genes

We further tested the feasibility of predicting the effectiveness of gRNA targeting two soybean genes by using the system described above. To this end, we first validated the gRNA effectiveness and then

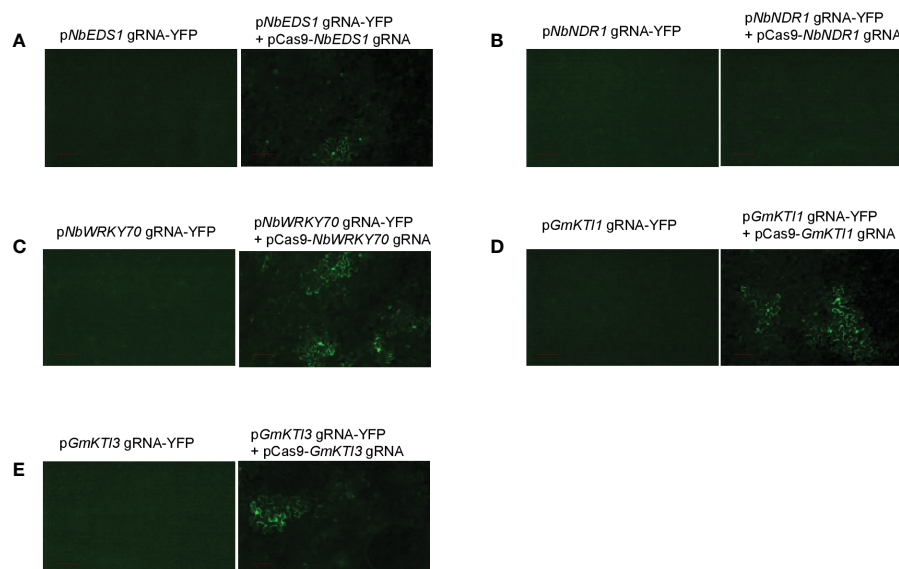


FIGURE 4

The pgRNA-YFP vectors are used to screen the effectiveness of gRNAs in *Agrobacterium*-mediated transient assays. Transient expression of pgRNA-YFP alone in *N. benthamiana* does not produce any fluorescent signals. (A) Co-expression of pNbEDS1 gRNA-YFP with pCas9-*NbNDR1*/*NbEDS1* gRNA restored the YFP signals in *N. benthamiana*. (B) Co-expression of pNbNDR1 gRNA-YFP with pCas9-*NbNDR1*/*NbEDS1* gRNA failed to restore the YFP signals. (C) Co-expression of pNbWRKY70 gRNA-YFP with pCas9-*NbWRKY70* gRNA restored the YFP signals. (D, E) Co-expression of either pGmKTI1 gRNA-YFP or pGmKTI3 gRNA-YFP with pCas9-*GmKTI1*/*GmKTI3* gRNA restored the YFP signals in *N. benthamiana*. Scale bar represents 100 μ m.

generated stable transgenic soybean plants. Although successful gene editing events in soybeans have been reported, the low transformation efficiency and a long-tissue culture process impede a broad application of the CRISPR/Cas9 system in soybean molecular breeding programs. Therefore, it is critical to validate the effectiveness of the gRNAs before the initiation of stable soybean transformation experiments.

Soybean Kunitz trypsin protease inhibitor (*KTI*) genes encode proteins that can reduce protein digestibility in animals and human beings (Gillman et al., 2015). Thus, a KTI-free or low KTI soybean cultivar with mutations on *KTI* genes is highly desirable. We attempted to knock out two seed-specific KTI genes, *KTI1* (Gm01g095000) and *KTI3* (Gm08g341500), via CRISPR/Cas9-mediated genome editing approach. The effectiveness of gRNAs targeting *KTI1* and *KTI3* was validated using the transient assay system described above (Figures 4D, E). Subsequently, the pCas9-*KTI1/3* gRNA construct was transformed into soybean cultivar William 82 via *Agrobacterium*-mediated transformation (Luth et al., 2015). Five independent transgenic plants were recovered and genotyped. Both *KTI1* and *KTI3* gene editing events were detected in all tested transgenic lines (Figures 5A, B).

Conclusion

In this study, we employed the *Agrobacterium*-mediated transient assays to test the effectiveness of gRNAs. The first method was based on the fact that Cas9 could edit a target gene fused with *YFP* and result in reduced fusion proteins and fluorescent signals in transformed plant cells. The second method, the p*XhoI*-*ccdB*-*YFP*/Cas9 system, was developed based on the fact that small indels could be introduced by the nonhomologous end-joining during the process of CRISPR/Cas9-mediated gene editing. Such small indels could correct the reading frame of a *YFP* gene that was disrupted by the insertion of a gRNA binding sequence. We demonstrated that the second method is more convenient and reliable by testing five gRNAs targeting *N. benthamiana* and soybean genes, respectively. The pre-validated gRNAs allow us to generate mutations on target genes in stable transgenic soybean plants (Wang et al., 2022 (under review)). Therefore, we expected the p*XhoI*-*ccdB*-*YFP*/Cas9 system to be broadly applicable for screening gRNA(s) for genome editing in crop plants.

Materials and methods

Escherichia coli, *Agrobacterium tumefaciens* and *Xanthomonas euvesicatoria* strains

E. coli strains *DH5a* and *ccdB survivor* (Thermo Fisher Scientific Inc.) were grown on Luria Broth (LB) medium supplemented with appropriate antibiotics at 37°C. *Agrobacterium tumefaciens* strain GV2260 was grown on LB medium at 28°C. *Xanthomonas euvesicatoria* strain Xe85-10 was grown on nutrient yeast glycerol agar (NYGA) medium at 28°C. Kanamycin 50 ug/ml (Sigma Inc.), Gentamicin 50 ug/ml (Sigma Inc.), and Rifamycin 100 ug/ml (Sigma Inc.) were used to select transformed bacterial cells carrying the designated plasmid constructs.

Plant materials and growth conditions

Soybean (*Glycine max*, cv. William 82 (WM82)) and *Nicotiana benthamiana* seeds were germinated and grown in Sunshine® Mix #1 in a growth chamber (14hr/10hr light/dark cycle at 25°C/20°C). The 4-week-old plants were used for the experiments detailed herein.

CRISPR/Cas9 plasmid construction

A previously described pCRISPR/Cas9-Kan construct (Liu et al., 2016) was modified for the application of genome editing in transgenic soybean plants. In brief, we replaced the *Kanamycin* resistance gene cassette with a *Bar* gene cassette (confers resistance to the herbicide Bialaphos). The *Bar* gene cassette which consisted of a mannopine synthase (MAS) promoter, the *Bar* gene open reading frame, and a MAS terminator, was amplified from the plasmid DNA of pEarleyGate101 (Earley et al., 2006). The PCR primers with annotations are listed in Table S1. The *Bar* gene cassette was cloned at the *PmeI*/*MauBI* sites of pCRISPR/Cas9-Kan construct using a Gibson Assembly® Cloning Kit (New England Biolabs Inc). The derived construct was named as pCas9. The Cas9 gene is expressed by using an Arabidopsis ubiquitin 10 gene promoter (Liu et al., 2016). The guiding RNA expression cassettes targeting *GmKTI1* and *GmKTI3*, *NbEDS1* and *NbNDR1*, and *NbWRKY70*, respectively, were synthesized at GenScript Biotech Corp. The expression of

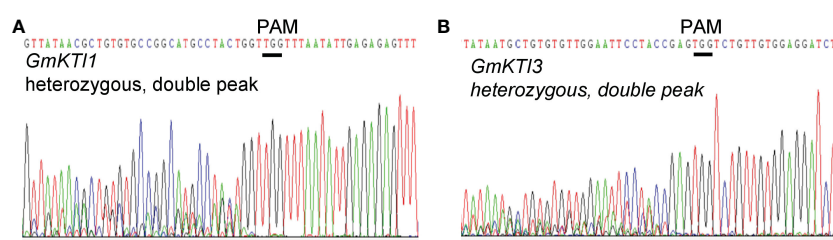


FIGURE 5

Sanger sequencing chromatograms display the DNA sequence of *GmKTI1* and *GmKTI3* that is targeted for gene editing. The open reading frames of *GmKTI1* and *GmKTI3* in the pCas9-*GmKTI1*/*GmKTI3* gRNA T0 transgenic soybean plants were amplified and sequenced. The double peaks in the sequencing chromatograms indicate the presence of both wild-type and mutant allele of sequences *GmKTI1* (A) and *GmKTI3* (B). The wild-type gene sequences are shown on the top of the sequencing chromatograms. The PAM sites are underlined.

gRNA is driven by an Arabidopsis U6 promoter (Liu et al., 2016). The gRNA cassette was cloned into the *PmeI* site of pCas9 using a Gibson Assembly® Cloning Kit. The derived constructs were transformed into *A. tumefaciens* strain GV2260 for stable transformation and transient gene expression assays.

Gene cloning and plasmid construction

The open reading frames (ORFs) of *NbEDS1* (Ordon et al., 2017), *NbNDR1* (Dong et al., 2022), and *NbWRKY70* (Liu et al., 2016) were amplified from the cDNA of *N. benthamiana* with primers listed in Table S1. The ORFs were cloned into pDonr207® using a BP® cloning kit (Thermo Fisher Scientific). The genes/fragments in donor vectors were subcloned into pEarleyGate101 using an LR® Gateway cloning kit (Thermo Fisher Scientific). The derived constructs are p35S:: *NbEDS1*-YFP, p35S:: *NbNDR1*-YFP, and p35S:: *NbWRKY70*-YFP. The expression of cloned genes is driven by a cauliflower mosaic virus 35S promoter (35S). The plant expression constructs were transformed into *A. tumefaciens* strain GV2260 via electroporation.

A binary vector p*XhoI*-YFP carrying a yellow fluorescent protein (YFP) gene, was constructed by amplifying and cloning a DNA fragment into the *XhoI* and *SpeI* sites of pEarleyGate101 (Earley et al., 2006). The primers are listed in Table S1. The cloned YFP gene has a unique in-frame *XhoI* site adjacent to the start codon of YFP gene. The *ccdB* cassette was amplified from pEarleyGate101 and inserted into the *XhoI* site of p*XhoI*-YFP (Table S1). The derived vector named p*XhoI*-ccdB-YFP maintained in *E. coli* strain *ccdB* survivor (Thermo Fisher Scientific Inc.), and it can be used for cloning of any synthesized gRNA binding sequences with reduced false-positive clones derived from incomplete digestion of p*XhoI*-ccdB-YFP DNA, which can be negatively selected by the *ccdB* gene in *E. coli* DH5α.

Gibson cloning

To clone a gRNA binding sequence into p*XhoI*-ccdB-YFP, two sequence complementary oligonucleotide primers of a given gRNA binding sequence (23 nt), including a PAM site, can be synthesized, annealed and cloned to the *XhoI* sites of p*XhoI*-ccdB-YFP via Gibson cloning. In brief, two nucleotide primers carrying the gRNA binding sequence, including a PAM site, were synthesized (Integrated DNA Technologies, Inc.). The two primers (100 pm) were annealed by using a thermal cycler programmed as 98 °C/30 sec, 56 °C/45 sec, repeated 10 cycles. The annealed gRNA linker was directly cloned into the *XhoI* site of p*XhoI*-ccdB-YFP by using a Gibson cloning kit (New England Inc) (Lei et al., 2014). In brief, 0.01 pmol of linearized vector p*XhoI*-ccdB-YFP, and 0.2 pmol of gRNA linker, were mixed with an equal volume of 2x Gibson master mix, and incubated in a thermal cycler programmed to hold at 50°C for 60 min (Jacobs and Martin, 2016). After incubation, the reaction mixture was transformed into *E. coli* DH5a using a heat shock approach. The derived binary constructs p*XhoI*-YFP was transformed into *Agrobacterium tumefaciens* strain GV2260 via electroporation.

Agrobacterium-mediated transient assay in *N. benthamiana* and western blotting analysis

Agrobacterium strains GV2260 harboring pCas9-gRNA plasmid, and p*XhoI*-YFP plasmid were co-infiltrated into *N. benthamiana* leaf mesophyll tissues for transient protein expression. The YFP signals were monitored using a fluorescent microscope (Axio Observer A.1, Zeiss) at 48 hr post infiltration. The YFP gene was fused with a C-terminal HA epitope tag, and the fusion proteins can be detected with anti-HA-HRP (Roche) antibodies at a dilution of 1:5,000. Plant-expressed proteins were detected by Western blot analysis following the procedure as previously described (Wang et al., 2018; Liu et al., 2020b).

Xanthomonas euvesicatoria disease assay on *N. Benthamiana*

Xe strain *Xe85-10* was cultivated on NYGA mediums supplemented with rifamycin (100 ug/ml) at 28 °C for two days. The bacterial cells were collected and suspended in 10 mM MgCl₂ and diluted to 4 × 10⁸ CFU ml⁻¹. The bacterial inoculum was infiltrated into the backside of *N. benthamiana* leaf using a blunt-end needleless syringe. The inoculated plants were maintained under 14 hr light/10 hr dark at room temperature. The phenotypes of programmed cell death were recorded at 48 hr post-inoculation.

Generation of transgenic tobacco and soybean plant by *Agrobacterium*

Agrobacterium strain GV2260 harboring plasmid pCas9, or pCas9-gRNA was used for tobacco transformation following a protocol as previously described (Liu et al., 2016). Soybean transformation was performed at the plant transformation facility at Iowa State University as previously described (Paz et al., 2006; Luth et al., 2015; Ge et al., 2016).

Data availability statement

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

Author contributions

BZhao conceived the project; ZW and BZhao designed the research procedure. ZW, ZS, KW, QL, KM, performed the experiments. ZW and BZhao analyzed the data. ZW, BZhao, and BZhao wrote the manuscript with input from all authors. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Virginia Soybean Board (#467059) and a seed grant from the Translational Plant Science Center at Virginia Tech. An integrated internal competitive grant from the Center for Advanced Innovation in Agriculture in College of Agriculture and Life Sciences at Virginia Tech, and Virginia Agricultural Experiment Station (VA160144).

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.

References

- Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., and Parker, J. E. (1998). Different requirements for *EDS1* and *NDR1* by disease resistance genes define at least two *r* gene-mediated signaling pathways in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 95 (17), 10306–10311. doi: 10.1073/pnas.95.17.10306
- Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573. doi: 10.1126/science.aaf5573
- Carlsen, F. M., Johansen, I. E., Yang, Z., Liu, Y., Westberg, I. N., Kieu, N. P., et al. (2022). Strategies for efficient gene editing in protoplasts of *Solanum tuberosum* theme: Determining gRNA efficiency design by utilizing protoplast (Research). *Front. Genome Ed.* 3, 795644. doi: 10.3389/fgeed.2021.795644
- Century, K. S., Holub, E. B., and Staskawicz, B. J. (1995). *NDR1*, a locus of *Arabidopsis thaliana* that is required for disease resistance to both a bacterial and a fungal pathogen. *Proc. Natl. Acad. Sci. U. S. A.* 92 (14), 6597–6601. doi: 10.1073/pnas.92.14.6597
- Chen, P. J., Hussmann, J. A., Yan, J., Knipping, F., Ravisankar, P., Chen, P. F., et al. (2021a). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* 184, 5635–5652.e29. doi: 10.1016/j.cell.2021.09.018
- Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049
- Chen, Z., Wu, Q., Tong, C., Chen, H., Miao, D., Qian, X., et al. (2021b). Characterization of the roles of SGT1/RAR1, EDS1/NDR1, NPR1, and NRC/ADR1/NRG1 in sw-5b-Mediated resistance to tomato spotted wilt virus. *Viruses* 13 (8), 1447. doi: 10.3390/v13081447
- Coll, N. S., Eppe, P., and Dangl, J. L. (2011). Programmed cell death in the plant immune system. *Cell Death Differ.* 18, 1247–1256. doi: 10.1038/cdd.2011.37
- Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., et al. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* 16, 280. doi: 10.1186/s13059-015-0846-3
- Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267. doi: 10.1038/nbt.3026
- Dong, X., Ai, G., Xia, C., Pan, W., Yin, Z., and Dou, D. (2022). Different requirement of immunity pathway components by oomycete effectors-induced cell death. *Phytopathol. Res.* 4, 4. doi: 10.1186/s42483-022-00109-1
- Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K., et al. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* 45, 616–629. doi: 10.1111/j.1365-3113X.2005.02617.x
- Ge, L., Yu, J., Wang, H., Luth, D., Bai, G., Wang, K., et al. (2016). Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proc. Natl. Acad. Sci. U. S. A.* 113 (44), 12414–12419. doi: 10.1073/pnas.1611763113
- Gillman, J. D., Kim, W. S., and Krishnan, H. B. (2015). Identification of a new soybean kunitz trypsin inhibitor mutation and its effect on bowman-birk protease inhibitor content in soybean seed. *J. Agric. Food Chem.* 63, 1352–1359. doi: 10.1021/jf505220p
- Graf, R., Li, X., Chu, V. T., and Rajewsky, K. (2019). sgRNA sequence motifs blocking efficient CRISPR/Cas9-mediated gene editing. *Cell Rep.* 26, 1098–1103.e3. doi: 10.1016/j.celrep.2019.01.024
- Jacobs, T. B., and Martin, G. B. (2016). High-throughput CRISPR vector construction and characterization of DNA modifications by generation of tomato hairy roots. *J. Vis. Exp.* 110, 53843. doi: 10.3791/53843
- Kim, H., Kim, S. T., Ryu, J., Kang, B. C., Kim, J. S., and Kim, S. G. (2017). CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* 8, 14406. doi: 10.1038/ncomms14406
- Knepper, C., Savory, E. A., and Day, B. (2011). The role of NDR1 in pathogen perception and plant defense signaling. *Plant Signaling Behav.* 6, 1114–1116. doi: 10.4161/psb.6.8.15843
- Konstantakos, V., Nentidis, A., Krithara, A., and Paliouras, G. (2022). CRISPR-Cas9 gRNA efficiency prediction: an overview of predictive tools and the role of deep learning. *Nucleic Acids Res.* 50, 3616–3637. doi: 10.1093/nar/gkac192
- Kumar, A., Birnbaum, M. D., Moorthy, B. T., Singh, J., Palovcak, A., Patel, D. M., et al. (2019). Insertion/deletion-activated frame-shift fluorescence protein is a sensitive reporter for genomic DNA editing. *BMC Genomics* 20, 609. doi: 10.1186/s12864-019-5963-z
- Lapin, D., Bhandari, D. D., and Parker, J. E. (2020). Origins and immunity networking functions of EDS1 family proteins. *Annu. Rev. Phytopathol.* 58, 253–276. doi: 10.1146/annurev-phyto-010820-012840
- Lei, Y., Lu, L., Liu, H.-Y., Li, S., Xing, F., and Chen, L.-L. (2014). CRISPR-p: A web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol. Plant* 7, 1494–1496. doi: 10.1093/mp/ssu044
- Liang, Y., Eudes, A., Yogiswara, S., Jing, B., Benites, V. T., Yamanaka, R., et al. (2019). A screening method to identify efficient sgRNAs in arabidopsis, used in conjunction with cell-specific lignin reduction. *Biotechnol. Biofuels* 12, 130. doi: 10.1186/s13068-019-1467-y

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2023.1111683/full#supplementary-material>

SUPPLEMENTARY FIGURE 2

Knock out of *NbEDS1* impairs the cell death trigger by interaction of *NbRoq1* and *XopQ*, but not *Rps2/AvrRpt2*. (A) The CRISPR/Cas9 construct harbors three necessary elements exhibited as below: the plant selection cassette consists of a MAS (mannopine synthase) promoter, Bar gene, and a MAS terminator; the Cas9 cassette consists of *Arabidopsis* ubiquitin 10 promoter, Cas9 gene, and OCS (octopine synthase) terminator; two guide RNA cassettes and each of them consists of a *Arabidopsis* U6 promoter, and one sgRNA. Sequence alignment of the wild type and mutant *EDS1* (B) or *NDR1* (C) alleles generated via CRISPR/Cas9-mediated genome editing displayed that the *NbEDS1* gene was successfully edited in all genotyped plants, while no edited *NbNDR1* genes were detected in any genotyped plants. (D) The pCas9- *NbNDR1/NbEDS1* gRNA transgenic plants and wild-type plants were inoculated with *Xanthomonas euvesicatoria* strain Xe85-10. By contrast, the co-inoculation of *A. tumefaciens* strains carrying p35S::RPS2 and p35S::avrRpt2, both with the concentration of 2×10^8 CFU ml⁻¹, can trigger cell death on both wild-type and transgenic *N. benthamiana* at 48 hpi (E). The pictures were taken from the front side and back side of the leaf sample.

SUPPLEMENTARY FIGURE 3

Validation of the pCas9 transgenic *N. benthamiana* plants. The pCas9 vector that doesn't carry any gRNAs was used to transform *N. benthamiana* plant. The Cas9 gene fragments were detected in transgenic lines #1 and #2 by PCR amplification but not in wild-type plants. The plasmid DNA pCas9 is served as the template for positive control. (1) 1 Kb marker, (2) pCas9 line 1, (3) pCas9 line 2, (4) wild-type plant, (5) pCas9 plasmid DNA.

- Liang, G., Zhang, H., Lou, D., and Yu, D. (2016). Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci. Rep.* 6, 21451. doi: 10.1038/srep21451
- Liu, L., and Fan, X. D. (2014). CRISPR-cas system: a powerful tool for genome engineering. *Plant Mol. Biol.* 85, 209–218. doi: 10.1007/s11103-014-0188-7
- Liu, Y., Miao, J., Traore, S., Kong, D., Liu, Y., Zhang, X., et al. (2016). SacB-SacR gene cassette as the negative selection marker to suppress agrobacterium overgrowth in agrobacterium-mediated plant transformation. *Front. Mol. Biosci.* 3, 70. doi: 10.3389/fmolb.2016.00070
- Liu, Y., Wang, K., Cheng, Q., Kong, D., Zhang, X., Wang, Z., et al. (2020b). Cysteine protease RD21A regulated by E3 ligase SINAT4 is required for drought-induced resistance to *Pseudomonas syringae* in *Arabidopsis*. *J. Exp. Bot.* 71, 5562–5576. doi: 10.1093/jxb/eraa255
- Liu, G., Zhang, Y., and Zhang, T. (2020a). Computational approaches for effective CRISPR guide RNA design and evaluation. *Comput. Struct. Biotechnol. J.* 18, 35–44. doi: 10.1016/j.csbj.2019.11.006
- Luth, D., Warnberg, K., and Wang, K. (2015). Soybean [*Glycine max* (L.) Merr]. *Methods Mol. Biol.* 1223, 275–284. doi: 10.1007/978-1-4939-1695-5_22
- Martin, R., Qi, T., Zhang, H., Liu, F., King, M., Toth, C., et al. (2020). Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* 370 (6521), eabd9993. doi: 10.1126/science.abd9993
- Matson, A. W., Hosny, N., Swanson, Z. A., Hering, B. J., and Burlak, C. (2019). Optimizing sgRNA length to improve target specificity and efficiency for the GGTA1 gene using the CRISPR/Cas9 gene editing system. *PloS One* 14, e0226107. doi: 10.1371/journal.pone.0226107
- Ordon, J., Gantner, J., Kemna, J., Schwalgun, L., Reschke, M., Streubel, J., et al. (2017). Generation of chromosomal deletions in dicotyledonous plants employing a user-friendly genome editing toolkit. *Plant J.* 89, 155–168. doi: 10.1111/tpj.13319
- Paz, M. M., Martinez, J. C., Kalvig, A. B., Fonger, T. M., and Wang, K. (2006). Improved cotyledonary node method using an alternative explant derived from mature seed for efficient agrobacterium-mediated soybean transformation. *Plant Cell Rep.* 25, 206–213. doi: 10.1007/s00299-005-0048-7
- Schultink, A., Qi, T., Lee, A., Steinbrenner, A. D., and Staskawicz, B. (2017). Roq1 mediates recognition of the *Xanthomonas* and *Pseudomonas* effector proteins XopQ and HopQ1. *Plant J.* 92, 787–795. doi: 10.1111/tpj.13715
- Wang, Z., Li, G., Sun, H., Ma, L., Guo, Y., Zhao, Z., et al. (2018). Effects of drought stress on photosynthesis and photosynthetic electron transport chain in young apple tree leaves. *Biol. Open* 7 (11), bio035279. doi: 10.1242/bio.035279
- Wang, Z., Shea, Z., Rosso, L., Shang, C., Li, J., Bewick, P., et al. (2022). Development of molecular markers of the Kunitz trypsin inhibitor mutant alleles generated by CRISPR/Cas9-mediated mutagenesis in soybean. *bioRxiv* 2022.08.22.504807. doi: 10.1101/2022.08.22.504807
- Wu, X., Kriz, A. J., and Sharp, P. A. (2014). Target specificity of the CRISPR-Cas9 system. *Quantitat Biol.* 2, 59–70. doi: 10.1007/s40484-014-0030-x
- Xiang, X., Corsi, G. I., Anthon, C., Qu, K., Pan, X., Liang, X., et al. (2021). Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning. *Nat. Commun.* 12, 3238. doi: 10.1038/s41467-021-23576-0
- Zong, Y., Liu, Y., Xue, C., Li, B., Li, X., Wang, Y., et al. (2022). An engineered prime editor with enhanced editing efficiency in plants. *Nat. Biotechnol.* 40 (9), 1394–1402. doi: 10.1038/s41587-022-01254-w



OPEN ACCESS

EDITED BY

Xue-Feng Ma,
Forage Genetics International,
United States

REVIEWED BY

Brian Ward,
Forage Genetics International,
United States
Alagu Manickavelu,
Central University of Kerala, India

*CORRESPONDENCE

Dongcheng Liu
✉ liudongcheng@hebau.edu.cn
Diaoguo An
✉ dgan@sjziam.ac.cn

SPECIALTY SECTION

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

RECEIVED 06 January 2023

ACCEPTED 08 March 2023

PUBLISHED 17 March 2023

CITATION

Ma F, Xu Y, Wang R, Tong Y, Zhang A, Liu D
and An D (2023) Identification of major
QTLs for yield-related traits with improved
genetic map in wheat.
Front. Plant Sci. 14:1138696.
doi: 10.3389/fpls.2023.1138696

COPYRIGHT

© 2023 Ma, Xu, Wang, Tong, Zhang, Liu and
An. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Identification of major QTLs for yield-related traits with improved genetic map in wheat

Feifei Ma¹, Yunfeng Xu¹, Ruifang Wang¹, Yiping Tong³,
Aimin Zhang^{2,3}, Dongcheng Liu^{2*} and Diaoguo An^{1*}

¹Center for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang, China, ²State Key Laboratory of North China Crop Improvement and Regulation, College of Agronomy, Hebei Agricultural University, Baoding, Hebei, China, ³State Key Laboratory of Plant Cell and Chromosome Engineering, National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Introduction: Identification of stable major quantitative trait loci (QTLs) for yield-related traits is important for yield potential improvement in wheat breeding.

Methods: In the present study, we genotyped a recombinant inbred line (RIL) population using the Wheat 660K SNP array and constructed a high-density genetic map. The genetic map showed high collinearity with the wheat genome assembly. Fourteen yield-related traits were evaluated in six environments for QTL analysis.

Results and Discussion: A total of 12 environmentally stable QTLs were identified in at least three environments, explaining up to 34.7% of the phenotypic variation. Of these, *QTKW-1B.2* for thousand kernel weight (TKW), *QPh-2D.1* (*QSL-2D.2/QScn-2D.1*) for plant height (PH), spike length (SL) and spikelet compactness (SCN), *QPh-4B.1* for PH, and *QTss-7A.3* for total spikelet number per spike (TSS) were detected in at least five environments. A set of Kompetitive Allele Specific PCR (KASP) markers were converted based on the above QTLs and used to genotype a diversity panel comprising of 190 wheat accessions across four growing seasons. *QPh-2D.1* (*QSL-2D.2/QScn-2D.1*), *QPh-4B.1* and *QTss-7A.3* were successfully validated. Compared with previous studies, *QTKW-1B.2* and *QPh-4B.1* should be novel QTLs. These results provided a solid foundation for further positional cloning and marker-assisted selection of the targeted QTLs in wheat breeding programs.

KEYWORDS

wheat, high-density linkage map, QTL mapping, yield-related traits, the wheat 660K SNP array

Introduction

Common wheat (*Triticum aestivum* L. $2n = 6x = 42$, AABBDD) is one of the important grain crops in the world. Given the food security challenges caused by gradual decrease in arable land and rapid increase in global population, increasing the yield potential of wheat through high-yield breeding programs has been a major focus of most wheat breeders around the world. Considering the complex inheritance and significant influence of environment, yield improvement remains to be a huge challenge. Achievement of this goal will require full identification of promising yield-related loci in wheat.

Wheat yield comprises three main components, viz. spike number per plant (SNPP), kernel number per spike (KNS) and thousand kernel weight (TKW). Of these, SNPP and KNS are more easily influenced by environment (Li et al., 2018). Lots of quantitative trait loci (QTLs) for these two traits have been detected. Two genes *TaTEF-7A* and *GNI1* related to KNS and one gene *TaD27-7B* related to SNPP were cloned (Zheng et al., 2014; Sakuma et al., 2019; Zhao et al., 2020). For TKW, due to its higher heritability, numerous QTLs have been mapped on all 21 wheat chromosomes, and more than 20 genes related to kernel weight have been cloned (Jiang et al., 2011; Zhang et al., 2012; Guo et al., 2013; Chang et al., 2014; Dong et al., 2014; Zhang et al., 2014; Hanif et al., 2015; Jiang et al., 2015; Ma et al., 2016; Zhang et al., 2017b; Zhang et al., 2018b; Yan et al., 2019; Yang et al., 2019). KNS and TKW can be affected by spike length (SL) and kernel weight per spike (KWS). The gene *Q* played an important role in domestication of polyploid wheat because it pleiotropically influenced many important domestication-related traits, including rachis fragility, threshability, glume tenacity, spike architecture, plant height and flowering time (Simons et al., 2006). The gene *Q* was reported to control SL (Jiang et al., 2019), whereas genetic studies on KWS were not given enough attention previously. KNS can be further divided into kernel number per spikelet and total spikelet number per spike (TSS), which was composed of fertile spikelet number per spike (FSS) and sterile spikelet number per spike (SSS). A wheat ortholog of rice gene *APO1* was reported to be the best candidate gene for a locus on 7AL affecting TSS (Kuzay et al., 2019; Kuzay et al., 2022). Plant height (PH) was significantly related to SL and affected the harvest index (HI) and grain yield. More than 20 major genes of PH have been identified and designated as reduced height (*Rht*) genes (Peng et al., 2011; Chen et al., 2014; Lu et al., 2015; Zhang et al., 2017a; Ford et al., 2018; Chai et al., 2022; Xiong et al., 2022). PH and SNPP can affect biomass yield per plant (BYP), which is composed of straw yield per plant (SYP) and kernel yield per plant (KYP). HI is the ratio of KYP to BYP and reflects the allocation of photosynthetic products between grain and vegetative organs. Improving HI is one of the important goals in wheat breeding programs.

To date, many QTLs related to yield traits have been identified on all 21 chromosomes of wheat (Liu et al., 2020). However, due to the large genome size, hexaploid nature and high percentage of repetitive regions of wheat, most QTLs were mapped by a low-density genetic linkage map with large confidence interval, and only

several yield-related QTLs have been fine mapped and cloned. A high-density genetic map based on an individual biparental mapping population would be of great value for high-resolution mapping and map-based cloning of a major targeted QTL. With the development of new sequencing technologies, high-density single nucleotide polymorphism (SNP) arrays technology has become a superior approach to construct genetic map and identify QTLs for yield-related traits in wheat. The high-density SNP assays including 9K (Wu et al., 2015), 35K (Allen et al., 2017), 55K (Ren et al., 2018), 90K (Wang et al., 2014), 660K (Cui et al., 2017) and 820K (Winfield et al., 2016), have become the best alternative to identify QTLs in wheat. For example, Cui et al. (2017) reported a high-density genetic map based on the 660K SNP array that was in good accordance with the released wheat genome assembly, providing a major resource for future genetic and genomic research. Using the high-density genetic map, a major QTL for KNS and eight putative additive QTL for PH were characterized (Cui et al., 2017; Zhang et al., 2017a). In addition, a high-density SNP genetic map is helpful to identify QTLs with major and stable effects. Converting the tightly linked SNP markers of QTLs into kompetitive allele specific PCR (KASP) markers that can be used for further validation in different genetic backgrounds, is important for marker-assisted transfer of these QTLs into elite breeding lines successfully. For instance, Liu et al. (2020) identified a QTL for kernel-related traits using a high-density genetic map based on the 660K SNP array and a KASP marker was developed for the QTL and verified by a natural population consisted of 141 cultivar/lines.

In our previous study, Xu et al. (2014) detected QTLs in a recombinant inbred line (RIL) population derived from 'Xiaoyan 54' and 'Jing 411' using a genetic map with 555 PCR-based markers and most QTLs were mapped in large confidence interval. Here, we used the Wheat 660K SNP array to re-genotype the 'Xiaoyan 54 × Jing 411' RIL population and identify QTLs for 14 yield-related traits across six environments. Our objectives were to: (i) construct a high-density genetic linkage map; (ii) identify key QTLs that were significantly associated with yield-related traits in at least five environments; (iii) develop KASP marker based on the key QTLs and validate the loci in a diversity panel; (iv) predict candidate genes for the key QTLs. The results may contribute key QTLs and user-friendly markers for marker-assisted selection, which can facilitate yield improvement in wheat breeding and provide further insights into the genetic basis of yield-related traits in wheat.

Materials and methods

Plant materials

One hundred and eighty-two F_{11} RILs derived from a cross between wheat cultivars 'Xiaoyan 54' and 'Jing 411' were used for QTL mapping. 'Xiaoyan 54' was derived from the wheat founder parent 'Xiaoyan 6', which was released in 1980 and has been widely cultivated in the main wheat growing regions of China in the past 30 years (Li et al., 2008). 'Jing 411' was a widely grown cultivar in the Northern Winter Wheat Region of China in the 1990s (Zhuang, 2003).

A diversity panel composed of 190 wheat genotypes, including 42 wheat founder parents and widely grown cultivars at different decades since 1950, 32 elite cultivars widely grown in Huang-huai wheat region in recent years, 68 Xiaoyan 6-derivatives or related cultivars, and 48 accessions from Chinese wheat mini-core collections, was used to validate key loci detected in this study.

Field trials and phenotyping

The trials were conducted at Luancheng Agroecosystem Experimental Station, the Chinese Academy of Sciences (37°53'15"N, 114°40'47"E, and elevation 50 m, located at the piedmont of the Taihang Mountains in the North China Plain).

The Xiaoyan 54/Jing 411 RIL population was planted during the 2006–2007 and 2007–2008 growing seasons. Three treatments were applied: low N (LN), low P (LP) and normal fertilized control (CK). Hereafter, '6LN', '6LP', '6CK', '7LN', '7LP' and '7CK' represent the six year \times treatment trials, respectively. A randomized complete-block design was employed, with three separate adjacent blocks as the main plots for the three treatments and subplots for the 182 RILs and their parents (Xu et al., 2014). For each Xiaoyan 54/Jing 411 RIL, ten plants in the middle of the two internal rows in each plot were sampled for phenotyping. PH, BYP, KYP and SNPP were determined from the mean of the ten plants; SL, KNS, SSS, FSS, TSS and KWS were determined from the mean of the main spikes of the ten plants. TKW was evaluated after harvest by weighing 500 kernels in triplicate. HI was calculated as KYP/BYP, spikelet compactness (SCN) as TSS/SL and SYP as BYP-KYP.

The diversity panel was grown in four wheat growing seasons from 2012–2013 to 2015–2016. A split-plot design was employed, with three separate adjacent blocks as the main plots for the three replications, and subplots for the genotypes. Each accession was planted in four 150 cm-long rows, 25 cm apart, with 30 seeds per row. Seeds were hand planted at the beginning of October, and plants were harvested in the middle of next June at physiological maturity. For each accession of the diversity panel, nine yield-related traits including PH, SNPP, SL, KNS, TSS, FSS, SSS, TKW and KYP were evaluated.

Map construction and QTL mapping

The 182 RILs as well as their parental cultivars were genotyped using the Wheat 660K SNP array (Sun et al., 2020). SNP flanking sequences were used to blast against the IWGSC wheat genome sequence (IWGSC RefSeq v1.1) to determine the physical locations of SNPs. SNP markers that had Call Rate > 97%, Heterozygote Rate < 20%, and Minor Allele Frequency > 5% were selected for map construction. To reduce the complexity of calculation, redundant markers (co-segregating markers) in the Xiaoyan 54/Jing 411 RIL population were removed using the BIN function in IciMapping 4.1 (<http://www.isbreeding.net/>) according to Meng et al. (2015). The unique SNP markers were sorted into linkage groups using the MAP function in IciMapping 4.1. The Kosambi mapping function was used to calculate genetic distances in centiMorgans (cM) with a LOD score

of 3.5 and a recombination fraction of 0.3. Markers with no linkage or linkage groups with less than five markers were discarded in the subsequent analysis. The 21 chromosomes and the marker order were confirmed according to physical position of most SNPs in the Chinese Spring reference genome sequence of wheat (RefSeq v1.1) (Appels et al., 2018). MapChart 2.2 (<http://www.biometris.nl/uk/Software/MapChart/>) was used to draw the genetic map. QTL mapping was conducted using the MAP function in IciMapping 4.1 with the inclusive composite interval mapping of additive (ICIM-ADD) QTL method, a walk speed of 1.0 cM, and a stepwise regression probability of 0.001. The LOD threshold 3.0 was set to declare a significant QTL.

Conversion of SNPs of key QTLs to KASP markers

Based on the flanking marker sequence of key QTL for yield-related traits that can be detected in at least five environments, eight SNPs were converted to KASP primers, which are specific for SNP genotyping (LGC Genomics LLC, Beverly, MA, USA). Newly designed KASP markers were evaluated for polymorphisms in reactions containing 5.0 μ l water, 5.0 μ l 2 \times KASPar reaction mix, 0.14 μ l assay mix, and 50 ng dried DNA, with a PCR profile of 94°C for 15 min (activation), followed by 10 cycles of 94°C for 20 s, 61–55°C for 60 s (drop 0.6°C per cycle), then 26 cycles of 94°C for 20 s and 55°C for 60 s. Fluorescence was measured as an end point reading at 37°C. KASP was performed in a BIO-RAD CFX Real-Time PCR system, and fluorescence was detected using Bio-Rad CFX Manage 3.1 software.

Results

Phenotypic variation and correlation analysis

Phenotypic performance of Xiaoyan 54/Jing 411 RIL population for the 14 yield-related traits is showed in Figure 1 and Supplementary Table S1. Jing 411 had higher PH, SL, TKW, KWS, BYP, KYP and SYP across all the environments (Figure 1; Supplementary Table S1). Conversely, Xiaoyan 54 had higher spikelet compactness (SCN) and SNPP in four and three environments, respectively. In the RIL population, phenotypic values showed continuous variation and transgressive segregation (Figure 1), indicating polygenic inheritance. Estimated correlation coefficients among the 14 traits are showed in Figure 2. For the three yield traits, TKW had a significant and negative correlation with KNS, HI, FSS, SCN and TSS, and a significant positive correlation with SYP, PH and BYP. KNS was positively correlated with HI, KWS, FSS, KYP and TSS, and was negatively correlated with SSS, PH and SYP. SNPP had a positive correlation with BYP, KYP and SYP, and had a negative correlation with KWS. For the spike-related traits, SL had a significant and negative correlation with SCN, and a significant and positive correlation with PH, SSS, TSS, SYP, BYP, FSS and KYP. SCN was positively correlated with HI, and was negatively correlated with PH, SYP, SSS, BYP, TKW and KYP. TSS had a significant and positive correlation with FSS and SSS. KWS was positively correlated

with KYP, HI, BYP, FSS and SYP, and was negatively correlated with SSS. Significant correlations were observed among BYP, KYP, SYP and SNPP. BYP had the highest positive correlation with SYP ($r = 0.894$), followed by BYP versus KYP ($r = 0.858$), BYP versus SNPP ($r = 0.585$), KYP versus SNPP ($r = 0.560$), KYP versus SYP ($r = 0.550$), and SYP versus SNPP ($r = 0.486$). HI had a positive correlation with KYP, and a negative correlation with SYP.

Linkage map construction

We constructed a high-density linkage map with 7,542 unique loci spanning 6153.8 cM (Figure 3, Table 1). Of these loci, 6,987 were SNP markers derived from the Wheat 660K SNP array, and the remaining 555 markers were reported by Xu et al. (2014). The 7,542 markers distributed unevenly on the 21 chromosomes, and the number ranged from 132 for chromosome 4D to 565 for chromosome 3B. The genetic coverage of each chromosome varied from 172.71 cM (4D) to 417.71 cM (2A). Altogether, the markers mapping on the A genome (3,142) were more than those on the B genome (2,878), and much fewer markers (1,522) were mapped on the D genome. Seven gaps (>30 cM) were observed on chromosomes 2B, 2D and 7A (Figure 3; Table 1). Of these, the largest gap was found on 2B, which was 35.7 cM. The marker density of the individual chromosomes ranged from 0.55 cM/marker for 3B to 1.48 cM/marker for 6D with an average marker density of 0.82 cM/marker in the whole genetic map (Table 1). Markers mapped on the A and B genomes had a marker density of 0.77 and 0.65 cM/marker, while those mapped on the D genome had a density of 1.23 cM/marker. Based on the SNP flanking sequences, 6,987 markers were

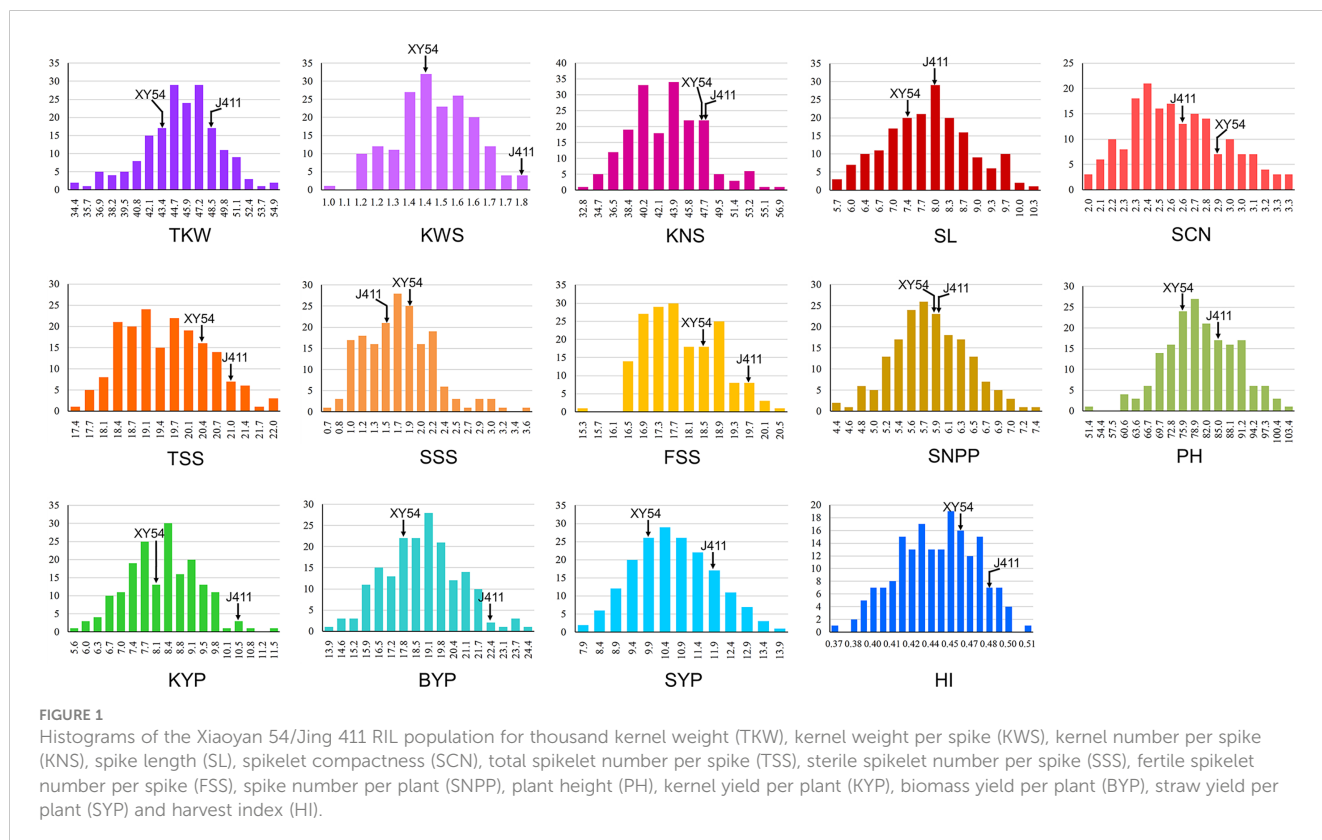
assigned to the wheat genome assembly (IWGSC RefSeq v1.1). SNP order in the present genetic map was in good agreement with that in the reference genome, except for chromosomes 2AL, 2BS and 2DS, in which a segment inversion was identified (Figures 3, 4).

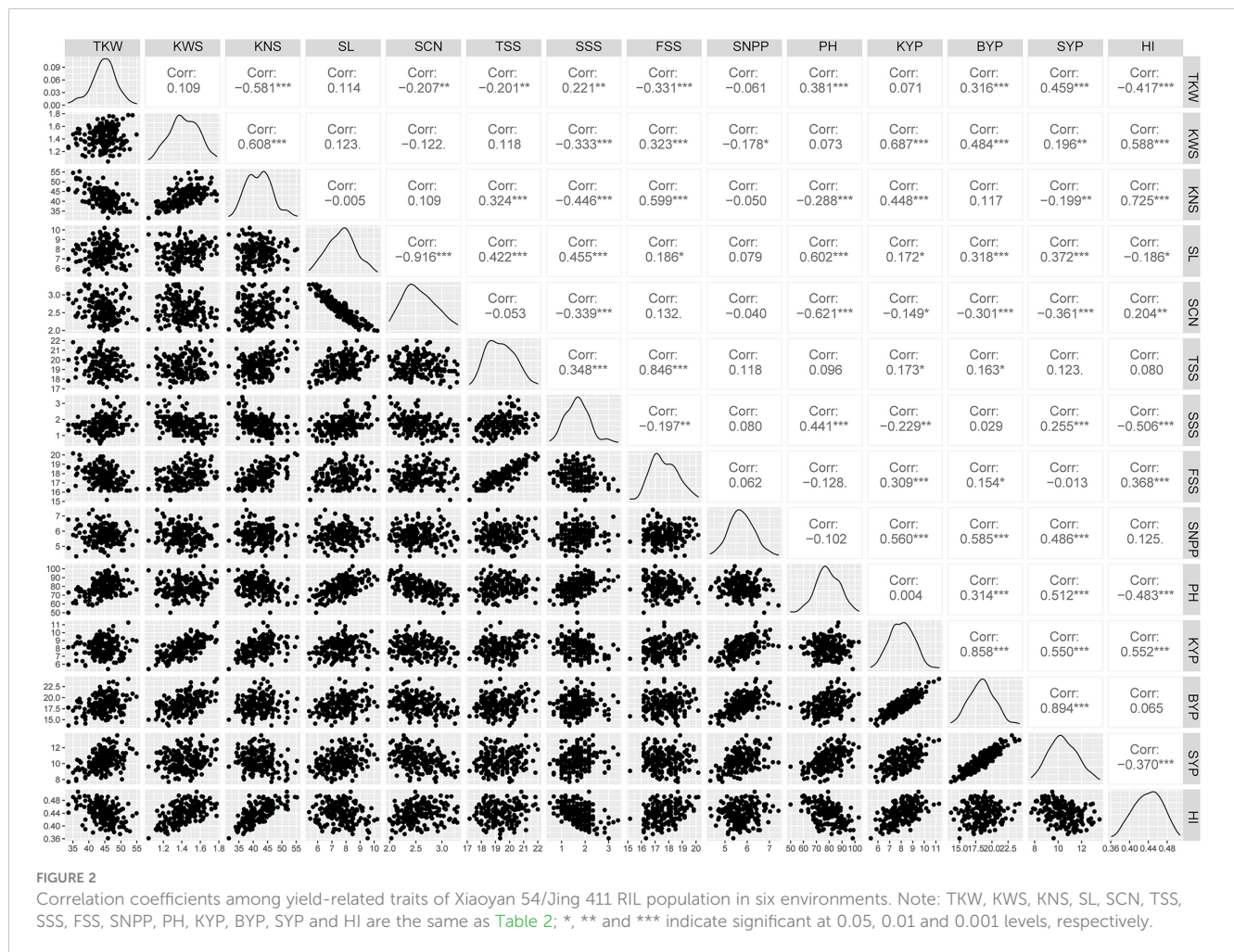
QTL mapping analysis

A total of 285 significant QTLs were detected for the 14 examined traits in six environments, explaining 0.6–34.7% of the phenotypic variation (Table 1; Supplementary Figure S1, Supplementary Table S2). One hundred and twenty-nine QTLs showed positive effect with the Xiaoyan 54 allele. For the remaining 156 QTLs, the positive phenotype was derived from Jing 411. The QTLs that could be detected in three or more environments were regarded as environmentally stable QTLs. Twelve environmentally stable QTLs were identified in this study (Table 2). The QTLs detected for each trait were showed in Supplementary Figure S1.

Kernel-related traits

Sixty-three QTLs for kernel-related traits (TKW and KWS) were detected on all chromosomes except for 7A, explaining 1.2–26.8% of the phenotypic variation (Table 2; Supplementary Figure S1, Supplementary Table S2). Of these, 23 QTLs showed positive effect with the Xiaoyan 54 allele, and 40 QTLs showed positive effect with the Jing 411 allele. Three environmentally stable QTLs for TKW were identified on chromosomes 1B, 4A and 4D (Table 2). The QTL *QTKw-1B.2* and *QTKw-4A.2* were significant in five and three environments, explaining 1.6–16.3% and 5.8–14.7% of the





phenotypic variation, respectively. Jing 411 contributed effect for increased TKW at these loci. The QTL *QTkw-4D.1* was identified in four environments, explaining 3.5–9.4% of the phenotypic variation. Xiaoyan 54 contributed effect for increased TKW at the locus.

Spike-related traits

One hundred and sixteen QTLs for spike-related traits (KNS, SL, SCN, TSS, SSS and FSS) were identified on all chromosomes, explaining 2.1–34.7% of the phenotypic variation (Table 2; Supplementary Figure S1, Supplementary Table S2). Of these, 56 QTLs showed positive effect with the Xiaoyan 54 allele and 60 QTLs showed positive effect with the Jing 411 allele. Six environmentally stable QTLs were detected on chromosomes 2D (3), 5A (2) and 7A. Of these, *QSL-2D.2* was significant for SL across all the six environments, explaining 23.9–34.7% of the phenotypic variation. The QTL was also significant for SCN across all the eleven environments (*QScn-2D.1*), explaining 13.6–22.9% of the phenotypic variation. *QSL-5A.1* was significant for SL in four environments, explaining 9.4–11.5% of the phenotypic variation. The QTL was also significant for SCN in three environments (*QScn-5A.2*), explaining 7.7–12.1% of the phenotypic variation. Jing 411 contributed effect for increased SL and SCN at the two loci. *QTss-7A.2* was significant for TSS in five environments, explaining 11.2–

23.1% of the phenotypic variation. Jing 411 contributed effect for an increased TSS at the locus. *QSSs-2D.2* was significant for SSS in three environments, explaining 9.3–13.2% of the phenotypic variation. Jing 411 contributed effect for an increased SSS at the locus.

Plant architecture-related traits

Thirty-nine QTLs for plant architecture-related traits (SNPP and PH) were detected on all chromosomes except for 1B, 6A and 7D, explaining 0.6–33.5% of the phenotypic variation (Table 2; Supplementary Figure S1, Supplementary Table S2). Of these, 23 QTLs showed positive effect with the Xiaoyan 54 allele and 16 QTLs showed positive effect with the Jing 411 allele. Two environmentally stable QTLs *QPh-2D.1* and *QPh-4B.1* were both significant for PH across all the six environments, explaining 12.5–21.0% and 20.5–33.5% of the phenotypic variation, respectively. Jing 411 contributed effect for increased PH at these two loci.

Yield-related traits

Sixty-seven QTLs for yield-related traits (KYP, BYP, SYP and HI) were detected on all chromosomes except for 1A, 4D, 5B, 5D and 7D, explaining 3.8–19.8% of the phenotypic variation (Table 2; Supplementary Figure S1, Supplementary Table S2). Of these, 28 QTLs showed positive effect with the Xiaoyan 54 allele and 39 QTLs

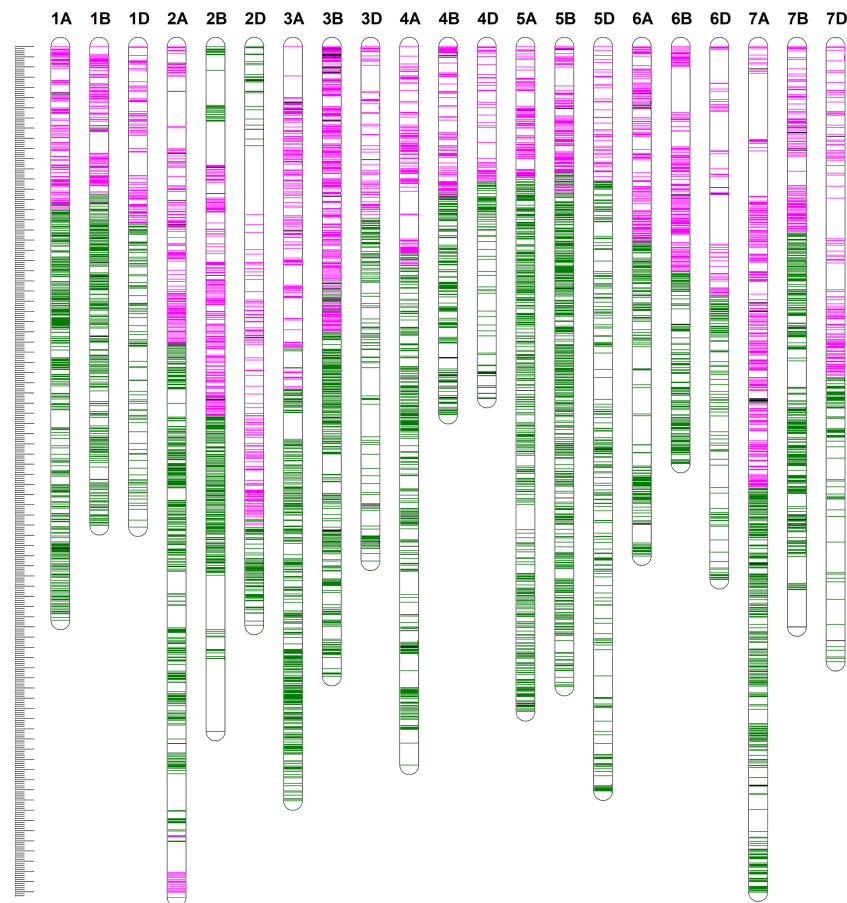


FIGURE 3

The high-density genetic linkage map of Xiaoyan 54/Jing 411 RIL population. For the redundant loci that showed co-segregation in the 182 RILs, only one unique informative marker is shown. The positions of the marker loci are indicated using a ruler on the left side. The names of the marker loci are listed to the right of the corresponding chromosomes. Loci in pink were best hits to Chinese Spring (CS) reference genome of the short arm of the corresponding chromosomes. Loci in green were best hits to CS reference genome of the long arm of the corresponding chromosomes. Loci in black were unknown.

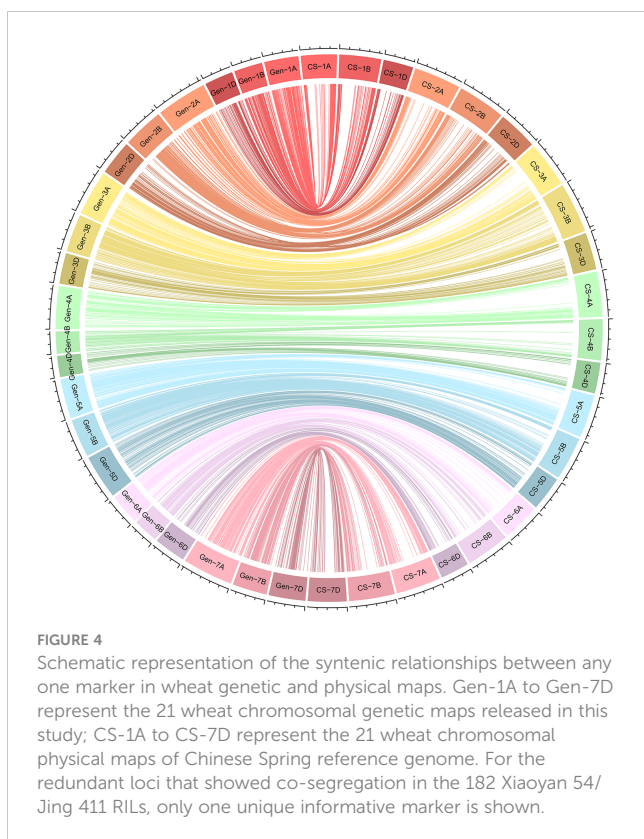
showed positive effect with the Jing 411 allele. The environmentally stable QTL *QHi-5A.1* was significant for HI in three environments, explaining 7.1–19.8% of the phenotypic variation. Jing 411 contributed effect for an increased HI at the locus.

Development of KASP markers to validate the key loci

In this study, four environmentally stable QTLs *QTKw-1B.2*, *QPh-2D.1* (*QSl-2D.2/QScn-2D.1*), *QPh-4B.1* and *QTss-7A.2* were detected in at least five environments (Figure 5). Based on the flanking marker sequence of these QTLs, eight KASP markers were designed and tested for polymorphism in the diversity panel (Supplementary Table S3, Supplementary Figure S2). Two-tailed t-test was conducted for each marker in the diversity panel for yield-related traits. Apart from the QTL *QTKw-1B.2*, the other three QTLs were validated successfully in the diversity panel.

For the QTL *QPh-2D.1* (*QSl-2D.2/QScn-2D.1*) that was significant for PH, SL and SCN across all the six environments,

two KASP markers KA196 and KA744 were developed based on the flanking SNP markers *AX-111021196* and *AX-111561744*, respectively (Figure 5; Supplementary Table S3, Supplementary Figure S2). In the diversity panel, KA196 and KA744 were significantly related to PH and SL in three to four environments (Figure 6). The accessions with Xiaoyan 54-derived alleles had lower PH and shorter SL, compared with the accessions with Jing 411-derived alleles. For the QTL *QPh-4B.1* that was significant for PH across all the six environments, two KASP markers KA058 and KA957 were developed based on the flanking SNP markers *AX-109850058* and *AX-110713957*, respectively (Supplementary Table S3, Supplementary Figure S2). In the diversity panel, the two KASP markers were significantly related to PH in four environments (Supplementary Figure S3). The accessions with Xiaoyan 54-derived alleles had lower PH, compared with the accessions with Jing 411-derived alleles. For the QTL *QTss-7A.3* that was significant for TSS in five environments, two KASP markers KA446 and KA493 were developed based on the flanking SNP markers *AX-108794446* and *AX-108848493*, respectively (Supplementary Table S3, Supplementary Figure S2). In the diversity panel, the KASP



marker KA493 were significantly related to TSS (Supplementary Figure S4). The accessions with Xiaoyan 54-derived alleles had more TSS than those with Jing 411-derived alleles.

Prediction of candidate genes in the four key loci

In the present study, four important loci were detected in at least five environments. Of these, the QTL *QTKw-1B.2* was identified for TKW in five environments (Figure 5). The confidence interval of *QTKw-1B.2* was bounded by SNP markers AX-109873144 and AX-108946001, corresponding to a physical distance of ~ 1.3 Mb (627,922,061– 629,168,892 bp, RefSeq v1.1), which contained 23 high-confidence annotated genes (Supplementary Table S4). Among these genes, *TraesCS1B01G396600*, which encodes the U4/U6 small nuclear ribonucleoprotein Prp31, showed higher expression level in developing wheat grain (Ramirez-Gonzalez et al., 2018) and may be the candidate gene for *QTKw-1B.2*. Further research is needed to firmly conclude the candidate gene. The QTL *QPh-2D.1* (*QSL-2D.2/QScn-2D.1*) was significantly related to PH, SL and SCN across all the six environments (Figure 5). In the diversity panel, two KASP markers based on the flanking SNP markers of the QTL were significantly related to PH and SL (Figure 6). The confidence interval of *QPh-2D.1* (*QSL-2D.2/QScn-2D.1*) was bounded by SNP markers AX-111021196 and AX-111561744, corresponding to a physical distance of ~ 0.9 Mb (22,498,824– 23,416,219 bp, RefSeq v1.1), which contained 45 high-confidence annotated genes

(Supplementary Table S5). The famous reduced height (*Rht*) gene *Rht8* was reported in this interval (Chai et al., 2019). Chai et al. (2022) isolated the candidate gene *TraesCSU02G024900* for *Rht8* via map-based gene cloning. The gene encodes a protein containing a zinc finger BED-type motif and an RNase H-like domain that regulates plant height via influencing bioactive gibberellin biosynthesis. Similar results were found by Xiong et al. (2022). Therefore, *TraesCSU02G024900* might be the candidate gene for *QPh-2D.1*. The QTL *QPh-4B.1* was significant for PH across all the six environments (Figure 5). In the diversity panel, two KASP markers were significantly related to PH in four environments (Supplementary Figure S3). The confidence interval of *QPh-4B.1* was bounded by SNP markers AX-109850058 and AX-110713957, corresponding to a physical distance of ~ 2.0 Mb (40,904,736– 42,912,713 bp, RefSeq v1.1), which contained 18 high-confidence annotated genes (Supplementary Table S6). Among these genes, *TraesCS4B02G053600* is an ortholog of the rice gene *Decrease in DNA Methylation 1* (*OsDDM1*), which was reported to be related to dwarf phenotypes (Higo et al., 2012). The QTL *QTss-7A.3* was significant for TSS in five environments (Figure 5). In the diversity panel, the KASP marker KA493 were significantly related to TSS (Supplementary Figure S4). The confidence interval of *QTss-7A.3* was bounded by SNP markers AX-108794446 and AX-108848493, corresponding to a physical distance of ~ 2.2 Mb (672,893,634– 675,112,612 bp, RefSeq v1.1), which contained 27 high-confidence annotated genes (Supplementary Table S7). Among these genes, *TraesCS7A02G481600*, which is the A-genome homeolog of *WHEAT ORTHOLOG OF APO1* (*WAPO-A1*), was reported to be the leading candidate gene for *QTss-7A.3* affecting spikelet number per spike (Kuzay et al., 2019, Kuzay et al., 2022).

Discussion

Comparison of the major QTLs with previous observations

In this study, we detect 285 QTLs for 14 yield-related traits using a high-density linkage map. Of these, the QTL *QTKw-1B.2* was identified for TKW in five environments and was located in the position interval 627.9–629.2 Mb of chromosome 1B (RefSeq v1.1). In the previous studies, Li et al. (2019) performed genome-wide association study (GWAS) in 166 wheat cultivars and identified a significant QTL associated with TKW in the interval 658.7–662.5 Mb of 1B based on the Wheat 90K and 660K SNP arrays. Quarrie et al. (2005) detected a QTL for TKW near the position 555.93 Mb using a doubled-haploid (DH) population derived from the cross of Chinese Spring and SQ1. Cui et al. (2014) also identified a QTL for TKW near the position 555.93 Mb using three related RIL populations. Zanke et al. (2015) performed GWAS in a wheat panel and detected three significant SNP loci on 1B associated with TKW. Pang et al. (2020) conducted a large-scale GWAS using a panel of 768 wheat cultivars and detected a significant QTL for TKW in the interval 667.9–668.1 Mb of 1B under three environments. We located the three SNP loci at the positions 14.1Mb, 560.5Mb and 649.1Mb of 1B by BLAST-searching

TABLE 1 Summary information of the Xiaoyan 54/Jing 411 RIL high-density genetic map.

| Chromosomes | Locus number | Map length (cM) | Average distance (cM) | Max distance (cM) |
|-------------|--------------|-----------------|-----------------------|-------------------|
| 1A | 460 | 281.69 | 0.61 | 8.52 |
| 1B | 398 | 235.21 | 0.59 | 10.48 |
| 1D | 215 | 235.86 | 1.10 | 11.50 |
| 2A | 485 | 417.71 | 0.86 | 18.19 |
| 2B | 420 | 336.30 | 0.80 | 35.66 |
| 2D | 234 | 284.15 | 1.21 | 33.83 |
| 3A | 457 | 370.32 | 0.81 | 14.31 |
| 3B | 565 | 309.22 | 0.55 | 10.07 |
| 3D | 219 | 252.58 | 1.15 | 15.96 |
| 4A | 381 | 352.85 | 0.93 | 11.78 |
| 4B | 234 | 180.62 | 0.77 | 7.01 |
| 4D | 132 | 172.71 | 1.31 | 8.86 |
| 5A | 453 | 326.46 | 0.72 | 12.22 |
| 5B | 537 | 314.21 | 0.59 | 9.32 |
| 5D | 285 | 365.45 | 1.28 | 12.86 |
| 6A | 345 | 250.10 | 0.72 | 14.20 |
| 6B | 314 | 204.58 | 0.65 | 22.42 |
| 6D | 177 | 261.74 | 1.48 | 24.34 |
| 7A | 561 | 415.14 | 0.74 | 32.39 |
| 7B | 410 | 284.86 | 0.69 | 18.17 |
| 7D | 260 | 302.01 | 1.16 | 19.93 |
| Group1 | 1073 | 752.76 | 0.70 | 35.66 |
| Group2 | 1139 | 1038.15 | 0.91 | 35.66 |
| Group3 | 1241 | 932.12 | 0.75 | 15.96 |
| Group4 | 747 | 706.17 | 0.95 | 11.78 |
| Group5 | 1275 | 1006.13 | 0.79 | 12.86 |
| Group6 | 836 | 716.42 | 0.86 | 24.34 |
| Group7 | 1231 | 1002.01 | 0.81 | 32.39 |
| GenomeA | 3142 | 2414.27 | 0.77 | 32.39 |
| GenomeB | 2878 | 1865.00 | 0.65 | 35.66 |
| GenomeD | 1522 | 1874.50 | 1.23 | 33.83 |
| Total | 7542 | 6153.76 | 0.82 | 35.66 |

against the Chinese Spring reference genome sequence. By comparison of the QTL position with previous observations, we found that the QTL *QTkw-1B.2* identified in this study was different from those from previous studies and may be a novel QTL for TKW, which represented a valuable target for map-based cloning and marker-assisted selection to enhance grain yield in wheat breeding.

The QTL *QPh-4B.1* was significant for PH across all the six environments. In the diversity panel, two KASP markers KA058

and KA957 were significantly related to PH in four environments. The QTL was located in the position interval 40.9–42.9 Mb of chromosome 4B (RefSeq v1.1). The “Green revolution” gene *Rht-B1b* was reported at 30.86 Mb of 4B (Xu et al., 2019), which is different from the QTL *QPh-4B.1*. The other height-reducing genes *Rht3* (*Rht-B1c*), *Rht11* (*Rht-B1e*) and *Rht17* (*Rht-B1p*) on 4B were allelic to *Rht-B1b* (Zhang et al., 2021). In the previous studies, Zhang et al. (2017a) detected a QTL for PH near the locus *Rht-B1b* in multi-environments using a 660K high-density map, and we

TABLE 2 Environmentally stable QTLs for each yield-related trait of Xiaoyan 54/Jing 411 RIL population in six environments.

| Traits | QTL | Env. | Chr. | Site (cM) | LeftMarker | RightMarker | LOD | PVE ^a (%) | Add ^b |
|-----------------------------------|------------------|------|------|-----------|--------------|--------------|------|----------------------|------------------|
| Thousand kernel weight | <i>QTkw-1B.2</i> | 6CK | 1B | 136 | AX-109873144 | AX-108946001 | 3.9 | 1.6 | -0.67 |
| (TKW) | | 6LN | 1B | 136 | AX-109873144 | AX-108946001 | 4.6 | 6.9 | -1.01 |
| | | 7CK | 1B | 136 | AX-109873144 | AX-108946001 | 6.0 | 4.5 | -1.10 |
| | | 7LN | 1B | 136 | AX-109873144 | AX-108946001 | 10.9 | 16.3 | -1.69 |
| | | 7LP | 1B | 136 | AX-109873144 | AX-108946001 | 7.1 | 14.4 | -1.52 |
| | <i>QTkw-4A.2</i> | 6CK | 4A | 38 | AX-94419996 | AX-108742845 | 25.8 | 14.7 | -2.01 |
| | | 6LN | 4A | 38 | AX-94419996 | AX-108742845 | 7.4 | 11.7 | -1.30 |
| | | 7LP | 4A | 38 | AX-94419996 | AX-108742845 | 3.0 | 5.8 | -0.95 |
| | <i>QTkw-4D.1</i> | 6LP | 4D | 36 | AX-108735064 | AX-110003964 | 4.1 | 3.5 | 1.00 |
| | | 7LP | 4D | 36 | AX-108735064 | AX-110003964 | 4.6 | 9.4 | 1.22 |
| | | 7CK | 4D | 38 | AX-110003964 | AX-109343336 | 4.5 | 3.5 | 0.96 |
| | | 7LN | 4D | 39 | AX-109343336 | AX-111048443 | 3.5 | 4.7 | 0.90 |
| Spike length | <i>QSl-2D.2</i> | 6CK | 2D | 145 | AX-111021196 | AX-111561744 | 16.7 | 23.9 | -0.52 |
| (SL) | | 6LN | 2D | 145 | AX-111021196 | AX-111561744 | 21.4 | 34.7 | -0.51 |
| | | 6LP | 2D | 145 | AX-111021196 | AX-111561744 | 17.6 | 25.7 | -0.52 |
| | | 7CK | 2D | 145 | AX-111021196 | AX-111561744 | 26.9 | 28.2 | -0.68 |
| | | 7LN | 2D | 145 | AX-111021196 | AX-111561744 | 14.7 | 28.0 | -0.51 |
| | | 7LP | 2D | 145 | AX-111021196 | AX-111561744 | 15.9 | 28.3 | -0.68 |
| | <i>QSl-5A.1</i> | 6CK | 5A | 107 | AX-108742477 | AX-108739527 | 7.5 | 9.5 | -0.33 |
| | | 6LN | 5A | 107 | AX-108742477 | AX-108739527 | 8.2 | 11.3 | -0.29 |
| | | 6LP | 5A | 107 | AX-108742477 | AX-108739527 | 8.8 | 11.5 | -0.34 |
| | | 7LN | 5A | 107 | AX-108742477 | AX-108739527 | 5.5 | 9.4 | -0.29 |
| Spikelet compactness | <i>QScn-2D.1</i> | 6CK | 2D | 145 | AX-111021196 | AX-111561744 | 10.4 | 19.4 | 0.16 |
| (SCN) | | 6LN | 2D | 145 | AX-111021196 | AX-111561744 | 10.1 | 14.2 | 0.12 |
| | | 6LP | 2D | 145 | AX-111021196 | AX-111561744 | 11.1 | 22.9 | 0.16 |
| | | 7CK | 2D | 145 | AX-111021196 | AX-111561744 | 7.7 | 15.0 | 0.11 |
| | | 7LN | 2D | 145 | AX-111021196 | AX-111561744 | 14.1 | 14.5 | 0.15 |
| | | 7LP | 2D | 145 | AX-111021196 | AX-111561744 | 7.9 | 13.6 | 0.12 |
| | <i>QScn-5A.2</i> | 6LN | 5A | 107 | AX-108742477 | AX-108739527 | 8.6 | 12.1 | 0.11 |
| | | 6LP | 5A | 107 | AX-108742477 | AX-108739527 | 6.2 | 12.1 | 0.11 |
| | | 7CK | 5A | 107 | AX-108742477 | AX-108739527 | 4.1 | 7.7 | 0.07 |
| Total spikelet number per spike | <i>QTss-7A.2</i> | 6CK | 7A | 300 | AX-108794446 | AX-108848493 | 4.7 | 11.2 | -0.47 |
| (TSS) | | 6LN | 7A | 300 | AX-108794446 | AX-108848493 | 12.3 | 22.4 | -0.72 |
| | | 6LP | 7A | 300 | AX-108794446 | AX-108848493 | 9.1 | 17.4 | -0.52 |
| | | 7CK | 7A | 300 | AX-108794446 | AX-108848493 | 8.8 | 17.8 | -0.60 |
| | | 7LN | 7A | 300 | AX-108794446 | AX-108848493 | 12.8 | 23.1 | -0.74 |
| Sterile spikelet number per spike | <i>QSSs-2D.2</i> | 6CK | 2D | 144 | Xwmc112 | AX-111021196 | 4.1 | 9.4 | -0.22 |
| (SSS) | | 6LP | 2D | 144 | Xwmc112 | AX-111021196 | 7.4 | 13.2 | -0.31 |
| | | 7LN | 2D | 145 | AX-111021196 | AX-111561744 | 4.6 | 9.3 | -0.31 |

(Continued)

TABLE 2 Continued

| Traits | QTL | Env. | Chr. | Site (cM) | LeftMarker | RightMarker | LOD | PVE ^a (%) | Add ^b |
|---------------|-----------------|------|------|-----------|--------------|--------------|------|----------------------|------------------|
| Plant height | <i>QPh-2D.1</i> | 6CK | 2D | 145 | AX-111021196 | AX-111561744 | 11.4 | 12.5 | -3.79 |
| (PH) | | 6LN | 2D | 145 | AX-111021196 | AX-111561744 | 12.8 | 16.2 | -4.14 |
| | | 7CK | 2D | 145 | AX-111021196 | AX-111561744 | 16.8 | 21.0 | -4.55 |
| | | 7LN | 2D | 145 | AX-111021196 | AX-111561744 | 11.5 | 15.2 | -3.49 |
| | | 6LP | 2D | 146 | AX-111561744 | AX-111500777 | 15.2 | 14.9 | -3.86 |
| | | 7LP | 2D | 146 | AX-111561744 | AX-111500777 | 9.8 | 14.8 | -3.54 |
| | <i>QPh-4B.1</i> | 6CK | 4B | 69 | AX-109850058 | AX-110713957 | 24.6 | 32.1 | -6.01 |
| | | 6LN | 4B | 69 | AX-109850058 | AX-110713957 | 23.0 | 33.4 | -5.88 |
| | | 6LP | 4B | 69 | AX-109850058 | AX-110713957 | 27.9 | 33.5 | -5.74 |
| | | 7CK | 4B | 69 | AX-109850058 | AX-110713957 | 20.6 | 27.6 | -5.17 |
| | | 7LN | 4B | 69 | AX-109850058 | AX-110713957 | 20.6 | 30.7 | -4.90 |
| | | 7LP | 4B | 69 | AX-109850058 | AX-110713957 | 13.1 | 20.5 | -4.15 |
| Harvest index | <i>QHi-5A.1</i> | 6CK | 5A | 174 | AX-108926070 | AX-108801270 | 6.9 | 7.1 | -0.01 |
| (HI) | | 6LN | 5A | 174 | AX-108926070 | AX-108801270 | 7.5 | 12.1 | -0.01 |
| | | 6LP | 5A | 174 | AX-108926070 | AX-108801270 | 11.0 | 19.8 | -0.02 |

^aPVE indicates phenotypic variation explained by each QTL.

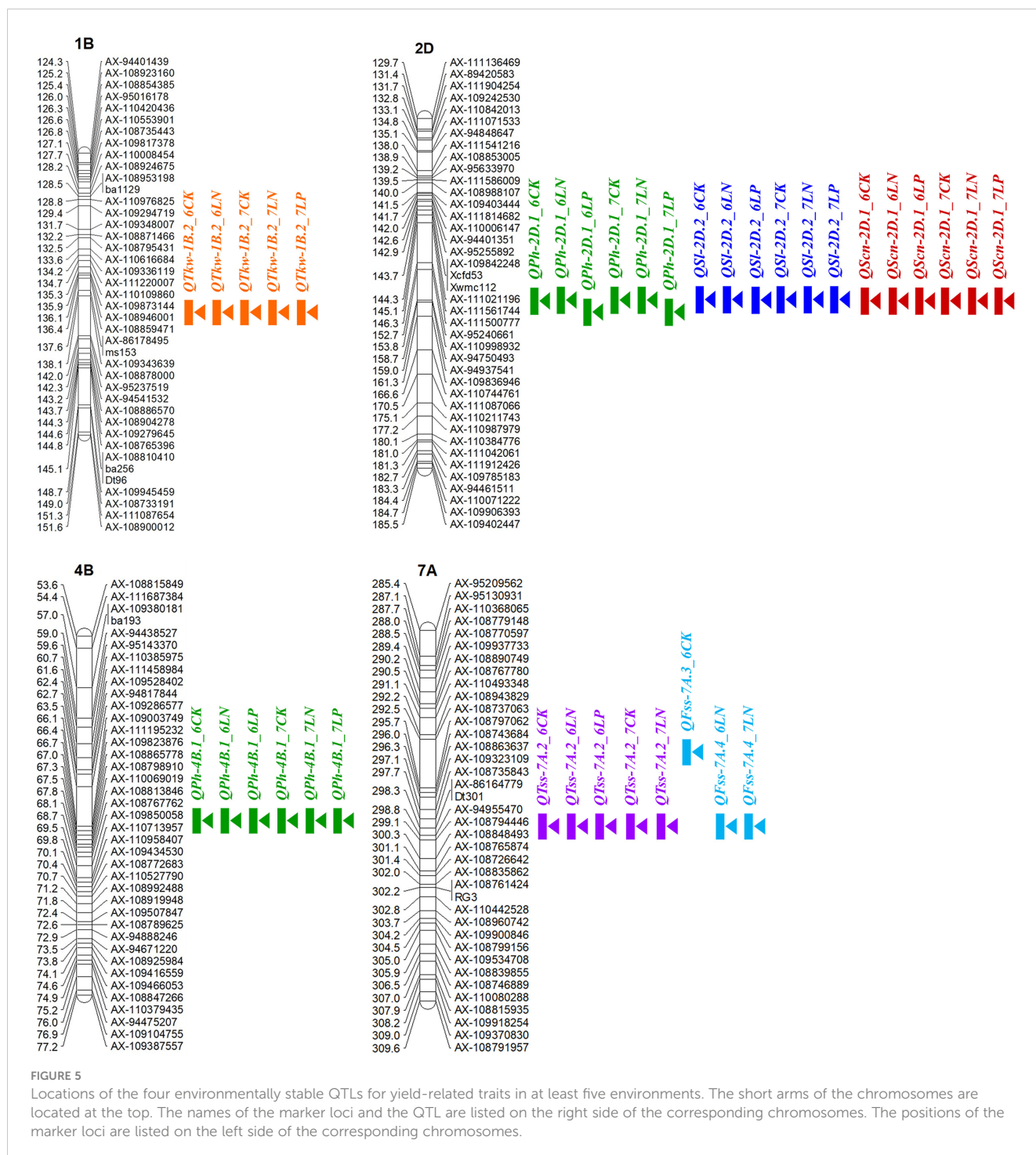
^bThe positive and negative additive values indicate Xiaoyan 54 and Jing 411 contributed increasing alleles for corresponding QTLs, respectively.

located it to the 27.4–28.9 Mb by BLAST-searching against the Chinese Spring reference genome sequence. Li et al. (2018) identified a QTL for PH at 25.8 Mb of 4B in three bread wheat populations using the Wheat 90K SNP array. Tadesse et al. (2015) detected a locus associated with PH at the position 68.1 Mb of 4B using a total of 120 elite wheat accessions. Guo et al. (2015) identified a locus *Xgwm495* associated with PH at the position 482.8 Mb using a set of 230 wheat cultivars by association mapping. Zhuang et al. (2021) cloned a PH-related gene *TaSRL1* from wheat at the position 585.8 Mb of 4B. By comparison of the QTL *QPh-4B.1* position with previous observations, we found that the QTL for PH detected in this study was different from those from previous studies and may be a novel QTL, which deserved for further studies including positional cloning and marker-assisted selection.

In this study, QTL *QPh-2D.1* (*QSL-2D.2/QScn-2D.1/QSss-2D.2/QTss-2D.1*) was identified for PH, SL and SCN across all the six environments, for SSS in three environments and for TSS in one environment. Two KASP markers based on the flanking SNP markers of the QTL were significantly related to PH and SL in the diversity panel (Figure 6). The QTL was located in the position interval 22.5–23.4 Mb of chromosome 2D (RefSeq v1.1). In the previous studies, the famous *Rht* gene *Rht8* was reported to be located on the same genomic interval (Chai et al., 2019). Chai et al. (2022) isolated the candidate gene for *Rht8* via map-based gene cloning and confirmed that loss of RNHL-D1 is responsible for semi-dwarf trait in *Rht8*-carrying wheat plants. Xiong et al. (2022) identified two new semi-dwarf wheat mutants that are allelic to *Rht8* and revealed the complexity and evolutionary history of *Rht8* in common wheat. Zhai et al. (2016) detected a pleiotropic QTL for PH, SL and SSS at the position 23.0 Mb on 2D using the RIL

population derived from Yumai 8679 and Jing 411. Xu et al. (2014) identified a QTL cluster in the interval 23.0–24.7 Mb of 2D controlling PH, SL, SSS and TSS using Xiaoyan 54/Jing 411 RIL population. Ma et al. (2007) identified a major QTL for SL on 2DS using Nanda 2419/Wangshuibai RIL population. Then the QTL was precisely mapped near the position 23.0 Mb (Wu et al., 2013). Zhou et al. (2017) detected a QTL for SL in the interval 22.9–23.7 Mb of 2D using a soft red winter wheat DH population. Using another DH population, Sourdille et al. (2003) detected a QTL for SL in the interval 20.4–24.3 Mb of 2D. Ma et al. (2018) identified a major QTL for FSS in the similar position through GWAS and found the QTL could affect SL, TSS and SSS. Therefore, it seems that the QTL on 2D in this study contained *Rht8* gene and was a pleiotropic locus that played an important role in affecting PH, SL, TSS, SSS and FSS.

The QTL *QTss-7A.3* (*QFss-7A.4*) was detected for TSS and FSS in five and two environments, respectively. In the diversity panel, the KASP marker KA493 were significantly related to TSS (Supplementary Figure S4). The QTL was located in the position interval 672.9–675.1 Mb of chromosome 7A (RefSeq v1.1). In the previous studies, Xu et al. (2014) identified a QTL for TSS, FSS and SSS in the interval 668.0–679.9 Mb of 7A using Xiaoyan 54/Jing 411 RIL population. Zhang et al. (2018a) identified a SNP for TSS at the position 674.3 Mb of 7A in a spring wheat panel and validated the SNP in a biparental population. Boeven et al. (2016) conducted GWAS and detected a QTL for TSS near the position 674.3 Mb of 7A in a diverse set of 209 winter bread wheat lines. Faris et al. (2014) identified a QTL for TSS in the interval 671.4–674.3 Mb of 7A using a RIL population derived from a cross between a cultivated emmer accession and a durum wheat cultivar. Voss-Fels et al. (2019) performed GWAS in a panel of 220 winter wheats and detected a



highly significant QTL for TSS in the interval 672.0–674.3 Mb of 7A. All these results suggested that the QTL on 7A was a key locus and showed significant effects for TSS in various environments. What's more, Kuzay et al. (2019) delimited this QTL to an 87-kb region (674,019,191–674,106,327 bp, RefSeq v1.1) containing four candidate genes and identified *WAPO-A1* as the most promising candidate gene. Loss-of-function mutations in the *WAPO-A1* gene reduced TSS and additional transgenic copies of this gene increased TSS. Haplotype analysis showed that H2 variant is associated with the largest increases in TSS and KNS in field experiments (Kuzay

et al., 2022). Therefore, the utilization of the *WAPO-A1* variant represents a promising opportunity to improve grain yield in wheat.

The high-density linkage map and comparative mapping

Constructing a high-quality and saturated genetic map is the prerequisite of QTL mapping. Based on the new sequencing technologies, a great number of SNPs have been identified and

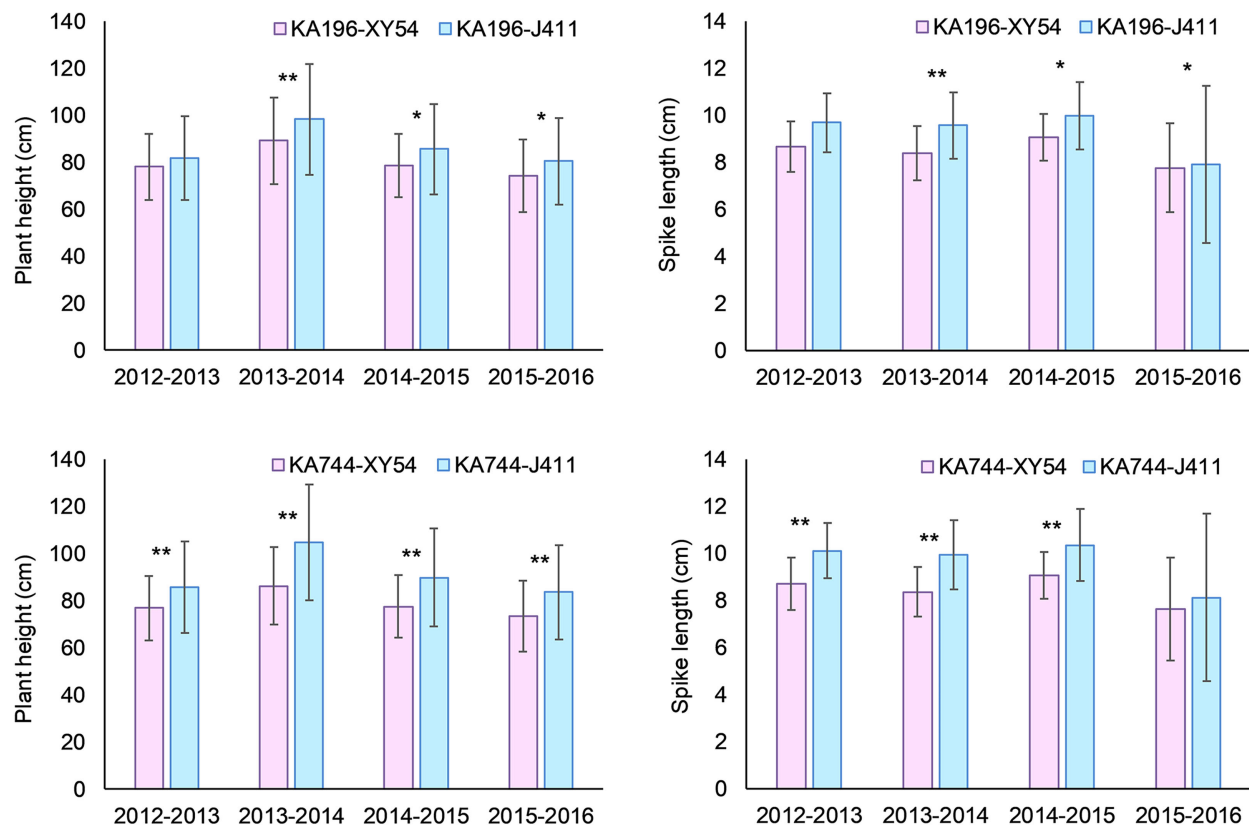


FIGURE 6

Mean difference in plant height (PH) and spike length (SL) between Xiaoyan 54 and Jing 411-derived alleles of KA196 and KA744 in the diversity panel. XY54 indicates the "Xiaoyan 54" allele; J411 indicates the "Jing 411" allele; * and ** indicate significant at 0.05 and 0.01 levels, respectively.

used for genetic map construction in wheat. Some high-density maps were reported using the high-throughput microarray genotyping method, such as the Wheat 9K, 90K and 660K arrays (Wu et al., 2015; Zhai et al., 2016; Cui et al., 2017). Sun et al. (2020) evaluated seven widely used high-throughput wheat arrays (Wheat 9K, 15K, 35K, 55K, 90K, 820K and 660K arrays) in terms of their SNP number, distribution, density, associated genes, heterozygosity, and application. The results suggested that the Wheat 660K SNP array is reliable and cost-effective and may be the best choice for targeted genotyping and marker-assisted selection in wheat genetic improvement. In the present study, we genotyped the Xiaoyan 54/Jing 411 RIL population using the Wheat 660K SNP array and developed a high-density linkage map of 7,542 polymorphisms markers. Based on the SNP flanking sequences, we assigned the markers to the reference genome of Chinese Spring. As shown in Figure 3, the genetic and physical positions of the mapped markers were generally in agreement. The SNP order in the present genetic map was also in good agreement with that in the physical position (Figure 4), which prompted us to search for candidate genes of major targeted QTLs. Using the high-density genetic map of Xiaoyan 54/Jing 411 RIL population, we identified 285 significant QTLs for the 14 examined traits in six environments, explaining 0.6–34.7% of the phenotypic variation (Supplementary Figure S1; Table 1; Supplementary Table S2). The number of QTLs significantly related to each trait ranged from 11 to 41. Xu et al.

(2014) genotyped the Xiaoyan 54/Jing 411 RIL population using gel-based markers and constructed a genetic linkage map with 555 polymorphic loci. Based on the genetic map, 89 QTLs for the same 14 yield-related traits were identified and the number of QTLs significantly related to each trait ranged from 2 to 14, which was much less than those detected in this study. The results showed that more QTLs could be identified using a high-density linkage map. What's more, Xu et al. (2014) identified a major QTL (*QTss-7A*) for TSS in the interval *Xbarc192-Xbarc253* in five environments, which could explain 7.1–20.5% of the phenotypic variation. We located the two flanking markers at the positions 668.0 Mb and 680.0 Mb of 7A (RefSeq v1.1). In this study, we also detected a QTL for TSS in five environments, explaining 11.2–23.1% of the phenotypic variation. The QTL was located in the position interval 672.9–675.1 Mb of 7A (RefSeq v1.1). By comparison of the positions the two QTLs, we found that the confidence intervals of QTLs identified using a high-density linkage map were much smaller than those identified using a low-density map. It is worth noting that Xu et al. (2014) detected a QTL for TKW on chromosome 1B in 7LN and 7LP two environments, explaining 9.4% and 9.9% of the phenotypic variation, respectively, while in this study, the QTL *QTkw-1B.2* was significant in five environments, explaining 16.3% and 14.4% of the phenotypic variation in the 7LN and 7LP environments, respectively, and 1.6–6.9% of phenotypic variation in other three environments. These results showed that QTLs with lower

phenotypic variance explained could be detected using a high-density linkage map. Taken together, it is more efficient to detect QTLs for yield-related traits using an improved high-density linkage genetic map.

Potential implications in wheat breeding

How to increase wheat yield has been a major focus of most wheat breeders. Wheat yield is significantly influenced by environment, which presents a major challenge to select high-yielding lines at the early stages of breeding programs. In contrast, yield-related traits, such as TKW, SL, PH and TSS, are less influenced by environment. Consequently, more effort has been put into yield-related traits to improve wheat yield. Identification of stable major QTLs for yield-related traits is of high importance in molecular breeding. In the present study, we detected four environmentally stable QTLs for yield-related traits in at least five environments using a high-density genetic map based on the Wheat 660K SNP array. Of these, the QTL *QPh-2D.1* (*QSl-2D.2/QScn-2D.1*) was identified for PH, SL and SCN across all the six environments. The QTL *QTss-7A.3* was detected for TSS in five environments. By comparison of the QTLs with previous observations, we found that the two QTLs showed constant effects on their corresponding yield-related traits in different genetic backgrounds and were strongly selected in breeding. The QTLs *QTKw-1B.2* and *QPh-4B.1* were identified for TKW and PH in five and six environments, respectively. The two QTLs were different from those from previous studies and might be novel QTLs. In the diversity panel, *QPh-2D.1* (*QSl-2D.2/QScn-2D.1*), *QTss-7A.3* and *QPh-4B.1* were validated successfully by developing KASP markers. These major QTLs represented a valuable target for marker-assisted selection to improve yield-related traits. The availability of time-saving and cost-effective KASP markers could facilitate their use in wheat breeding. With the application of high-density linkage maps in QTL detection and user-friendly flanking markers, wheat breeding by molecular design is not a distant goal.

Conclusion

We constructed a high-density genetic map using an RIL population with the Wheat 660K SNP array. The genetic map showed high collinearity with the wheat genome assembly. Using the high-density genetic map, we conducted QTL mapping for 14 yield-related traits in six environments. Four major QTLs, *QTKw-1B.2*, *QPh-2D.1* (*QSl-2D.2/QScn-2D.1*), *QPh-4B.1* and *QTss-7A.3*, were detected in at least five environments. Of these, *QPh-2D.1* (*QSl-2D.2/QScn-2D.1*), *QPh-4B.1* and *QTss-7A.3* were successfully validated in the natural population based on the developed KASP markers. By comparing with results from previous studies, we found that *QTKw-1B.2* and *QPh-4B.1* should be novel QTLs. The

identified QTLs and the developed KASP marker will be valuable for further positional cloning and marker-assisted selection in wheat breeding programs.

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

DA, DL, YT and AZ conceived the research. DL, YT and AZ constructed the RIL population. FM, YX and RW performed phenotypic assessments. FM and YX carried out statistics analysis, QTL mapping and developed the KASP markers. FM wrote the manuscript. DA and DL supervised and revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was financially supported by the National Key Research and Development Program of China (no. 2021YFD1200600), the National Natural Science Foundation of China (no. 32101686), and the Hebei Province Key Research and Development Program (no. 22326306D).

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.2023.1138696/full#supplementary-material>

References

- Allen, A. M., Winfield, M. O., Burrridge, A. J., Downie, R. C., Benbow, H. R., Barker, G. L., et al. (2017). Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (*Triticum aestivum*). *Plant Biotechnol. J.* 15, 390–401. doi: 10.1111/pbi.12635
- Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361, eaar7191. doi: 10.1126/science.aar7191
- Boeven, P. H., Longin, C. F., Leiser, W. L., Kollers, S., Ebmeyer, E., and Wurschum, T. (2016). Genetic architecture of male floral traits required for hybrid wheat breeding. *Theor. Appl. Genet.* 129, 2343–2357. doi: 10.1007/s00122-016-2771-6
- Chai, L., Chen, Z., Bian, R., Zhai, H., Cheng, X., Peng, H., et al. (2019). Dissection of two quantitative trait loci with pleiotropic effects on plant height and spike length linked in coupling phase on the short arm of chromosome 2D of common wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 1815–1831. doi: 10.1007/s00122-019-03318-z
- Chai, L., Xin, M., Dong, C., Chen, Z., Zhai, H., Zhuang, J., et al. (2022). A natural variation in ribonuclease h-like gene underlies *Rht8* to confer "Green revolution" trait in wheat. *Mol. Plant* 15, 377–380. doi: 10.1016/j.molp.2022.01.013
- Chang, J., Hao, C., Chang, X., Zhang, X., and Jing, R. (2014). HapIII of *TaSAP1-A1*, a positively selected haplotype in wheat breeding. *J. Integr. Agr.* 13, 1462–1468. doi: 10.1016/s2095-3119(14)60808-x
- Chen, S., Gao, R., Wang, H., Wen, M., Xiao, J., Bian, N., et al. (2014). Characterization of a novel reduced height gene (*Rht23*) regulating panicle morphology and plant architecture in bread wheat. *Euphytica* 203, 583–594. doi: 10.1007/s10681-014-1275-1
- Cui, F., Zhang, N., Fan, X. L., Zhang, W., Zhao, C. H., Yang, L. J., et al. (2017). Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. *Sci. Rep.* 7, 3788. doi: 10.1038/s41598-017-04028-6
- Cui, F., Zhao, C., Ding, A., Li, J., Wang, L., Li, X., et al. (2014). Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. *Theor. Appl. Genet.* 127, 659–675. doi: 10.1007/s00122-013-2249-8
- Dong, L., Wang, F., Liu, T., Dong, Z., Li, A., Jing, R., et al. (2014). Natural variation of *TaGASR7-A1* affects grain length in common wheat under multiple cultivation conditions. *Mol. Breed.* 34, 937–947. doi: 10.1007/s11032-014-0087-2
- Faris, J. D., Zhang, Q., Chao, S., Zhang, Z., and Xu, S. S. (2014). Analysis of agronomic and domestication traits in a durum × cultivated emmer wheat population using a high-density single nucleotide polymorphism-based linkage map. *Theor. Appl. Genet.* 127, 2333–2348. doi: 10.1007/s00122-014-2380-1
- Ford, B. A., Foo, E., Sharwood, R., Karafiatova, M., Vrana, J., MacMillan, C., et al. (2018). *Rht18* semidwarfism in wheat is due to increased GA 2-oxidaseA9 expression and reduced GA content. *Plant Physiol.* 177, 168–180. doi: 10.1104/pp.18.00023
- Guo, J., Hao, C., Zhang, Y., Zhang, B., Cheng, X., Qin, L., et al. (2015). Association and validation of yield-favored alleles in Chinese cultivars of common wheat (*Triticum aestivum* L.). *PLoS One* 10, e0130029. doi: 10.1371/journal.pone.0130029
- Guo, Y., Sun, J., Zhang, G., Wang, Y., Kong, F., Zhao, Y., et al. (2013). Haplotype, molecular marker and phenotype effects associated with mineral nutrient and grain size traits of *TaGS1a* in wheat. *Field Crops Res.* 154, 119–125. doi: 10.1016/j.fcr.2013.07.012
- Hanif, M., Gao, F., Liu, J., Wen, W., Zhang, Y., Rasheed, A., et al. (2015). *TaTGW6-A1*, an ortholog of rice *TGW6*, is associated with grain weight and yield in bread wheat. *Mol. Breed.* 36, 1. doi: 10.1007/s11032-015-0425-z
- Higo, H., Tahir, M., Takashima, K., Miura, A., Watanabe, K., Tagiri, A., et al. (2012). *DDM1* (decrease in DNA methylation) genes in rice (*Oryza sativa*). *Mol. Genet. Genomics* 287, 785–792. doi: 10.1007/s00438-012-0717-5
- Jiang, Y. F., Chen, Q., Wang, Y., Guo, Z. R., Xu, B. J., Zhu, J., et al. (2019). Re-acquisition of the brittle rachis trait via a transposon insertion in domestication gene *Q* during wheat de-domestication. *New Phytol.* 224, 961–973. doi: 10.1111/nph.15977
- Jiang, Q., Hou, J., Hao, C., Wang, L., Ge, H., Dong, Y., et al. (2011). The wheat (*T. aestivum*) sucrose synthase 2 gene (*TaSus2*) active in endosperm development is associated with yield traits. *Funct. Integr. Genomics* 11, 49–61. doi: 10.1007/s10142-010-0188-x
- Jiang, Y., Jiang, Q., Hao, C., Hou, J., Wang, L., Zhang, H., et al. (2015). A yield-associated gene *TaCWI*, in wheat: its function, selection and evolution in global breeding revealed by haplotype analysis. *Theor. Appl. Genet.* 128, 131–143. doi: 10.1007/s00122-014-2417-5
- Kuzay, S., Lin, H., Li, C., Chen, S., Woods, D. P., Zhang, J., et al. (2022). *WAPO-A1* is the causal gene of the 7AL QTL for spikelet number per spike in wheat. *PLoS Genet.* 18, e1009747. doi: 10.1371/journal.pgen.1009747
- Kuzay, S., Xu, Y., Zhang, J., Katz, A., Pearce, S., Su, Z., et al. (2019). Identification of a candidate gene for a QTL for spikelet number per spike on wheat chromosome arm 7AL by high-resolution genetic mapping. *Theor. Appl. Genet.* 132, 2689–2705. doi: 10.1007/s00122-019-03382-5
- Li, Z., Li, B., and Tong, Y. (2008). The contribution of distant hybridization with decaploid *Agropyron elongatum* to wheat improvement in China. *J. Genet. Genomics* 35, 451–456. doi: 10.1016/s1673-8527(08)60062-4
- Li, F., Wen, W., He, Z., Liu, J., Jin, H., Cao, S., et al. (2018). Genome-wide linkage mapping of yield-related traits in three Chinese bread wheat populations using high-density SNP markers. *Theor. Appl. Genet.* 131, 1903–1924. doi: 10.1007/s00122-018-3122-6
- Li, F., Wen, W., Liu, J., Zhang, Y., Cao, S., He, Z., et al. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* 19, 168. doi: 10.1186/s12870-019-1781-3
- Liu, H., Zhang, X., Xu, Y., Ma, F., Zhang, J., Cao, Y., et al. (2020). Identification and validation of quantitative trait loci for kernel traits in common wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 20, 529. doi: 10.1186/s12870-020-02661-4
- Lu, Y., Xing, L., Xing, S., Hu, P., Cui, C., Zhang, M., et al. (2015). Characterization of a putative new semi-dominant reduced height gene, *Rht_NM9*, in wheat (*Triticum aestivum* L.). *J. Genet. Genomics* 42, 685–698. doi: 10.1016/j.jgg.2015.08.007
- Ma, L., Li, T., Hao, C., Wang, Y., Chen, X., and Zhang, X. (2016). *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. *Plant Biotechnol. J.* 14, 1269–1280. doi: 10.1111/pbi.12492
- Ma, F., Xu, Y., Ma, Z., Li, L., and An, D. (2018). Genome-wide association and validation of key loci for yield-related traits in wheat founder parent xiaoyan 6. *Mol. Breed.* 38, 91. doi: 10.1007/s11032-018-0837-7
- Ma, Z., Zhao, D., Zhang, C., Zhang, Z., Xue, S., Lin, F., et al. (2007). Molecular genetic analysis of five spike-related traits in wheat using RIL and immortalized F₂ populations. *Mol. Genet. Genomics* 277, 31–42. doi: 10.1007/s00438-006-0166-0
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., et al. (2020). High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol. Plant* 13, 1311–1327. doi: 10.1016/j.molp.2020.07.008
- Peng, Z. S., Li, X., Yang, Z. J., and Liao, M. L. (2011). A new reduced height gene found in the tetraploid semi-dwarf wheat landrace aiganfanmai. *Genet. Mol. Res.* 10, 2349–2357. doi: 10.4238/2011.October.5.5
- Quarrie, S. A., Steed, A., Calestani, C., Semikhodskii, A., Lebreton, C., Chinoy, C., et al. (2005). A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese spring × SQ1 and its use to compare QTLs for grain yield across a range of environments. *Theor. Appl. Genet.* 110, 865–880. doi: 10.1007/s00122-004-1902-7
- Ramirez-Gonzalez, R. H., Borrill, P., Lang, D., Harrington, S. A., Brinton, J., Venturini, L., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* 361, eaar6089. doi: 10.1126/science.aar6089
- Ren, T., Hu, Y., Tang, Y., Li, C., Yan, B., Ren, Z., et al. (2018). Utilization of a wheat 55K SNP array for mapping of major QTL for temporal expression of the tiller number. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00333
- Sakuma, S., Golan, G., Guo, Z., Ogawa, T., Tagiri, A., Sugimoto, K., et al. (2019). Unleashing floret fertility in wheat through the mutation of a homeobox gene. *Proc. Natl. Acad. Sci. U. S. A.* 116, 5182–5187. doi: 10.1073/pnas.1815465116
- Simons, K. J., Fellers, J. P., Trick, H. N., Zhang, Z., Tai, Y. S., Gill, B. S., et al. (2006). Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172, 547–555. doi: 10.1007/s00122-002-1044-8
- Sourdille, P., Cadalen, T., Guyomarc'h, H., Snape, J. W., Perretant, M. R., Charmet, G., et al. (2003). An update of the Courtot × Chinese Spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat. *Theor. Appl. Genet.* 106, 530–538. doi: 10.1007/s00122-002-1044-8
- Sun, C., Dong, Z., Zhao, L., Ren, Y., Zhang, N., and Chen, F. (2020). The wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. *Plant Biotechnol. J.* 18, 1354–1360. doi: 10.1111/pbi.13361
- Tadesse, W., Ogbonnaya, F. C., Jighly, A., Sanchez-Garcia, M., Sohail, Q., Rajaram, S., et al. (2015). Genome-wide association mapping of yield and grain quality traits in winter wheat genotypes. *PLoS One* 10, e0141339. doi: 10.1371/journal.pone.0141339
- Voss-Fels, K. P., Keeble-Gagnere, G., Hickey, L. T., Tibbits, J., Nagorny, S., Hayden, M. J., et al. (2019). High-resolution mapping of rachis nodes per rachis, a critical determinant of grain yield components in wheat. *Theor. Appl. Genet.* 132, 2707–2719. doi: 10.1007/s00122-019-03383-4
- Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183
- Winfield, M. O., Allen, A. M., Burrridge, A. J., Barker, G. L., Benbow, H. R., Wilkinson, P. A., et al. (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1206. doi: 10.1111/pbi.12485
- Wu, Q. H., Chen, Y. X., Zhou, S. H., Fu, L., Chen, J. J., Xiao, Y., et al. (2015). High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 × Beinnong6. *PLoS One* 10, e0118144. doi: 10.1371/journal.pone.0118144

- Wu, X., Cheng, R., Xue, S., Kong, Z., Wan, H., Li, G., et al. (2013). Precise mapping of a quantitative trait locus interval for spike length and grain weight in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* 33, 129–138. doi: 10.1007/s11032-013-9939-4
- Xiong, H., Zhou, C., Fu, M., Guo, H., Xie, Y., Zhao, L., et al. (2022). Cloning and functional characterization of *Rht8*, a "Green revolution" replacement gene in wheat. *Mol. Plant* 15, 373–376. doi: 10.1016/j.molp.2022.01.014
- Xu, Y., Wang, R., Tong, Y., Zhao, H., Xie, Q., Liu, D., et al. (2014). Mapping QTLs for yield and nitrogen-related traits in wheat: influence of nitrogen and phosphorus fertilization on QTL expression. *Theor. Appl. Genet.* 127, 59–72. doi: 10.1007/s00122-013-2201-y
- Xu, D., Wen, W., Fu, L., Li, F., Li, J., Xie, L., et al. (2019). Genetic dissection of a major QTL for kernel weight spanning the *Rht-B1* locus in bread wheat. *Theor. Appl. Genet.* 132, 3191–3200. doi: 10.1007/s00122-019-03418-w
- Yan, X., Zhao, L., Ren, Y., Dong, Z., Cui, D., and Chen, F. (2019). Genome-wide association study revealed that the *TaGW8* gene was associated with kernel size in Chinese bread wheat. *Sci. Rep.* 9, 2702. doi: 10.1038/s41598-019-38570-2
- Yang, J., Zhou, Y., Wu, Q., Chen, Y., Zhang, P., Zhang, Y., et al. (2019). Molecular characterization of a novel *TaGL3-5A* allele and its association with grain length in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 132, 1799–1814. doi: 10.1007/s00122-019-03316-1
- Zanke, C. D., Ling, J., Plieske, J., Kollers, S., Ebmeyer, E., Korzun, V., et al. (2015). Analysis of main effect QTL for thousand grain weight in European winter wheat (*Triticum aestivum* L.) by genome-wide association mapping. *Front. Plant Sci.* 6, doi: 10.3389/fpls.2015.00644
- Zhai, H., Feng, Z., Li, J., Liu, X., Xiao, S., Ni, Z., et al. (2016). QTL analysis of spike morphological traits and plant height in winter wheat (*Triticum aestivum* L.) using a high-density SNP and SSR-based linkage map. *Front. Plant Sci.* 7, 1617. doi: 10.3389/fpls.2016.01617
- Zhang, N., Fan, X., Cui, F., Zhao, C., Zhang, W., Zhao, X., et al. (2017a). Characterization of the temporal and spatial expression of wheat (*Triticum aestivum* L.) plant height at the QTL level and their influence on yield-related traits. *Theor. Appl. Genet.* 130, 1235–1252. doi: 10.1007/s00122-017-2884-6
- Zhang, J., Gizaw, S. A., Bossolini, E., Hegarty, J., Howell, T., Carter, A. H., et al. (2018a). Identification and validation of QTL for grain yield and plant water status under contrasting water treatments in fall-sown spring wheats. *Theor. Appl. Genet.* 131, 1741–1759. doi: 10.1007/s00122-018-3111-9
- Zhang, P., He, Z., Tian, X., Gao, F., Xu, D., Liu, J., et al. (2017b). Cloning of *TaTPP-6AL1* associated with grain weight in bread wheat and development of functional marker. *Mol. Breed.* 37, 78. doi: 10.1007/s11032-017-0676-y
- Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., et al. (2018b). Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. *Plant J.* 94, 857–866. doi: 10.1111/tpj.13903
- Zhang, Y., Liu, J., Xia, X., and He, Z. (2014). *TaGS-D1*, an ortholog of rice *OsGS3*, is associated with grain weight and grain length in common wheat. *Mol. Breed.* 34, 1097–1107. doi: 10.1007/s11032-014-0102-7
- Zhang, Y., Liu, H., and Yan, G. (2021). Characterization of near-isogenic lines confirmed QTL and revealed candidate genes for plant height and yield-related traits in common wheat. *Mol. Breed.* 41, 4. doi: 10.1007/s11032-020-01196-8
- Zhang, L., Zhao, Y. L., Gao, L. F., Zhao, G. Y., Zhou, R. H., Zhang, B. S., et al. (2012). *TaCKX6-D1*, the ortholog of rice *OsCKX2*, is associated with grain weight in hexaploid wheat. *New Phytol.* 195, 574–584. doi: 10.1111/j.1469-8137.2012.04194.x
- Zhao, B., Wu, T. T., Ma, S. S., Jiang, D. J., Bie, X. M., Sui, N., et al. (2020). *TaD27-b* gene controls the tiller number in hexaploid wheat. *Plant Biotechnol. J.* 18, 513–525. doi: 10.1111/pbi.13220
- Zheng, J., Liu, H., Wang, Y., Wang, L., Chang, X., Jing, R., et al. (2014). *TEF-7A*, a transcript elongation factor gene, influences yield-related traits in bread wheat (*Triticum aestivum* L.). *J. Exp. Bot.* 65, 5351–5365. doi: 10.1093/jxb/eru306
- Zhou, Y., Conway, B., Miller, D., Marshall, D., Cooper, A., Murphy, P., et al. (2017). Quantitative trait loci mapping for spike characteristics in hexaploid wheat. *Plant Genome* 10, 2. doi: 10.3835/plantgenome2016.10.0101
- Zhuang, Q. (2003). *Chinese Wheat improvement and pedigree analysis (in Chinese)* (Beijing: Chinese Agricultural Press).
- Zhuang, M., Li, C., Wang, J., Mao, X., Li, L., Yin, J., et al. (2021). The wheat SHORT ROOT LENGTH 1 gene *TaSRL1* controls root length in an auxin-dependent pathway. *J. Exp. Bot.* 72, 6977–6989. doi: 10.1093/jxb/erab357



OPEN ACCESS

EDITED BY

Hairul Roslan,
Universiti Malaysia Sarawak, Malaysia

REVIEWED BY

Tetsuya Yamada,
Hokkaido University, Japan
Pankaj Kumar Bhowmik,
National Research Council Canada (NRC),
Canada
Raj Kumar Joshi,
Rama Devi Women's University, India

*CORRESPONDENCE

Bingyu Zhao
✉ bzhao07@vt.edu
Bo Zhang
✉ zhang76@vt.edu

[†]These authors have contributed equally to this work

RECEIVED 29 November 2022

ACCEPTED 31 March 2023

PUBLISHED 08 May 2023

CITATION

Wang Z, Shea Z, Rosso L, Shang C, Li J, Bewick P, Li Q, Zhao B and Zhang B (2023) Development of new mutant alleles and markers for *KTI1* and *KTI3* via CRISPR/Cas9-mediated mutagenesis to reduce trypsin inhibitor content and activity in soybean seeds.
Front. Plant Sci. 14:1111680.
doi: 10.3389/fpls.2023.1111680

COPYRIGHT

© 2023 Wang, Shea, Rosso, Shang, Li, Bewick, Li, Zhao and Zhang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of new mutant alleles and markers for *KTI1* and *KTI3* via CRISPR/Cas9-mediated mutagenesis to reduce trypsin inhibitor content and activity in soybean seeds

Zhibo Wang^{1†}, Zachary Shea^{1†}, Luciana Rosso ¹, Chao Shang¹, Jianyong Li², Patrick Bewick¹, Qi Li¹, Bingyu Zhao ^{1*} and Bo Zhang ^{1*}

¹School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States,

²Department of Biochemistry, Virginia Tech, Blacksburg, VA, United States

The digestibility of soybean meal can be severely impacted by trypsin inhibitor (TI), one of the most abundant anti-nutritional factors present in soybean seeds. TI can restrain the function of trypsin, a critical enzyme that breaks down proteins in the digestive tract. Soybean accessions with low TI content have been identified. However, it is challenging to breed the low TI trait into elite cultivars due to a lack of molecular markers associated with low TI traits. We identified Kunitz trypsin inhibitor 1 (*KTI1*, Gm01g095000) and *KTI3* (Gm08g341500) as two seed-specific TI genes. Mutant *kti1* and *kti3* alleles carrying small deletions or insertions within the gene open reading frames were created in the soybean cultivar *Glycine max* cv. *Williams 82* (*WM82*) using the CRISPR/Cas9-mediated genome editing approach. The KTI content and TI activity both remarkably reduced in *kti1/3* mutants compared to the *WM82* seeds. There was no significant difference in terms of plant growth or maturity days of *kti1/3* transgenic and *WM82* plants in greenhouse condition. We further identified a T1 line, #5-26, that carried double homozygous *kti1/3* mutant alleles, but not the Cas9 transgene. Based on the sequences of *kti1/3* mutant alleles in #5-26, we developed markers to co-select for these mutant alleles by using a gel-electrophoresis-free method. The *kti1/3* mutant soybean line and associated selection markers will assist in accelerating the introduction of low TI trait into elite soybean cultivars in the future.

KEYWORDS

soybean, anti-nutritional factor, Kunitz trypsin inhibitor (KTI), *KTI1*, *KTI3*, CRISPR/Cas9, allele-based selection marker

Introduction

Soybean meal provides an excellent source of protein in animal feed since it is rich in amino acids with a high nutritional profile (Cromwell et al., 1991). For instance, soy makes up 26% and 50% of swine and poultry feed, respectively (Gillman et al., 2015). However, reduced feed efficiency has been observed due to anti-nutritional and biologically active factors in raw soybean seeds (Liener, 1996). Among these factors, trypsin inhibitor (TI) accounts for a substantial amount of this effect that cannot be ignored (Hymowitz, 1986). TI restrains the activity of trypsin in monogastric animals. Because this enzyme is essential for optimal protein digestion, its restriction can lead to animal growth inhibition of 30–50% due to pancreatic hypertrophy/hyperplasia when raw soybeans are used in feed (Hymowitz, 1986; Cook et al., 1988; Liener, 1994). In soybean meal processing facilities, TI in soybean meal is deactivated *via* a heating process at 90.5°C–100°C with the presence of 1% NaOH (Chen et al., 2014). This process not only reduces the nutritional value of soybean meal due to thermal destruction of amino acids, but also increases the energy cost of meal production by 25% (Chang et al., 1987).

With the recent increase in feed price and shipping cost, livestock farmers and grain operations are reconsidering soybean varieties with low-TI or TI-free traits as a way to reduce farm expenses by using raw soybeans as feed. Raising low-TI or TI-free soybeans on farms creates a niche market for integrated crop and livestock farmers, increasing their farm's profitability. The use of soybean lines with genetically reduced levels of TI has proven as an effective strategy for improving animal growth. For instance, chicks fed soy-based diets with raw, unprocessed, low-KTI soymeal had higher feed efficiency ratios than chicks fed diets containing raw, unprocessed, conventional soybean meal (Batal and Parsons, 2004). Thus, soybean cultivars with low TI content in the seeds is a long-term breeding goal for higher protein digestibility, better economic benefits, reduced environmental pollution caused by phosphorus, and the pursuit of sustainability for humanity and nature.

Plants have evolved a group of TI genes encoding proteins that can suppress the enzyme activities of proteases found in plants, herbivores, animals and human beings (Jofuku and Goldberg, 1989; Schuler et al., 1999). The TIs in soybean can be classified into two families: the 21 kDa Kunitz trypsin inhibitor protein family (KTI) and the 7–8 kDa Bowman-Birk inhibitor protein family (BBTI) (Kunitz, 1945; Wei, 1983; Gillman et al., 2015). KTI proteins are thought to be largely specific for trypsin inhibition, while the major isoform of BBTI contains domains that interact with and inhibit both trypsin and chymotrypsin (Hwang et al., 1977; Gillman et al., 2015). Currently, only the KTI genes are targeted for selection of low-TI soybeans because KTI serves as the major contributor to trypsin inhibitor activity in soybeans. By far, the most significant success in reducing TI activity in soybean was the identification of a soybean accession (PI 157740) with dramatically reduced (~40%) TI activity (Gillman et al., 2015). A frameshift mutation that results in premature termination during translation of *KTI3* (Gm08g341500) was identified in PI 157740, which is responsible for the low TI phenotype (Jofuku et al., 1989). This frameshift

mutation is caused by the alteration of three nucleotides located at positions +481, +482, and +486. PI 157740 has been used in feeding trials, and it was found that raw extruded protein meal with lower *KTI3* protein is superior for animal weight gain when compared to raw soybean meal harboring functional *KTI3* (Cook et al., 1988; Perez-Maldonado et al., 2003). However, weight gain for young animals fed with non-heat-treated soybean materials including nonfunctional *KTI3* soybean materials is still inferior to those fed with heat-treated soybeans (Cook et al., 1988; Perez-Maldonado et al., 2003). Another soybean germplasm accession (PI 68679) was identified to carry a nonfunctional mutation on *KTI1* (Gm01g095000) gene, where a single base deletion which introduces (GGG356 → 358GG, relative to start codon) a frameshift mutation (Gillman et al., 2015). *KTI1* and *KTI3* genes were determined to be synergistically controlling the TI content in soybean seeds (Gillman et al., 2015). Therefore, it is desirable to breed new soybean cultivars carrying both *kti1* and *kti3* mutant alleles. However, it is time-consuming to breed low TI soybean cultivars by selecting progenies derived from crosses between PI 157740, PI 68679, and elite varieties. Besides, the linkage drags associated with *KTI1* and *KTI3* may introduce undesirable agronomic traits, which could be difficult to remove by backcrossing. Although Kompetitive Allele Specific PCR (KASP) markers associated with *KTI3* and its mutant allele with 86% efficiency are available (Rosso et al., 2021), attempts at developing molecular markers associated with *KTI1* have not been successful (Gillman et al., 2015). Therefore, it is highly desirable to develop new *KTI1* mutant alleles that can be tagged with convenient molecular markers.

CRISPR/Cas9 mediated genome editing employs a Cas9 endonuclease and an 18–22 bp small guide RNA (sgRNA) that have a region that is complementary to a target gene sequence. The sgRNA binds to Cas9 and recruits the complex to target a gene. The Cas9 endonuclease generates DNA breaks, leading to mis-repaired target genes that contain deletions or insertions that disrupt gene function. In addition, several sgRNAs can be co-expressed in a single cell with Cas9, which allows the multiplex mutations of different genes simultaneously (Liang et al., 2016). Because genome edited plants without transgenes are not considered as genetically modified organisms (GMO) (Kim and Kim, 2016), mutant plants can be either directly released for field test or served as valuable resources for further breeding selection. Thus far, the CRISPR/Cas9 mediated genome editing technology has been widely used for targeted gene mutagenesis in diverse crop plant species, including soybean (Haun et al., 2014; Cai et al., 2015; Jacobs et al., 2015), rice (Xu et al., 2015), wheat (Upadhyay et al., 2013), maize (Chen et al., 2018), tomato (Vu et al., 2020), cotton (Gao et al., 2017), citrus (Peng et al., 2017), apple (Osakabe et al., 2018), grape (Osakabe et al., 2018), potato (Nakayasu et al., 2018), and banana (Shao et al., 2020) to improve their agronomic performances. For example, Jacobs et al. (2015) reported the first targeted mutagenesis in soybean using the CRISPR/Cas9 technology (Jacobs et al., 2015). Haun et al. (2014) generated a high oleic acid content soybean variety without transgenic components and improved the quality of soybean (Haun et al., 2014). A soybean mutant with a late flowering

In this study, we aimed to (1) simultaneously knockout *KTI1* and *KTI3* genes in soybean cultivar Williams 82 via CRISPR/Cas9-mediated genome editing and (2) develop molecular markers associated with *kti1* and *kti3* mutant alleles that can be used for marker-assisted selection (MAS). We successfully recovered transgenic soybean plants that are carrying both *kti1* and *kti3* mutations. KTI content and trypsin inhibition activities (TIA) are dramatically decreased in the *kti1* and *kti3* mutant lines. In addition, we also developed molecular markers for co-selection of the new *kti1* and *kti3* mutant alleles. These *kti1* and *kti3* mutant lines and the newly developed selection markers have great potential for breeding the low TI trait into elite soybean varieties in the future.

KTI gene family consists of multiple members with distinct expression patterns

To identify the *KTI* genes expressed in the seed, we analyzed the expression patterns of all *KTI* genes in cv. WM82 based on the expression data acquired through the Gene Networks in Seed Development database (<http://seedgenenetwork.net/sequence>) (Figure 1A). According to the expression patterns of *KTI* genes in

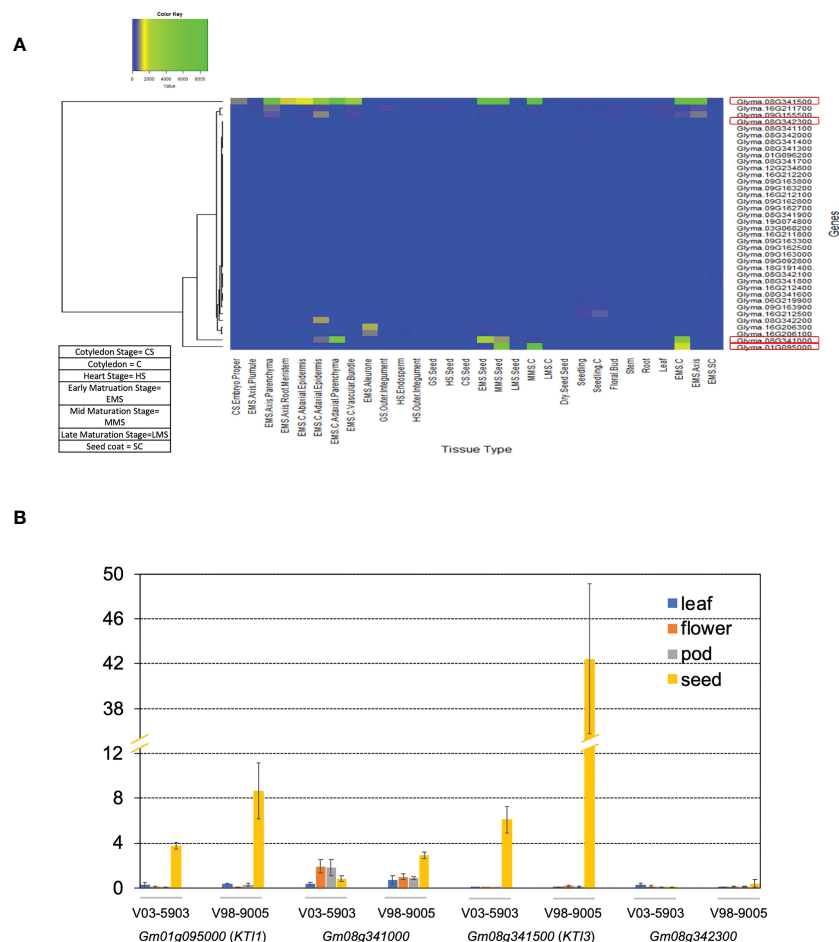


FIGURE 1
Expression levels of *KTI* genes in *WM82*. **(A)** RNA sequencing data of 38 *KTI* genes in 30 different tissue types of cv. *Williams 82* acquired from Phytozome soybean database was used to construct the heatmap to visualize their expression patterns. **(B)** The expressions of four soybean *KTI* genes were monitored by real-time PCR. Samples of leaf, flower, pod, and seed tissues from 2 breeding lines, V98-9005 (normal-TI line) and V03-5903 (low-TI line), were collected for RNA extraction. After reverse transcription, real-time PCR was used to evaluate the expressions of 4 genes including *Gm01g095000*, *Gm08g341000*, *Gm08g342300*, and *Gm08g341500* in different tissues with the *ELF1B* (*Gm02g276600*) as the reference gene. The expression data was normalized as ΔCT and shown as mean \pm s.e. Experiments were repeated three times and obtained similar results.

various soybean tissues as displayed in [Figure 1A](#), four *KTI* genes (Gm01g095000 (*KTI1*), Gm08g341000, Gm08g342300, and Gm08g341500 (*KTI3*)) were identified as seed-specific *KTI* genes. Soybean breeding lines V98-9005 (normal TI) and V03-5903 (low TI), presenting significantly different amounts of KTI concentration in seeds, were used to validate the tissue specific expressions of the four *KTI* genes by real-time PCR. Gm01g095000 (*KTI1*) and Gm08g341500 (*KTI3*) were predominately expressed in seeds compared to other tissues ([Figure 1B](#)). Both Gm08g341000 and Gm08g342300 had a relatively lower expression level in seeds than *KTI1* and *KTI3* but had higher expression in other tissue types ([Figure 1B](#)). Interestingly, both *KTI1* and *KTI3* had a relatively low expression level in the seeds of V03-5903 (low TI line), but higher expression in V98-9005 (normal TI line). Thus, we conclude that *KTI1* and *KTI3* are two major genes that may directly contribute to the TI contents in soybean seeds.

Development of CRISPR/Cas9-based binary vector for genome-editing in soybean

To knock out the *KTI1* and *KTI3* genes from cv. WM82 genome and create a new soybean cultivar with low TI content in soybean seeds, we developed a CRISPR/Cas9 construct, pBAR-Cas9-*kti13*, where the nuclease gene *Cas9* is expressed by Arabidopsis ubiquitin 10 (U10) promoter. A *bar* gene driven by a MAS (mannopine synthase) promoter was used for selection of the putative transformants with bialaphos or phosphinothricin ([Figure 2A](#)). A tandem array of two sgRNAs targeting *KTI1* and one sgRNA targeting *KTI3* was expressed by the U6 RNA promoter ([Figure 2B](#)).

KTI1 and KTI3 genes are knocked out by CRISPR/Cas9 mediated gene editing

pBAR-Cas9-*kti1/kti3* was transformed into WM82 via *Agrobacterium*-mediated transformation (Plant Transformation Facility at Iowa State University). Seventeen putative transgenic shoots were regenerated. Six shoots elongated and were transferred to rooting mediums. After further selection, they were transplanted into soil. Four lines, No. #2, #5, #11 and #17 were confirmed to be true transformants by positive amplification of the *bar* gene and a part of the *Cas9* gene ([Figures 2C, D](#)). The gene editing events in T0 plants were identified by amplification and sequencing of DNA fragments covering the sgRNA binding sites of *KTI1* and *KTI3*. The double peaks in the sequencing chromatograms suggest that both *KTI1* and *KTI3* genes were mutated and resulted in heterozygous alleles in the edited plant cells ([Figure 2E](#)). T0 seeds were harvested from T0 lines #2, #5, #11 and #17. Four T0 seeds of each line were randomly picked for DNA extraction and genotyping of the *KTI1* and *KTI3* genes via PCR amplification and DNA sequencing. The *KTI1* gene editing was completed and resulted in homozygous mutant alleles in all tested T0 seeds of the four lines. In addition, an identical gene editing pattern in *KTI1* was detected in all tested T0 seeds, in which a small DNA fragment (66bp) between two sgRNAs was lost after the gene editing ([Figures 3A–C](#)). Homozygous *KTI3*

mutant alleles were only detected in T0 seeds from 2-3, #5-4, #11-2, and #11-4 ([Figures 3D–G](#)). The gene editing patterns in *KTI3* included both small deletions and insertions that all resulted in frameshift mutations in *KTI3*.

TI content and activity dramatically declined in the edited soybean seeds

T0 seeds were also used for quantification of the KTI content by using a HPLC-based approach ([Rosso et al., 2018](#)). The tested seeds of #2-3, #5-4, #11-2, and #11-4, which carried mutations on both *KTI1* and *KTI3* genes had the lowest KTI content ([Figure 4](#)). The tested seeds of #2-1, #5-1, #11-1, and #17-1, with only the *KTI1* mutation, also had lower KTI content than the wild-type WM82 seeds ([Figure 4](#)). The KTI content in other genotyped seeds with editing only on *KTI1* was also lower than that in WM82 seeds (data not shown). We further tested the trypsin inhibition activity (TIA) using crude protein extracts from the T0 seeds. As shown in [Figure 5](#), the crude proteins of seeds with mutant *kti1* and *kti3* had the lowest TIA. The seeds with mutant *kti1* only also had reduced TIA ([Figure 5](#)) in comparison with WM82 and Glenn (a commercial soybean cultivar as a control). The KTI content and TIA were ranked in order as: *kti1/3* double mutant < *kti1* single mutant ≤ PI 547656 (low TI accession) < WM82 < Glenn. Taken together, we conclude that *KTI1* and *KTI3* are two major genes responsible for the KTI content and TIA in soybean seeds. Therefore, knockout of *KTI1* and *KTI3* reduced the KTI content and impaired the TIA in soybean seeds.

The edited *KTI1* gene lost 66 bp that may result in mutant proteins with deletion of 22 amino acids. Truncated *KTI1* may still possess some TIA. To rule out this possibility, we also tested the TIA of truncated *KTI1*_{Δ22aa} protein *in vitro*. To this end, we cloned the open reading frames of *KTI1*_{Δ66bp} and wild-type *KTI1* and *KTI3* into a protein expression vector, in which a 6xHis tag is fused to C-terminus of the expressed proteins. The purified proteins were subjected to a TIA assay which showed that while *KTI1* and *KTI3* both could inhibit trypsin activity, the truncated *KTI1*_{Δ22aa} failed to suppress trypsin activity ([Figures S2A, B](#)). Therefore, the new *kti1* allele (*KTI1*_{Δ66bp}) encodes a truncated protein that loses its TI function.

Knockout KTI1 and KTI3 did not affect plant growth and maturity period days of soybean

To examine whether the mutant *kti1/3* could significantly affect plant growth and maturity period days of soybean, we planted T0 seeds of line #2 and #5, and WM82 in a greenhouse. By bialaphos-mediated screening, we classified the T1 plants from line #2 and #5 as transgene-free plants or transgenic plants. We measured the agronomic traits of the transgenic plants including plant height, the number of main branches per plant, number of pods bearing branches, number of pods, leaf length, leaf width and petiole length. There was no significant difference in terms of all measured agronomic traits among the plants of WM82, Line 2 and Line 5 ([Table 1](#)). We also measured the maturity period days of

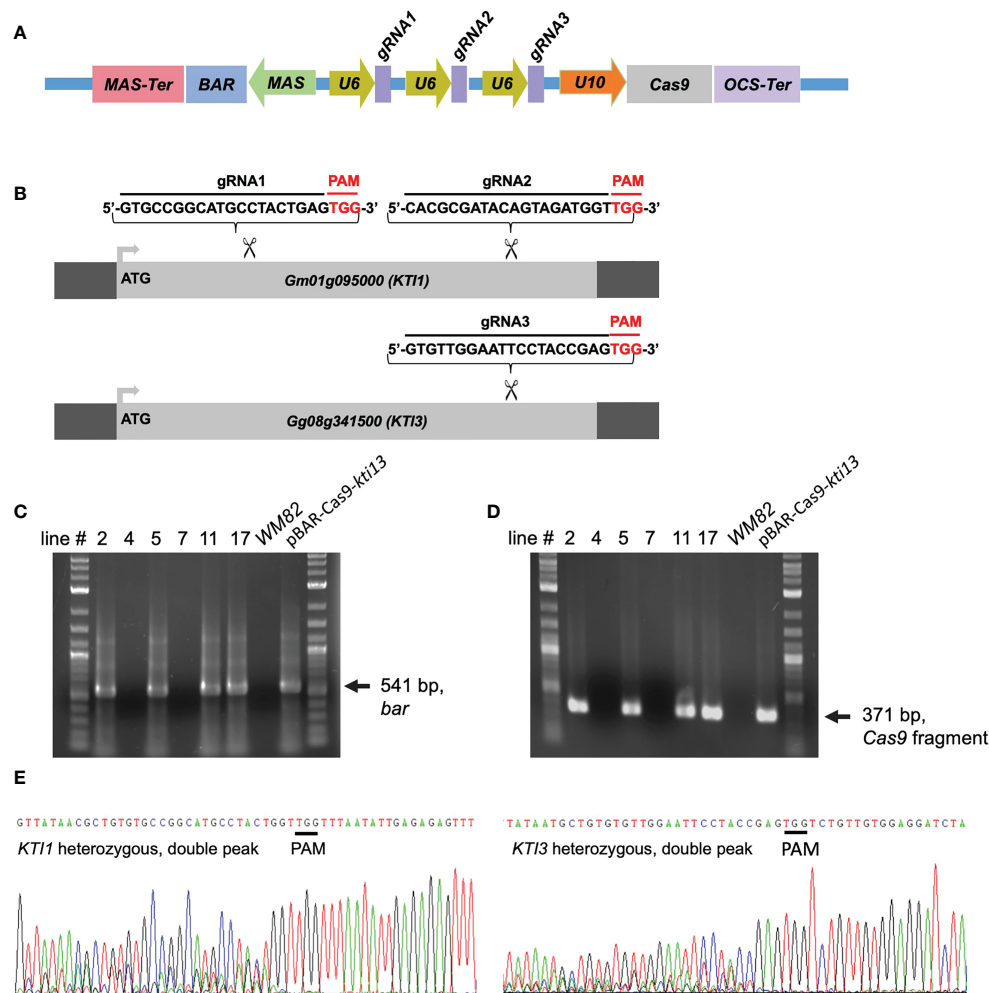


FIGURE 2

The scheme of binary vector used for CRISPR/Cas9 mediated gene editing on *KTI1/KTI3*, and transgenes and gene editing have been detected in the leaves of four T0 soybean plants. (A) The CRISPR/Cas9 construct harbors three necessary elements exhibited as below: the selection cassette consists of MAS promoter, *Bar* gene (soybean transformation selection marker), and MAS terminator; the Cas9 cassette consists of U10 promoter, Cas9 gene, and OCS terminator; three guide RNA cassettes and each of them consists of a U6 promoter, and one sgRNA. (B) The sequences of three sgRNAs is shown here. Two sgRNAs were designed, synthesized, and assembled to the plasmid to target on *KTI1*, while one sgRNA was designed, synthesized, assembled to the plasmid to target on *KTI3*. The fragments of two transgenes, (C) *Cas9* and (D) *Bar*, have both been detected in lines #2, #5, #11, and #17 by PCR, but not lines #4 and #7. The WM82 gDNA serves as the template for negative control, while the plasmid DNA serves as the template for positive control. (E) The gene editing on *KTI1* and *KTI3* has also been observed in the leaf tissues of plants at T0 generation. The double peak sequence around the sgRNA region indicates the gene editing was ongoing but not completed.

the soybean plants by recording the dates from planting to beginning bloom (R1), to beginning pod (R3), to beginning seed (R5), to full seed (R6), to maturity (R8), and the total lifespan (from planting to maturity). There were no remarkable differences in terms of R1, R3, R5, R6, R8 and total life span among all tested plants (Table 1). Therefore, we conclude that knockout of *KTI1* and *KTI3* did not alter plant growth or the maturity period of soybean lines tested.

Development of molecular markers for selection of the *kti1* and *kti3* alleles

A double homozygous *kti1* and *kti3* mutant plant #5-26 that did not carry the Cas9 transgene was selected from T1 generation plants

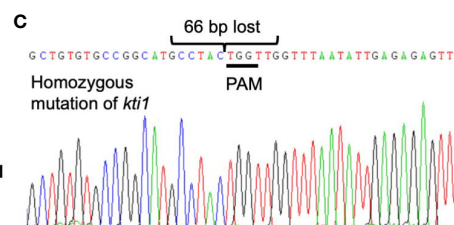
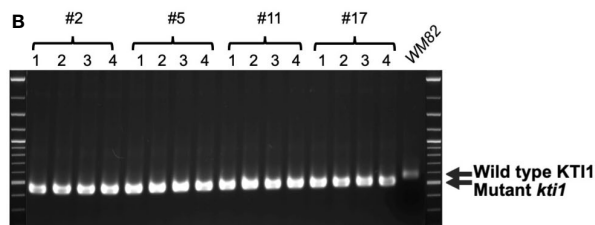
(Figures 3A, C, D). The 'transgene-free' soybean plants can be used to breed the low TI trait into other elite soybean cultivars. In order to co-select the *kti1* and *kti3* mutant alleles in the derived progenies, we attempted to develop co-dominate molecular markers that can distinguish between wild-type *KTI1/KTI3* and mutant *kti1/kti3* alleles.

In the genome of #5-26, the mutant allele of *kti1* had a 66 bp deletion. We designed three PCR primers, ZW1, ZW2 and ZW3 (Figure 6A and Table S2). ZW1 was a common reverse primer that can bind to the same region in both *KTI1* and *kti1* alleles, while ZW2 and ZW3 were both forward primers binding to unique sequences of *KTI1* and *kti1* alleles, respectively (Figure 6A). Soybean cultivars that carry the wild-type *KTI1* gene (WM82) amplified a 180 bp DNA fragment when hybridized with ZW1 and ZW2, but failed to amplify any fragments when hybridized with

A Alignment of *ktt1* mutants

```

WM82      GGCATGCCTACTGACTGGGCTATTGTGGAGAGAGGGTCTACAAGCTGTAAACTTGCTGCACGCGATACAGTAGATGGTTGGTTTAATATT
Line2-1    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line2-2    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line2-3    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line2-4    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line5-1    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line5-2    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line5-3    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line5-4    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line7-1    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line7-2    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line7-3    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line7-4    GGCATGCCTAC-----TGGTTGGTTTAATATT
Line11-1   GGCATGCCTAC-----TGGTTGGTTTAATATT
Line11-2   GGCATGCCTAC-----TGGTTGGTTTAATATT
Line11-3   GGCATGCCTAC-----TGGTTGGTTTAATATT
Line11-4   GGCATGCCTAC-----TGGTTGGTTTAATATT
  
```



D Alignment of *ktt3* mutants

```

WM82      TGCAGTTATAAT-----GCTGTGTGTTGGAATTCCTACCGAGTGGTCTGTTGTGGA
2-3       TGCAGTTATAATCCACACACAGCAGTTATAATCCACACACAGCATTATAATTGCAATGAATCGAACTTAATCCTTT-----GAGTGGTCTGTTGTGGA
5-4       TGCAGTTATAAT-----GCTGTGTGTTGGAATTCCTAC--GAGTGGTCTGTTGTGGA
11-2      TGCAGTTATAAT-----GCTGTGTGTTGGAAT-----GAGTGGTCTGTTGTGGA
11-4      TGCAGTTATAAT-----GCTGTGTGTTGGAAT-----GAGTGGTCTGTTGTGGA
  
```

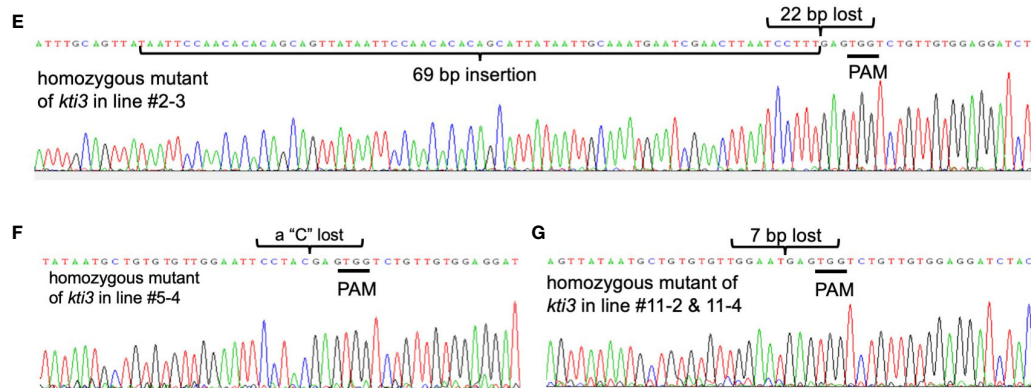


FIGURE 3

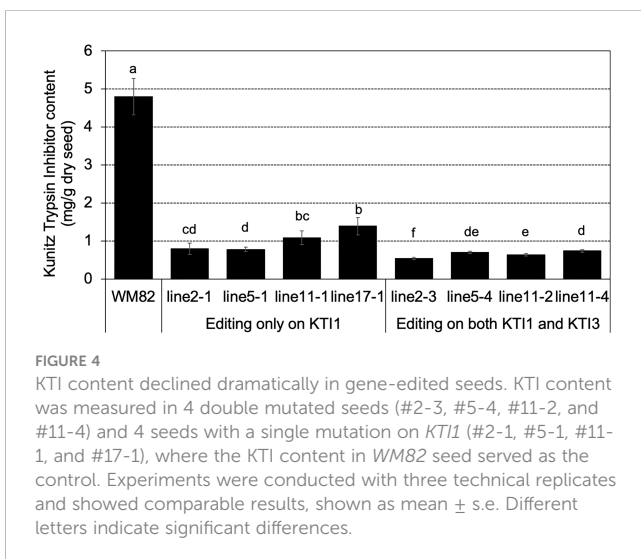
Gene editing on *KT11* has been completed for all seeds of T0 generation while it has been completed on *KT13* for some seeds of T0 generation. From each transgenic line (#2, #5, #11 and #17), four seeds of T0 generation were selected randomly for genotyping. (A) The gr of mutant *ktt1* in T0 seeds and T1 plant (#5-26) leaf, where the wild type *KT11* in WM82 was the control. (B) Gel electrophoresis of *ktt1* PCR products showed 16 seeds from line #2, #5, #11, and #17 had the same mutant on *ktt1*, in which 66 nucleotides are lost between two sgRNAs. (C) Sanger sequencing result displayed the identical mutant *ktt1*. (D) The alignment of mutant *ktt3* in T0 seeds (#2-3, #5-4, #5-26, #11-2 and #11-4) and T1 plant (#5-26) leaf, where the wild type *KT11* in WM82 was the control. (E–G) showed the sanger sequencing results of *ktt3* mutant in #2-3, #5-4, #11-2, #11-4, and #5-26.

ZW1 and ZW3. On the contrary, the soybean lines carrying a homozygous *ktt1* mutant allele (#5-26) amplified a 134 bp DNA fragment with ZW1 and ZW3, but failed to amplify any fragments with ZW1 and ZW2 (Figure 6B).

In the genome of line #5-26, the mutant *ktt3* allele had a 38 bp deletion, which allowed us to design PCR primers ZW4, ZW5 and ZW6 (Figure 6A and Table S2). ZW4 was a common reverse primer for both *KT13* and *ktt3*, while ZW5 and ZW6 were forward primers matched with unique sequences of *KT13* and *ktt3*, respectively (Figure 6A). A soybean cultivar carrying the wild-type *KT13* gene (MW82) amplified a 264 bp DNA fragment with primers ZW4 and ZW5, but not with primers ZW4 and ZW6. In contrast, the soybean line carrying homozygous *ktt3* allele (#5-26) can amplify a 233 bp

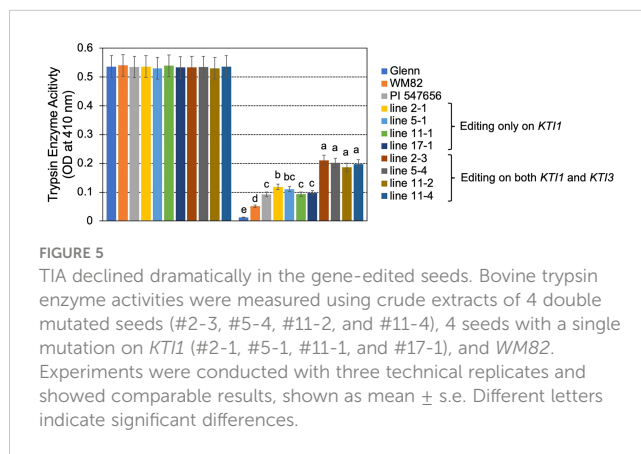
DNA fragment with primers ZW4 and ZW6, but not ZW4 and ZW5 (Figure 6B).

We further tested these PCR primers by amplifying DNA fragments from two T1 plants that were genotyped by DNA sequencing. Line #5-9 had homozygous *ktt1* alleles and heterozygous *KT13/ktt3* alleles, where the *ktt3* allele was identical to the one in #5-26 (Figure S3). Line #2-30 had homozygous *ktt3* alleles and heterozygous *KT11/ktt1* alleles, where the *ktt1* allele was identical to the one in #5-26 (Figure S3). As shown in Figure 6B, PCR amplification with the different combinations of ZW1, ZW2, ZW3, ZW4, ZW5 and ZW6 can accurately identify the *KT11/ktt1* and *KT13/ktt3* genotypes of #5-9 and #2-30 (Figure 6B). Therefore, we successfully developed molecular markers to select the *ktt1* and



kti3 mutant alleles generated by CRISPR/Cas9 mediated mutagenesis. These molecular markers can assist in the breeding selection of low TI soybean plants harboring *kti1/3*.

To simplify the procedure of marker-aided selection, we tested a gel-electrophoresis-free protocol that can be implemented for high throughput screening of progenies derived from a cross between a soybean cultivar carrying wild-type *KTI1/3* and one carrying the *kti1/3* mutant. In brief, all PCR products as described above were mixed with 1X SYBR Green and heated at 75°C for 10 mins and then visualized under UV light. As shown in Figure 6B, the fluorescent signals were the indications of positive amplifications in *WM82* with primers ZW1/ZW2 and ZW4/ZW5, while in #5-26, the fluorescent signals can only be observed with primers ZW1/ZW3 and ZW4/ZW6 (Hirotsu et al., 2010). Therefore, we identified



the homozygous *kti1* and *kti3* alleles by directly staining the PCR products without the need of gel electrophoresis, which can significantly reduce the cost of labor and time.

Discussion

In this study, we optimized a CRISPR/cas9-vector for genome editing in soybean (Figure 2A). The modified vector allowed us to simultaneously knock out two seed specific KTI genes (*KTI1* and *KTI3*). The *kti1/3* mutant plants grew normally in greenhouse conditions, and the seeds of *kti1/3* mutant had dramatically reduced KTI content and TI activities in comparison with wild type seed of *WM82*.

Soybean is one of the important sources of protein for animal and human consumption. However, in their evolution, soybeans have developed diverse defense components to protect seeds from

TABLE 1 Knockout of *KTI1* and *KTI3* does not alter the plant growth indicator and maturity period days of cv. *WM82*.

| Genotypes | Plant growth indicator | | | | | | | Maturity period days | | | | | |
|-----------------------|------------------------|-------------------------|----------------------|-----------------------|----------------|---------------|----------------|----------------------|----------------|----------------|----------------|----------------|-----------------|
| | Plant height | Main branches per plant | Pod bearing branches | No. of pods per plant | Leaf length | Leaf width | Petiole length | Plant to R1 | R1 to R3 | R3 to R5 | R5 to R6 | R6 to R8 | Planting to R8 |
| <i>WM82</i> | 29.5 \pm 2.2 | 2.2 \pm 0.4 | 16.0 \pm 1.6 | 30.0 \pm 2.9 | 11.2 \pm 1.1 | 6.8 \pm 0.8 | 12.4 \pm 1.5 | 43.8 \pm 3.1 | 23.6 \pm 2.1 | 18.6 \pm 2.1 | 30.8 \pm 2.6 | 19.6 \pm 2.7 | 136.4 \pm 4.7 |
| Line 2 transgenic | 29.2 \pm 1.2 | 2.4 \pm 0.5 | 16.4 \pm 1.8 | 29.4 \pm 3.8 | 11.6 \pm 1.8 | 7.0 \pm 1.0 | 11.4 \pm 1.9 | 40.8 \pm 3.3 | 22.4 \pm 2.7 | 20.2 \pm 3.3 | 32 \pm 2.5 | 21.4 \pm 2.3 | 136.8 \pm 6.1 |
| Line 2 non-transgenic | 29.8 \pm 1.2 | 2.5 \pm 0.5 | 16.6 \pm 1.7 | 30.7 \pm 3.2 | 11.3 \pm 1.5 | 7.1 \pm 1.0 | 11.9 \pm 2.0 | 42.1 \pm 3.8 | 23.0 \pm 3.0 | 19.4 \pm 3.1 | 31.4 \pm 2.2 | 20.1 \pm 2.3 | 136.0 \pm 5.5 |
| Line 5 transgenic | 29.1 \pm 1.3 | 2.2 \pm 0.4 | 17.2 \pm 1.6 | 28.6 \pm 3.4 | 11.4 \pm 1.3 | 7.2 \pm 0.8 | 11.2 \pm 2.4 | 41.0 \pm 4.1 | 20.6 \pm 2.9 | 20.8 \pm 2.8 | 31 \pm 4.2 | 20.0 \pm 2.1 | 133.4 \pm 8.9 |
| Line 5 non-transgenic | 29.0 \pm 1.5 | 2.4 \pm 0.3 | 16.9 \pm 1.8 | 29.1 \pm 3.5 | 11.0 \pm 1.4 | 6.9 \pm 0.8 | 11.6 \pm 1.8 | 41.7 \pm 4.4 | 22.5 \pm 2.8 | 20.5 \pm 2.6 | 30.7 \pm 3.2 | 19.9 \pm 2.4 | 135.3 \pm 6.7 |

R1: From planting to beginning bloom; R3: beginning bloom to beginning pod; R5: beginning pod to beginning seed; R6: beginning seed to full seed; R8: full seed to maturity. For each genotype, 5 plants were utilized for measuring plant growth indicators and maturity period days. An ANOVA test was employed here for statistical analysis.

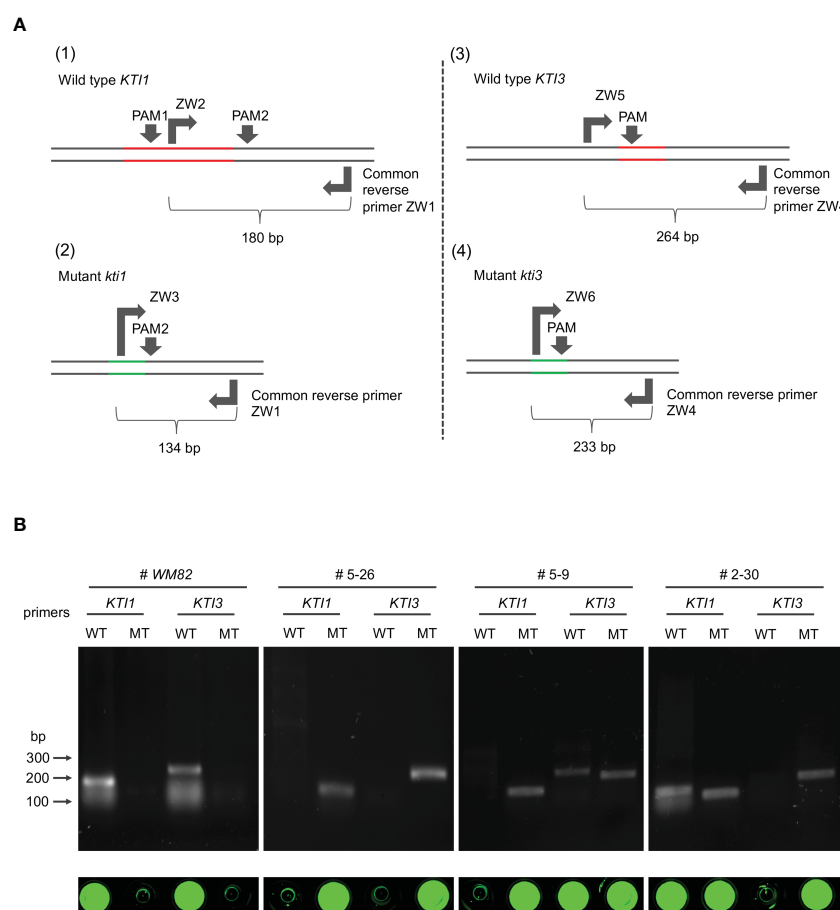


FIGURE 6

The development of selection markers for breeding low KTI soybean varieties based on the *kTi1* and *kTi3* mutants generated by CRISPR/Cas9-mediated gene editing. **(A)** Schematic development of primers for amplification of wild type *KTI1* (1), mutant *KTI1* (2), wild type *KTI3* (3), and mutant *KTI3* (4). The red lines indicate the lost fragment in *KTI1* or *KTI3* during gene editing. The green lines indicate new DNA regions in *kTi1* or *kTi3* generated by splicing two fragments. **(B)** The 4 pairs of primers in **(A)** were utilized to amplify the alleles of *KTI1*, *kTi1*, *KTI3*, and *kTi3* with gDNA of four different soybean genotypes, including WM82, three transgenic lines #5-26, #5-9, and #2-30. Based on our genotyping data, #5-26 has homozygous mutations of *kTi1* and *kTi3*; #5-9 only has a homozygous mutation of *kTi1* but carries the heterozygous mutation of *kTi3*; #2-30 only has a homozygous mutation of *kTi3* but carries the heterozygous mutation of *kTi1*. Thus, it was clear that the pair of ZW1/ZW2 can amplify wild type *KTI1* from WM82 and #2-30 gDNA in PCR tests, while the pair of ZW1/ZW3 can amplify mutant *kTi1* from #5-9 and #5-26 gDNA. Also, the pair of ZW4/ZW5 can amplify wild type *KTI3* from WM82 and #5-9 gDNA, while ZW4/ZW6 can amplify mutant *kTi3* from #2-30 and #5-26 gDNA. As shown in the bottom panel, only the positive PCR products incubated with the dye of sybrgreen at 75°C can display the fluorescent signals, suggesting the reliability of the developed gel-electrophoresis-free method for screening mutant alleles of *kTi1* and *kTi3*.

being eaten by insects and animals including trypsin inhibitor, phytate acid, and raffinose family of oligosaccharides (RFOs). In the agricultural practice, the anti-nutritional and biologically active factors are responsible for reduced feed efficiency when raw soybeans are fed to animals. Therefore, it is of great significance to increase feed efficiency, especially the protein digestibility via assembling gene function exploration, application, and advanced genetically engineering together into the soybean industry. Proteinaceous plant trypsin inhibitors are a diverse family of (poly)peptides that play diverse roles in plant growth such as maintaining physiological homeostasis and serving the innate defense machinery (Li et al., 2008; Junker et al., 2012; Arnaiz et al., 2018; Zhao et al., 2019). Since TI proteins exert direct effects on pests and herbivores by interfering with their physiology, any food containing TI proteins will be avoided by these organisms. In alfalfa, the trypsin inhibitors Msti-94 and Msti-

16 were demonstrated to act as a stomach poison, significantly reducing the survival and reproduction rates of aphid (Zhao et al., 2019). Mutant plants with reduced TI are usually more susceptible to pests. In wheat, α -amylase/trypsin inhibitors (ATIs) CM3 and 0.19 were identified as pest-resistance molecules, activating innate immune responses in monocytes, macrophages, and dendritic cells (Junker et al., 2012). For example, the *Arabidopsis* lines containing silenced *atkti4* and *atkti5* were found to have a higher susceptibility to *T. urticae* (Spider mite) than wild-type plants (Arnaiz et al., 2018). RNAi silencing of the *AtKTI01* gene resulted in enhanced lesion development after infiltration of leaf tissue with the programmed cell death eliciting fungal toxin fumonisin B1 or the avirulent bacterial pathogen *Pseudomonas syringae* pv. *tomato* DC3000 carrying *avrB* (Li et al., 2008). Although similar defense functions have not been reported on TI genes in the soybean genome, it is reasonably suspected that certain members of the

KTI gene family have comparable protecting roles for soybean plants.

As previously discussed, the high concentration of TI proteins in soybean meal restricts the function of trypsin, which causes low digestibility and reduces its nutritional value. Thus, cultivars have been developed by introgression of low TI traits into elite cultivars. We previously developed a low TI line *via* conventional breeding: V12-4590. During field trials in 2017 and 2018, we observed that this low TI line is indeed more susceptible to multiple phytopathogens such as: all races of Soybean Cyst Nematode (SCN) (*Heterodera glycines*), Stem Canker (*Diaporthe aspalathi*), Cercospora leaf blight (*Cercospora kukuchii*), Soybean vein necrosis virus, and Downy Mildew (*Peronospora manshurica*) (Zhang, unpublished data). This suggests that the soybean KTI genes that are negatively selected do indeed have a role in plant immunity. It is also possible that some plant immunity related genes that genetically link with KTI genes are negatively selected during the breeding process. As shown in Figure S1, at least 13 KTI genes are clustered in a small region on chromosome 8. Interestingly, we also identified a putative TGACG-Binding (TGA) transcription factor (TF) that is tightly linked to the KTI gene cluster at Chr 8. Arabidopsis TGA TFs play a positive role in systemic acquired resistance (SAR) that is crucial in plant immunity (Hussain et al., 2018). Therefore, breeding of low TI soybean lines resulting in the loss or mutation of both of the KTI genes and the TGA TF gene, leading to an increased susceptibility of soybean challenged by phytopathogens.

The *ktt1/3* mutant soybean line generated *via* CRISPR/Cas9-mediated mutagenesis is an isogenic line of wild type WM82 (Table 1). Therefore, it will be an ideal test subject to see if *KTT1/3* has a direct role in plant immunity. Since *KTT1/3* were almost only expressed in seeds (Figure 1) (Gillman et al., 2015), the knockout of these two genes may not interfere with plant immunity in non-seed tissue, which deserves to be further investigated in the future.

In this study, we identified *ktt1/3* double homozygous mutants as well as *ktt1* single homozygous mutants. Our findings indicate that the *ktt1/3* double homozygous mutant exhibits the lowest levels of both KTI content and trypsin inhibition activity, as illustrated in Figures 4, 5. It is noteworthy that a significant reduction in KTI content was observed in seeds where only KTI1 was fully edited, implying that most of the wild-type *KTI3* alleles in those seeds had been mutated. These results indicate that KTI1 and KTI3 act in synergy to contribute to the KTI content and TI activity in soy proteins. The previous report suggests that the soybean line carrying natural mutations of *ktt1* and *ktt3* has increased BBTI content (Gillman et al., 2015). It is unclear if the increased BBTI content is caused by un-intentional selection during the breeding process or if the expression of *BBTI* genes is increased because of the mutations of two *KTI* genes. Therefore, it will be interesting to test the BBTI content and activity in the seed proteins of the *ktt1/3* mutant generated in this study.

Despite the fact that the CRISPR/Cas9 technique has been successfully utilized to generate various soybean mutants, the current *agrobacterium*-mediated soybean transformation protocol is inefficient and genotype dependent. This limits the wide implementation of CRISPR/Cas9 technique in soybean breeding programs (Yamada et al., 2012). The soybean transformation protocol employs bialaphos as the selection agent (Luth et al., 2015).

The *Bar* gene is used as the selection marker gene and encodes a phosphinothricin acetyltransferase protein that can confer the transformants' resistance to bialaphos. It has been reported that the *Bar* gene expression must be fine-tuned in order to successfully select true transgenic plants (Testroet et al., 2017). The original CRISPR/Cas9 vector, pCut, has used a MAS promoter to express the *Bar* gene (Peterson et al., 2016). However, for unknown reasons, the vector does not work well, even in *Arabidopsis thaliana* (Liu and Zhao, unpublished data).

With the intention of improving the transformation system, we modified the bialaphos selection vector in the pMU3T (Liu et al., 2016). Specifically, we replaced the Kanamycin selection marker gene with the *Bar* gene, whose expression was driven by a MAS promoter (Figure 2A). The MAS promoter is known to be most active in the roots of emerging seedlings and very active in the cotyledons and lower leaves (Langridge et al., 1989). Despite the MAS promoter having a lower level of expression than p35S, populations of transformants created with this promoter show normally distributed expression levels (Perez-Gonzalez and Caro, 2019). Thus, the MAS promoter can be used for functional screening of positive transformants in both of our shoot regeneration and rooting medium supplemented with bialaphos.

It is noteworthy that, before the initiation of stable transformation, we evaluated the effectiveness of gRNAs by using a convenient *Agrobacterium*-mediated transient assay method (Wang et al., 2023). As soybean plants have a long-life cycle (4–6 months), the estimation of the gRNAs' effectiveness helps to avoid the waste of time and enhance the possibility of obtaining authentic gene-edited plants.

Although the soybean cultivars with natural variations on either *KTT1* (PI 68679) or *KTI3* (PI 542044) have been discovered, conventional breeding to develop new cultivars stacking with two mutant alleles *via* crossing will take a long time. In addition, linkage drag might lead to interference with the functions of genes located at the flanking sequences of mutant *ktt1* or *ktt3*. The limited genetic background of natural *ktt1* and *ktt3* mutants may also reduce the genetic diversity of soybean breeding lines with low TI trait, and it can be difficult to stack low TI trait with a bundle of various, desirable traits. In the present study, the *ktt1/3* mutant line was created using cv. WM82, which has a genetic background distinct from accessions that harbor natural *ktt1* and/or *ktt3* mutations. Therefore, it offers a new recourse for breeding low TI traits in soybean practice.

Current soybean transformation protocol is genotype dependent, and only a few cultivars (WM82, Jack, Thorne, etc.) can be efficiently transformed (Yamada et al., 2012). A mutant allele must be created in those transformable cultivars and bred into other elite cultivars *via* marker-assisted selection (MAS). Thus, creating a mutant allele tagged with convenient molecular markers is essential for MAS (Hasan et al., 2021). CRISPR/Cas9-based genome editing can introduce small deletions/insertions to targeted genes, enabling us to develop molecular markers based on the sequences of the insertion and deletion mutation regions. In this study, we tested using single sgRNA and two sgRNAs for generation of mutagenesis on *KTT1* and *KTI3*, respectively (Figure 2B). Interestingly, we observed that all genotyped mutant lines carried an identical gene editing pattern of the *ktt1* gene, where 66 nucleotides between the two gRNAs were deleted. The homozygous *ktt1* allele can be

identified in all tested seeds of the T0 generation, while homozygous *ktt3* alleles were identified in some of those genotyped T0 seeds (Figure 3). Therefore, it is possible that two sgRNAs are more efficient for triggering the gene editing events in early generations of transgenic plants.

MAS has been widely implemented in plant breeding including soybean programs (Hasan et al., 2021). The selection marker of *ktt3* has been developed based on its natural mutant allele, but the molecular marker for the natural mutation of *ktt1* in PI 68679 is still not available (Gillman et al., 2015). Therefore, it is challenging to breed the natural *ktt1/3* mutant alleles into a new cultivar via MAS. In this study, we created co-dominant markers that can distinguish between the wild and mutant alleles of *KTT1/KTT3* and *ktt1/ktt3* based on small deletions created by CRISPR/Cas9 machinery (Figure 6). In addition, a simple gel-electrophoresis-free method can be used to identify plants carrying mutant *ktt1* and *ktt3* alleles (Hirotzu et al., 2010). Thus, MAS makes it possible to effectively breed the new mutant *ktt1/3* alleles into other elite cultivars. Taken together, the whole experimental design may serve as a practical example of how to create and select mutant alleles in crop plants in the future.

Conclusions

The present study developed non-transgenic, low TI soybean mutant in cv. William 82. The mutant gene alleles are tagged with convenient molecular markers that are suitable for high throughput marker-aided selection. We expect the low TI soybean mutant will be widely used to breed low-TI or TI-free soybean cultivars for commercial production in value-added meal industry and for stacking with other valuable agronomic traits in the future.

Methods

Plant materials and growth conditions

Soybean plants were grown in 2.5-gallon pots using Miracle-Gro all-purpose potting soil mix in Keck Greenhouse at Virginia Tech (14h/10h light/dark cycle at 25°C/20°C) for the experiments described herein. The plants were watered by an automatic irrigation system. Soybean transformation was performed at the plant transformation facility at Iowa State University as previously described (Paz et al., 2006; Luth et al., 2015; Ge et al., 2016). The plant growth indicators and maturity period days of WM82 and progeny plants of the T1 generation derived from lines #2 and #5 were measured in the green house. The 4-week-old T1 soybean plants were used for genotyping. The seeds of T1 plants were used for seed weight analysis.

Constructing a soybean KTI gene map

The gene map showing locations of *KTI* genes on soybean chromosomes was made using MapInspect. Locations of all *KTI* genes were obtained from the Phytozome database

(<https://phytozome-next.jgi.doe.gov/>) and plotted on their respective chromosomes.

Bacterial growth

E. coli strains DH5 α and C41 (DE3) (Lucigen, Middleton, WI) were grown on Luria agar medium at 37°C. *Agrobacterium tumefaciens* (*A. tumefaciens*) EHA105 was grown on Luria agar medium at 28°C (Zhao et al., 2011; Traore et al., 2019). *E. coli* antibiotic selections used in this study were as follows: 50 μ g/ml kanamycin, 100 μ g/ml carbenicillin, 100 μ g/ml spectinomycin. *A. tumefaciens* antibiotic selection were 100 μ g/ml rifampicin, and/or 100 μ g/ml spectinomycin.

Cloning

The open reading frames (ORFs) of *KTT1* and *KTT3*, were amplified from the genomic DNA of WM82. The *KTT1* _{Δ 66bp}, truncated ORF of *KTT1*, was amplified from the genomic DNA of mutant soybean plant #2-1. All PCR primers with annotations are listed in Table S2. The genes/fragments were then cloned into a pDonr207 plasmid (Thermo Fisher Scientific) for future use.

T1 plant genomic DNA (gDNA) was used as the templates to amplify *KTT1* and/or its mutant allele, and *KTT3* and/or its mutant allele by PCR. The purified PCR fragments were used for genotyping by Sanger sequencing at Virginia Tech Genomic Sequencing Center and cloned to the PCR8/GW/TOPO vector by TA cloning (Invitrogen) for molecular marker tests.

In order to apply the CRISPR/Cas9 system to gene editing in soybean, we modified our current CRISPR/Cas9 construct (Liu et al., 2016). The cassette consists of a MAS promoter, the bialaphos resistant gene, and a MAS terminator that was amplified using plasmid DNA of pEarleyGate101 as the template. All PCR primers with annotations are listed in Table S2. The cassette was assembled to the backbone of CRISPR/Cas9 construct using Gibson Assembly[®] Cloning Kit (New England Biolabs Inc). Since 38 *KTI* genes in soybean share conserved sequences, an alignment was conducted for these genes to design gRNAs that can exclusively target on *KTT1* and *KTT3*. The gRNAs were synthesized in one cassette at GenScript Biotech Corp. The backbone of the new CRISPR/Cas9 construct and the fragment of gRNAs were assembled together using Gibson Assembly[®] Cloning Kit.

Expression analysis of KTI genes in WM82

RNA sequencing data, in FPKM (fragments per kilobase of transcript per million fragments mapped), of 38 *KTI* genes in 30 different tissue types from Williams 82 were acquired through the Gene Networks in Seed Development database (<http://seedgenenetwork.net/sequence>). Construction of the heatmap to visualize expression data was done using the heatmap.2 function from the ggplot2 package in R. A green/blue color gradient was chosen to show expression with blue representing little to no

expression and green representing high expression. The code for the heatmap is as follows:

```
heatmap.2(x=KTI Expression, main = "KTI Expression In
Different Soybean Tissue", notecol="black", density.info="none",
trace="none", margins = c(12,9), col=my_palette,
breaks=col_breaks, dendrogram="row", Colv="NA", ylab= "Genes",
xlab= "Tissue Type", cexCol=.9,cexRow = .8)
```

RNA isolation and real-time PCR

All RNA was extracted from V98-9005 and V03-5903 seeds using TRIzol reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. Any DNA residue was eliminated by treating with UltraPure DNase I (Thermo Fisher Scientific). The integrity and quantity of total RNA were determined by electrophoresis in 1% agarose gel and a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). cDNA synthesis was performed using the SuperScript III First-Strand RT-PCR Kit (Thermo Fisher Scientific) with an oligo-dT primer based on the manufacturer's instructions. Real-time PCR was conducted with cDNA as the template using the Quantitect SYBR Green PCR kit (Qiagen) according to the manufacturer's protocol. Oligo primers are listed in Table S2. The soybean *ELF1B* gene (Gm02g276600) was used as reference gene, and data is presented as Δ CT (Jian et al., 2008).

Expression and purification of KTI1, KTI1 Δ 22aa, and KTI3 proteins

The *KTI1*, *KTI1* Δ 66bp, and *KTI3* genes in pDonr207 were subcloned into a Gateway compatible pET28a destination vector via a LR[®] Gateway cloning kit (Thermo Fisher Scientific) (Earley et al., 2006; Liu et al., 2016). The plasmids were transformed into *E. coli* C41 cells (Lucigen). KTI1, KTI1 Δ 22aa, and KTI3 proteins were expressed and purified following a procedure as previously described (Liu et al., 2020). Protein purity was evaluated by SDS-PAGE. The protein concentration was determined by a protein assay kit (Bio-Rad) using bovine serum albumin as standard (Han et al., 2015).

Standard bioassay to measure trypsin inhibitor activity

A TI activity bioassay was performed following American Association of Cereal Chemists Official Method 22-40 (AACC, 1999) with some modifications previously reported (Rosso et al., 2018). Briefly, 30 mg of finely ground soybean seed powder was mixed with 3 mL of 9 mM HCl (pH 2.0). The mixture was shaken for 1 h at room temperature. 2 mL of the extracts was centrifuged at 10,350 rpm for 20 min at room temperature, and the supernatant was diluted by 10 times with 9 mM HCl for measuring TI activity. A TI activity assay was performed in a 96-well plate format following the same steps described previously (Rosso et al., 2018). Each sample row was repeated three times. Portions of diluted HCl extracts (0, 20, 30,

40, and 60 μ L) or 50 μ g recombinant proteins of KTI1, KTI1 Δ 22aa, and KTI3 were pipetted into the microplate wells, and the volume was adjusted to 60 μ L with 9 mM HCl. 60 μ L of extractant was used as a sample blank and 60 μ L of water were used as a substrate blank. 60 μ L of trypsin (from bovine pancreas, Sigma-Aldrich T8003) solution was added to each sample well, and the microplates were placed in an oven at 37°C for 15 min. After the incubation, 150 μ L of BAPNA substrate pre-warmed at 37°C was added to all wells, and the plates were incubated for exactly 10 min at 37°C. The reaction was stopped by adding 30 μ L of acetic acid solution to all wells. The absorbance of each well was read on a plate reader (FLOUstar Omega, BMG Labtech) at 410 nm for 30 s after shaking at 700 rpm.

HPLC method to quantify Kunitz trypsin inhibitor

The HPLC method to quantify *KTI* was performed following the method developed previously (Rosso et al., 2018). Briefly, 10 mg of finely ground soybean seed powder was mixed with 1.5 mL of 0.1 M sodium acetate buffer (pH 4.5). Samples were vortexed and shaken for 1 h at room temperature. The sample was centrifuged at 12,000 rpm for 15 min. 1 mL of the supernatant was filtered through a syringe with an IC Millex-LG 13-mm mounted 0.2-mm low protein binding hydrophilic millipore (polytetrafluoroethylene [PTFE]) membrane filter (Millipore Ireland). The *KTI* in solution was separated on an Agilent 1260 Infinity series (Agilent Technologies) equipped with a guard column (4.6 x 5 mm) packed with POROS R2 10-mm Self Pack Media and a Poros R2/H perfusion analytical column (2.1 x 100 mm, 10 μ m). The mobile Phase A consisted of 0.01% (v/v) trifluoroacetic acid in Milli-Q water, and the mobile Phase B was 0.085% (v/v) trifluoroacetic acid in acetonitrile. The injection volume was 10 μ L and the detection wavelength was 220 nm.

Development of molecular selection markers with a gel electrophoresis free method for high throughput screening

The transgene free and double homozygous mutant line, #5-26, was selected for the development of molecular selection markers. Based on the genotyping data of #5-26, two pairs of markers were designed: ZW1 with ZW2 or ZW3. ZW1 is the common reverse primer for both *KTI1* and *kTI1*, while ZW2 and ZW3 are two reverse primers matched with unique sequences in *KTI1* and *kTI1*, respectively. Similarly, two pairs of molecular markers, ZW4 with ZW5 or ZW6, were designed. ZW4 is the common reverse primer for both *KTI3* and *kTI3*, while ZW5 and ZW6 are two reverse primers matched with unique sequences in *KTI3* and *kTI3*, respectively. The gDNA of WM82, #5-26 (homozygous mutants of both *kTI1* and *kTI3*), #5-9 (homozygous mutant of *kTI1* while heterozygous mutant of *kTI3*) and #2-30 (homozygous mutant of *kTI3* while heterozygous mutant of *kTI1*) were used as templates to test the efficiency and reliability of these markers in PCR.

PCR amplifications were performed in a total volume of 20 μ L containing 50 ng of gDNA, 0.5 μ M each of forward and reverse primers

(Table S2), 10 µl 2X BioMix Red (Bioline) and ddH₂O. The PCR program was set to be 95°C for 5 min for pre-denature, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 30 s, followed by final extension at 72°C for 5 min.

In order to screen the large-scale progenies derived from crosses between soybean elite cultivars carrying wild type *KTI1/3* and the newly developed mutant plant carrying *ktil/3*, a simple gel electrophoresis-free method was designed. 1X sybreen dye (Thermo Fisher) was added to complete PCR reactions, and the solution was incubated for 10 mins at 75°C before placing in the gel doc (Biorad) for imaging the fluorescent signals. Only the positive PCR products with the dye will display fluorescent signals while the failed PCR will not show signals.

Statistical data analysis

Analytical experiments were performed with at least three technical replicates. Statistical significance was based on one-way ANOVA test for multiple comparisons. Data was analyzed using JMP Pro14. Values of $P < 0.05$ were considered significant.

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

BYZ and ZW designed the experiments and analyzed data. ZW performed experiments, analyzed data, and wrote the manuscript. ZS, LR, CS, JL, and PB also performed experiments. BYZ, BZ, ZS, LR, and PB edited the manuscript. BZ supervised the project. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Virginia Soybean Board (#467059) and a seed grant from the Translational Plant Science

Center at Virginia Tech. An integrated internal competitive grant from the College of Agriculture and Life Sciences at Virginia Tech and Virginia Agricultural Experiment Station (VA160144).

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.2023.1111680/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Physical mapping of 38 *GmKTI* genes on soybean 20 chromosomes. The gene map showing locations of *KTI* genes on soybean chromosomes was made using MapInspect. As displayed in the map, 38 *KTI* genes are located on 9 out of 20 chromosomes.

SUPPLEMENTARY FIGURE 2

SDS-PAGE of purified recombinant proteins and in-frame mutated protein of *KTI1_{A22aa}* nearly lost the TIA. (A) SDS-PAGE was used to assess the purity of three recombinant proteins, *KTI1*, *KTI3*, and *KTI1_{A22aa}*. (B) Purified proteins of *KTI1* and *KTI3*, but not *KTI1_{A22aa}* were able to inhibit the trypsin activity *in vivo*. Experiments were conducted with three technical replicates and obtained similar results.

SUPPLEMENTARY FIGURE 3

Sequence information of *KTI1* and/or *KTI3* in the T1 plants used for the development of selection markers. (A) The alignment of mutant *ktil* in T1 plant leaves (#5-9 and #5-26), where the wild type *KTI1* in WM82 was the control. (B) The alignment of mutant *ktil3* in T1 plant leaves (#2-30 and #5-26), where the wild type *KTI1* in WM82 was the control.

References

- Arnaiz, A., Talavera-Mateo, L., Gonzalez-Melendi, P., Martinez, M., Diaz, I., and Santamaria, M. E. (2018). Arabidopsis kunitz trypsin inhibitors in defense against spider mites. *Front. Plant Sci.* 9, 986. doi: 10.3389/fpls.2018.00986
- Batal, A. B., and Parsons, C. M. (2004). Utilization of various carbohydrate sources as affected by age in the chick. *Poultry Sci.* 83, 1140–1147. doi: 10.1093/ps/83.7.1140
- Cai, Y., Chen, L., Liu, X., Guo, C., Sun, S., Wu, C., et al. (2018). CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soya bean. *Plant Biotechnol. J.* 16, 176–185. doi: 10.1111/pbi.12758
- Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., et al. (2015). CRISPR/Cas9-mediated genome editing in soybean hairy roots. *PLoS One* 10, e0136064. doi: 10.1371/journal.pone.0136064
- Chang, C. J., Tanksley, T. D., Knabe, D. A. Jr., and Zebrowska, T. (1987). Effects of different heat treatments during processing on nutrient digestibility of soybean meal in growing swine. *J. Anim. Sci.* 65, 1273–1282. doi: 10.2527/jas1987.6551273x
- Chen, R., Xu, Q., Liu, Y., Zhang, J., Ren, D., Wang, G., et al. (2018). Generation of transgene-free maize Male sterile lines using the CRISPR/Cas9 system. *Front. Plant Sci.* 9, 1180. doi: 10.3389/fpls.2018.01180

- Chen, Y., Xu, Z., Zhang, C., Kong, X., and Hua, Y. (2014). Heat-induced inactivation mechanisms of kunitz trypsin inhibitor and Bowman-Birk inhibitor in soy milk processing. *Food Chem.* 154, 108–116. doi: 10.1016/j.foodchem.2013.12.092
- Cook, D. A., Jensen, A. H., Fraley, J. R., and Hymowitz, T. (1988). Utilization by growing and finishing pigs of raw soybeans of low kunitz trypsin inhibitor content. *J. Anim. Sci.* 66, 1686–1691. doi: 10.2527/jas1988.6671686x
- Cromwell, G. L., Stahly, T. S., and Monegue, H. J. (1991). Amino acid supplementation of meat meal in lysine-fortified, corn-based diets for growing-finishing pigs. *J. Anim. Sci.* 69, 4898–4906. doi: 10.2527/1991.69124898x
- Earley, K. W., Haag, J. R., Pontes, O., Oppen, K., Juehne, T., Song, K., et al. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* 45, 616–629. doi: 10.1111/j.1365-3113X.2005.02617.x
- Gao, W., Long, L., Tian, X., Xu, F., Liu, J., Singh, P. K., et al. (2017). Genome editing in cotton with the CRISPR/Cas9 system. *Front. Plant Sci.* 8, 1364. doi: 10.3389/fpls.2017.01364
- Ge, L., Yu, J., Wang, H., Luth, D., Bai, G., Wang, K., et al. (2016). Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proc. Natl. Acad. Sci. United States America* 113, 12414–12419. doi: 10.1073/pnas.1611763113
- Gillman, J. D., Kim, W. S., and Krishnan, H. B. (2015). Identification of a new soybean kunitz trypsin inhibitor mutation and its effect on Bowman-Birk protease inhibitor content in soybean seed. *J. Agric. Food Chem.* 63, 1352–1359. doi: 10.1021/jf505220p
- Han, Q., Zhou, C., Wu, S., Liu, Y., Triplett, L., Miao, J., et al. (2015). Crystal structure of xanthomonas AvrXo1-ORF1, a type III effector with a polynucleotide kinase domain, and its interactor AvrXo1-ORF2. *Structure* 23, 1900–1909. doi: 10.1016/j.str.2015.06.030
- Hasan, N., Choudhary, S., Naaz, N., Sharma, N., and Laskar, R. A. (2021). Recent advancements in molecular marker-assisted selection and applications in plant breeding programmes. *J. Genet. Eng. Biotechnol.* 19, 128. doi: 10.1186/s43141-021-00231-1
- Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J.* 12, 934–940. doi: 10.1111/pbi.12201
- Hirotsu, N., Murakami, N., Kashiwagi, T., Ujiie, K., and Ishimaru, K. (2010). Protocol: a simple gel-free method for SNP genotyping using allele-specific primers in rice and other plant species. *Plant Methods* 6, 12. doi: 10.1186/1746-4821-6-12
- Hussain, R. M. F., Sheikh, A. H., Haider, I., Qureshi, M., and Linthorst, H. J. M. (2018). Arabidopsis WRKY50 and TGA transcription factors synergistically activate expression of PR1. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00930
- Hwang, D. L., Foard, D. E., and Wei, C. H. (1977). A soybean trypsin inhibitor: crystallization and x-ray crystallographic study. *J. Biol. Chem.* 252, 1099–1101. doi: 10.1016/S0021-9258(19)75211-9
- Hymowitz, T. (1986). Genetics and breeding of soybeans lacking the kunitz trypsin inhibitor. *Adv. Exp. Med. Biol.* 199, 291–298. doi: 10.1007/978-1-4757-0022-0_18
- Jacobs, T. B., Lafayette, P. R., Schmitz, R. J., and Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15, 16. doi: 10.1186/s12896-015-0131-2
- Jian, B., Liu, B., Bi, Y., Hou, W., Wu, C., and Han, T. (2008). Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol. Biol.* 9, 59. doi: 10.1186/1471-2199-9-59
- Jofuku, K. D., and Goldberg, R. B. (1989). Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. *Plant Cell* 1, 1079–1093. doi: 10.1105/tpc.1.11.1079
- Jofuku, K. D., Schipper, R. D., and Goldberg, R. B. (1989). A frameshift mutation prevents kunitz trypsin inhibitor mRNA accumulation in soybean embryos. *Plant Cell* 1, 567. doi: 10.1105/tpc.1.4.427
- Junker, Y., Zeissig, S., Kim, S. J., Barisani, D., Wieser, H., Leffler, D. A., et al. (2012). Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. *J. Exp. Med.* 209, 2395–2408. doi: 10.1084/jem.20102660
- Kim, J., and Kim, J. S. (2016). Bypassing GMO regulations with CRISPR gene editing. *Nat. Biotechnol.* 34, 1014–1015. doi: 10.1038/nbt.3680
- Kunitz, M. (1945). Crystallization of a trypsin inhibitor from soybean. *Science* 101, 668–669. doi: 10.1126/science.101.2635.668
- Langridge, W. H., Fitzgerald, K. J., Koncz, C., Schell, J., and Szalay, A. A. (1989). Dual promoter of agrobacterium tumefaciens mannopine synthase genes is regulated by plant growth hormones. *Proc. Natl. Acad. Sci. United States America* 86, 3219–3223. doi: 10.1073/pnas.86.9.3219
- Li, J., Brader, G., and Palva, E. T. (2008). Kunitz trypsin inhibitor: an antagonist of cell death triggered by phytopathogens and fumonisins in Arabidopsis. *Mol. Plant* 1, 482–495. doi: 10.1093/mp/ssn013
- Liang, G., Zhang, H., Lou, D., and Yu, D. (2016). Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci. Rep.* 6, 21451. doi: 10.1038/srep21451
- Liener, I. E. (1994). Implications of antinutritional components in soybean foods. *Crit. Rev. Food Sci. Nutr.* 34, 31–67. doi: 10.1080/10408399409527649
- Liener, I. E. (1996). Effects of processing on antinutritional factors in legumes: the soybean case. *Archivos Latinoamericanos Nutricion* 44, 48S–54S.
- Liu, Y., Miao, J., Traore, S., Kong, D., Liu, Y., Zhang, X., et al. (2016). SacB-SacR gene cassette as the negative selection marker to suppress agrobacterium overgrowth in agrobacterium-mediated plant transformation. *Front. Mol. Biosci.* 3, 70. doi: 10.3389/fmolb.2016.00070
- Liu, Y., Wang, K., Cheng, Q., Kong, D., Zhang, X., Wang, Z., et al. (2020). Cysteine protease RD21A regulated by E3 ligase SINAT4 is required for drought-induced resistance to pseudomonas syringae in Arabidopsis. *J. Exp. Bot.* 71, 5562–5576. doi: 10.1093/jxb/eraa255
- Luth, D., Warnberg, K., and Wang, K. (2015). Soybean [Glycine max (L.) Merr]. *Methods Mol. Biol.* 1223, 275–284. doi: 10.1007/978-1-4939-1695-5_22
- Nakayasu, M., Akiyama, R., Lee, H. J., Osakabe, K., Osakabe, Y., Watanabe, B., et al. (2018). Generation of alpha-solanine-free hairy roots of potato by CRISPR/Cas9 mediated genome editing of the St16DOX gene. *Plant Physiol. Biochem.* 131, 70–77. doi: 10.1016/j.plaphy.2018.04.026
- Osakabe, Y., Liang, Z., Ren, C., Nishitani, C., Osakabe, K., Wada, M., et al. (2018). CRISPR-Cas9-mediated genome editing in apple and grapevine. *Nat. Protoc.* 13, 2844–2863. doi: 10.1038/s41596-018-0067-9
- Paz, M. M., Martinez, J. C., Kalvig, A. B., Fonger, T. M., and Wang, K. (2006). Improved cotyledonary node method using an alternative explant derived from mature seed for efficient agrobacterium-mediated soybean transformation. *Plant Cell Rep.* 25, 206–213. doi: 10.1007/s00299-005-0048-7
- Peng, A., Chen, S., Lei, T., Xu, L., He, Y., Wu, L., et al. (2017). Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnol. J.* 15, 1509–1519. doi: 10.1111/pbi.12733
- Perez-Gonzalez, A., and Caro, E. (2019). Benefits of using genomic insulators flanking transgenes to increase expression and avoid positional effects. *Sci. Rep.* 9, 8474. doi: 10.1038/s41598-019-44836-6
- Perez-Maldonado, R. A., Mannion, P. F., and Farrell, D. J. (2003). Effects of heat treatment on the nutritional value of raw soybean selected for low trypsin inhibitor activity. *Br. Poultry Sci.* 44, 299–308. doi: 10.1080/0007166031000085463
- Peterson, B. A., Haak, D. C., Nishimura, M. T., Teixeira, P. J., James, S. R., Dangl, J. L., et al. (2016). Genome-wide assessment of efficiency and specificity in CRISPR/Cas9 mediated multiple site targeting in Arabidopsis. *PLoS One* 11, e0162169. doi: 10.1371/journal.pone.0162169
- Rosso, M. L., Shang, C., Correa, E., and Zhang, B. (2018). An efficient HPLC approach to quantify kunitz trypsin inhibitor in soybean seeds. *Crop Sci.* 58, 1616–1623. doi: 10.2135/cropsci2018.01.0061
- Rosso, M. L., Shang, C., Song, Q., Escamilla, D., Gillenwater, J., and Zhang, B. (2021). Development of breeder-friendly KASP markers for low concentration of kunitz trypsin inhibitor in soybean seeds. *Int. J. Mol. Sci.* 22(5): 2675. doi: 10.3390/ijms22052675
- Schuler, T. H., Poppy, G. M., Kerry, B. R., and Denholm, I. (1999). Potential side effects of insect-resistant transgenic plants on arthropod natural enemies. *Trends Biotechnol.* 17, 210–216. doi: 10.1016/S0167-7799(98)01298-0
- Shao, X., Wu, S., Dou, T., Zhu, H., Hu, C., Huo, H., et al. (2020). Using CRISPR/Cas9 genome editing system to create MaGA2ox2 gene-modified semi-dwarf banana. *Plant Biotechnol. J.* 18, 17–19. doi: 10.1111/pbi.13216
- Testroet, A., Lee, K., Luth, D., and Wang, K. (2017). Comparison of transformation frequency using the bar gene regulated by the CaMV 35S or NOS promoter in agrobacterium-mediated soybean (Glycine max L.) transformation. *In Vitro Cell. Dev. Biol. - Plant* 53, 188–199. doi: 10.1007/s11627-017-9810-0
- Traore, S. M., Ekshtain-Levi, N., Miao, J., Castro sparks, A., Wang, Z., Wang, K., et al. (2019). Nicotiana species as surrogate host for studying the pathogenicity of acidovorax citrulli, the causal agent of bacterial fruit blight of cucurbits. *Mol. Plant Pathol.* 20, 800–814. doi: 10.1111/mpp.12792
- Upadhyay, S. K., Kumar, J., Alok, A., and Tuli, R. (2013). RNA-Guided genome editing for target gene mutations in wheat. *G3* 3, 2233–2238. doi: 10.1534/g3.113.008847
- Vu, T. V., Sivankalyani, V., Kim, E. J., Doan, D. T. H., Tran, M. T., Kim, J., et al. (2020). Highly efficient homology-directed repair using CRISPR/Cpf1-geminiviral replicon in tomato. *Plant Biotechnol. J.* 18(10): 2133–2143. doi: 10.1111/pbi.13373
- Wang, Z., Shea, Z., Li, Q., Wang, K., Mills, K., Zhang, B., et al. (2023). Evaluate the guide RNA effectiveness via agrobacterium-mediated transient assays in Nicotiana benthamiana. *Front. Plant Sci.* 14, 1111683. doi: 10.3389/fpls.2023.1111683
- Wei, C. H. (1983). Crystallization of two cubic forms of soybean trypsin inhibitor E-I, a member of the Bowman-Birk inhibitor family. *J. Biol. Chem.* 258, 9357–9359. doi: 10.1016/S0021-9258(17)44675-8
- Xu, R. F., Li, H., Qin, R. Y., Li, J., Qiu, C. H., Yang, Y. C., et al. (2015). Generation of inheritable and “transgene clean” targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.* 5, 11491. doi: 10.1038/srep11491
- Yamada, T., Takagi, K., and Ishimoto, M. (2012). Recent advances in soybean transformation and their application to molecular breeding and genomic analysis. *Breed. Sci.* 61, 480–494. doi: 10.1270/jsbbs.61.480
- Zhao, B., Dahlbeck, D., Krasileva, K. V., Fong, R. W., and Staskawicz, B. J. (2011). Computational and biochemical analysis of the xanthomonas effector AvrBs2 and its role in the modulation of xanthomonas type three effector delivery. *PLoS Pathog.* 7, e1002408. doi: 10.1371/journal.ppat.1002408
- Zhao, H., Ullah, H., McNeill, M. R., Du, G., Hao, K., Tu, X., et al. (2019). Inhibitory effects of plant trypsin inhibitors msti-94 and msti-16 on thrips (Homoptera: Aphididae) in alfalfa. *Insects* 10(6): 154. doi: 10.3390/insects10060154

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

