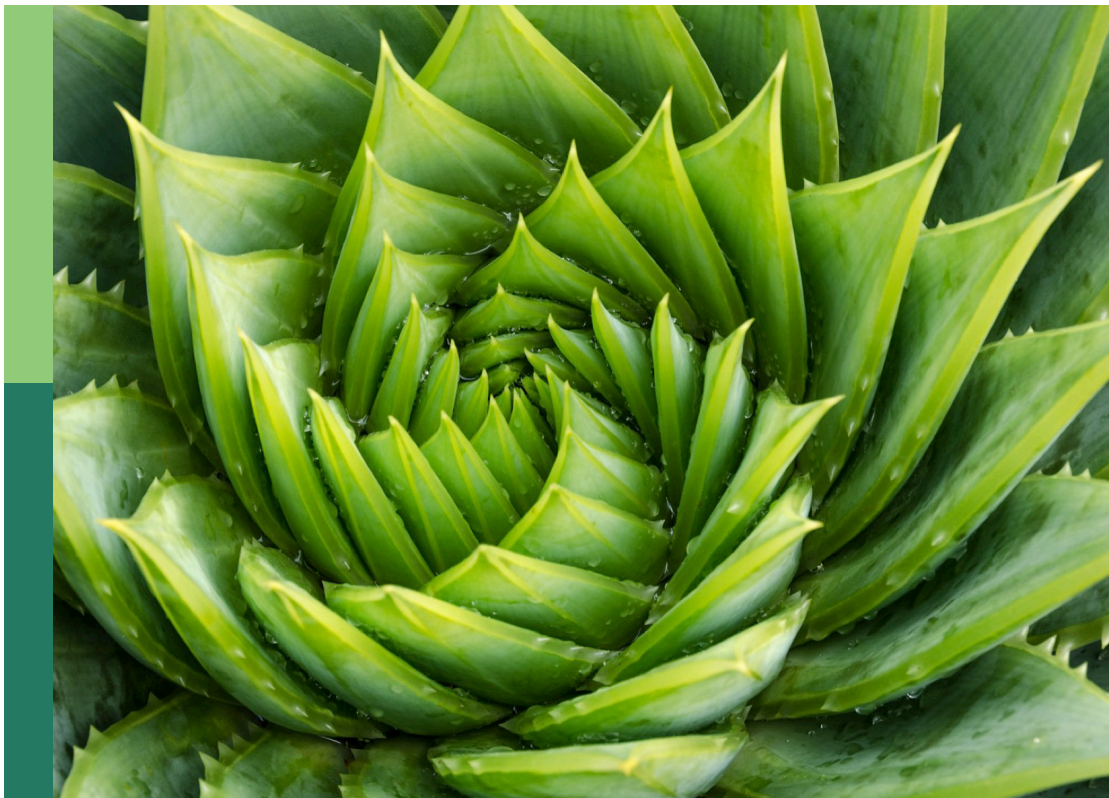
frontiers | Research Topics

Omics in seed development: challenges and opportunities for improving of seed quality and yield in model and crop plants

Edited by
Daoquan Xiang, Zhaorong Hu and Raju Datla

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 978-2-8325-6072-3
DOI 10.3389/978-2-8325-6072-3

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

Omics in seed development: challenges and opportunities for improving of seed quality and yield in model and crop plants

Topic editors
Daoquan Xiang — National Research Council Canada (NRC), Canada
Zhaorong Hu — China Agricultural University, China
Raju Datla — Global Institute for Food Security (GIFS), Canada

Citation
Xiang, D., Hu, Z., Datla, R., eds. (2025). *Omics in seed development: challenges and opportunities for improving of seed quality and yield in model and crop plants*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6072-3

Table of contents

05 Editorial: Omics in seed development: challenges and opportunities for improving of seed quality and yield in model and crop plants
Changye Yang, Zhaorong Hu, Raju Datla and Daoquan Xiang

08 Transcriptome analysis of differential sugar accumulation in the developing embryo of contrasting two *Castanea mollissima* cultivars
Ruimin Huang, Fei Peng, Dongsheng Wang, Fei Cao, Chunlei Guo, Liyang Yu, Jingzheng Zhang and Yuedong Yang

21 Metabolomics and transcriptomics analyses provide new insights into the nutritional quality during the endosperm development of different ploidy rice
Lin Xian, Jiaqi Tian, Yanxi Long, Huijin Ma, Min Tian, Xiangdong Liu, Guoying Yin and Lan Wang

34 Multifaceted roles of transcription factors during plant embryogenesis
Hai Ying Yuan, Sateesh Kagale and Alison M. R. Ferrie

49 SGR-YOLO: a method for detecting seed germination rate in wild rice
Qiong Yao, Xiaoming Zheng, Guomin Zhou and Jianhua Zhang

65 Applying a non-GMO breeding approach with an identified natural variation to reduce food allergen Len c3 in *Lens culinaris* seeds
Jingpu Song, Ioannis Mavraganis, Wenyun Shen, Hui Yang and Jitao Zou

72 Omics-driven advances in the understanding of regulatory landscape of peanut seed development
Zhihui Wang, Yong Lei and Boshou Liao

81 A systems genomics and genetics approach to identify the genetic regulatory network for lignin content in *Brassica napus* seeds
Wentao Zhang, Erin E. Higgins, Stephen J. Robinson, Wayne E. Clarke, Kerry Boyle, Andrew G. Sharpe, Pierre R. Fobert and Isobel A. P. Parkin

94 Understanding grain development in the Poaceae family by comparing conserved and distinctive pathways through omics studies in wheat and maize
Yuanyuan Ji, Thulani Hewavithana, Andrew G. Sharpe and Lingling Jin

105 Genome-wide identification and expression analysis of the U-box E3 ubiquitin ligase gene family related to bacterial wilt resistance in tobacco (*Nicotiana tabacum* L.) and eggplant (*Solanum melongena* L.)
Rui Chen, Gang Gu, Binghui Zhang, Chaofan Du, Xiaolu Lin, Weiwei Cai, Yan Zheng, Tong Li, Ruiqi Wang and Xiaofang Xie

121 Voice from both sides: a molecular dialogue between transcriptional activators and repressors in seed-to-seedling transition and crop adaptation
Dongeun Go, Bailan Lu, Milad Alizadeh, Sonia Gazzarrini and Liang Song

141 Pulse protein quality and derived bioactive peptides
Matthew G. Nosworthy, Bianyun Yu, L. Irina Zaharia, Gerardo Medina and Nii Patterson

149 Applications of synchrotron light in seed research: an array of x-ray and infrared imaging methodologies
Paula Ashe, Kaiyang Tu, Jarvis A. Stobbs, James J. Dynes, Miranda Vu, Hamid Shaterian, Sateesh Kagale, Karen K. Tanino, Janitha P. D. Wanasundara, Sally Vail, Chithra Karunakaran and Teagen D. Quilichini

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 20 February 2025
DOI 10.3389/fpls.2025.1568039

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Ling-Ling Chen,
Guangxi University, China

*CORRESPONDENCE
Daoquan Xiang
daoquan.xiang@nrc-cnrc.gc.ca

RECEIVED 28 January 2025
ACCEPTED 03 February 2025
PUBLISHED 20 February 2025

CITATION
Yang C, Hu Z, Datla R and Xiang D (2025) Editorial: Omics in seed development: challenges and opportunities for improving of seed quality and yield in model and crop plants. *Front. Plant Sci.* 16:1568039. doi: 10.3389/fpls.2025.1568039

COPYRIGHT
© 2025 His Majesty the King in Right of Canada, as represented by the National Research Council of Canada. 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: Omics in seed development: challenges and opportunities for improving of seed quality and yield in model and crop plants

Changye Yang[1], Zhaorong Hu[2], Raju Datla[3] and Daoquan Xiang[1]*

[1]Aquatic and Crop Resource Development, National Research Council Canada, Saskatoon, SK, Canada, [2]Frontiers Science Center for Molecular Design Breeding, Key Laboratory of Crop Heterosis and Utilization, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing, China, [3]Global Institute for Food Security, University of Saskatchewan, Saskatoon, SK, Canada

KEYWORDS

seed development, omics, genomics, transcriptomics, proteomics, metabolomics

Editorial on the Research Topic
Omics in seed development: challenges and opportunities for improving of seed quality and yield in model and crop plants

Introduction

Seeds are a fundamental component in the life cycle of sexually reproducing plants, marking both the beginning of a new generation through germination and the culmination of the reproductive phase through seed production. Plant species exhibit remarkable diversity in seed characteristics, particularly in size, number, and composition, with seeds developing through precisely coordinated programs across embryo, endosperm, and seed coat compartments. In crop plants, seeds represent the most economically significant products, directly influencing both crop quality and yield. Recent advances in omics technologies including genomics, transcriptomics, proteomics, and metabolomics have dramatically enhanced our understanding of seed biology (Liu et al., 2022). These comprehensive findings have provided unprecedented insights into the genetic and molecular mechanisms underlying seed development, germination, and composition (Chen et al., 2023; Yu et al., 2023; Klcova et al., 2024). Seed omics studies have facilitated the identification of key genes and pathways associated with essential traits, contributing to the development of advanced breeding strategies that improve desirable attributes such as nutritional content, fertilization efficiency, and yield (Yuan et al., 2024). Moreover, these studies play a crucial role in optimizing seed quality and enhancing crop resilience to various environmental and climatic challenges. As global food demand continues to rise, insights from seed omics research have become increasingly vital for achieving sustainable agriculture and food security goals.

This Research Topic focuses on recent advances in seed omics research, comprising twelve articles that explore diverse aspects of the field. Five review articles examine recent progress in seed omics, while seven research articles present findings on omics data processing, new methodological developments, and seed trait analysis, as detailed below.

New technology and methods

Several articles in this Research Topic have developed new approaches in seed structural imaging, machine learning detection, and breeding procedures. The study by Ashe et al., presented a suite of Synchrotron radiation (SR) related imaging methodologies for applications in research studies using plants seeds. The datasets generated from this study represent diverse plant species that include *Citrullus* sp. (watermelon), *Brassica* sp. (canola), *Pisum sativum* (pea), and *Triticum durum* (wheat). The authors have introduced the SR micro-computed tomography (SR-μCT) non-destructive imaging method, with advanced capabilities for unveiling detailed internal seed microstructures and their three-dimensional morphologies. Additionally, presented methods for using synchrotron X-rays, including X-ray absorption spectroscopy (XAS) and X-ray fluorescence (XRF) imaging to reveal elemental structural distributions that allow the spatial mapping of micronutrients in seed sub-compartments to determine their speciation characteristics.

Once the large omics datasets are acquired, the researchers focus on processing technologies. In this context, a deep learning network SGR-YOLO, proposed by Yao et al., specifically designed to detect seed germination rates in wild rice. The backbone of the network is based on YOLOv7 with the addition of a bi-directional feature pyramid network (BiFPN) and ECA lightweight attention mechanism. The trained model presented in this study will facilitate processing of images of wild rice grains in hydroponic boxes and Petri dishes, outputting bounding boxes that identify each grain, irrespective of germination status.

Moreover, Song et al. employed a non-GMO breeding approach to reduce food allergen *Len c3* in *Lens culinaris* seeds. The study, first identified a lipid transfer protein (LTP) gene *Lcu.2RBY.4g013600* that encodes the lentil allergen *Len c3*. The authors of this study introduced gene screening to search for natural mutations of the *Len c3* allergen-encoding gene for mitigation of food allergen *Len c3*. This research resulted in the selection of lentil hybrids with reduced allergenic traits.

Reviews on current advancements

Nearly every important trait of seeds, such as nutritional content, is regulated by a multitude of factors. For instance, Nosworthy et al. reviewed studies related to the protein content and quality of pulses protein crops, reflecting the growing interest in the development of plant-based proteins. This review highlighted the quality of pulse proteins is likely determined by the Protein Efficiency Ratio and Protein Digestibility Corrected Amino Acid Score as well as bioactive properties of specific bioactive peptides related to amelioration of hypertension and diabetes. In another review, Yuan et al. explored the diverse roles of transcription factors (TFs) during zygotic embryogenesis (ZE) and somatic embryogenesis (SE). This review focused on the master TFs involved in embryogenesis, such as *BABY BOOM* (*BBM*) from the *APETALA2*/Ethylene-Responsive Factor (*AP2*/*ERF*) family, *WUSCHEL* and *WUSCHEL*-related homeobox (*WOX*) from the homeobox family, etc. This review also covered several plant species and discussed future perspectives on the role of transcription factors in plant zygotic embryogenesis. Furthermore, Go et al. from the University of British Columbia reviewed the role of transcriptional activators and repressors in epigenetic repression of the embryonic and seed maturation programs in seedlings. Their review illustrates the roles of core proteins and accessory components in the epigenetic machinery, as well as the similarities in the activation and repression of the embryonic and seed maturation programs.

Extensive research has also been conducted on generating various omics data for seed research. Wang et al. summarized the achievements in peanut seed development regulation and trait analysis based on reference genome-guided omics studies. Following the completion of the peanut reference genome, numerous studies have utilized omics data to elucidate the quantitative trait loci (QTL) associated with seed weight, oil content, and protein content. This review highlights the significant progresses in the understanding of molecular basis of peanut seed development and demonstrates the importance of omics data in elucidating the regulatory mechanisms involved. In another review, Ji et al. summarized the recent studies using the application of omics technologies such as genomics, transcriptomics, proteomics and metabolomics. These studies shed light on the mechanisms underlying seed development in wheat and maize, representing the C3 and C4 plants in *Poaceae* family. The *Poaceae* family, commonly known as the grass family, due to its diverse range of species, represents economically the most important crops for human society. The diversity represented in these species also offers photosynthetic pathways of C3 and C4 plants which introduces intriguing variations in their physiological and biochemical processes, potentially affecting their respective seed developmental programs.

Seed omics data processing

In this Research Topic, several studies also presented their research on analyzing seed behavior using omics data. Huang et al. conducted transcriptome analysis of differential sugar accumulation in the embryos of *Castanea mollissima* (Chinese chestnut). This study utilized the metabolomic and transcriptomic datasets to investigate metabolites and genes related to sugar in two Chinese chestnut cultivars representing the high and low sugar variants. The conclusion is that the high-sugar cultivar promoted the conversion of starch to sucrose, partly due to a strong increase of the activity of the SUS-synthetic enzyme. Beyond Chinese chestnuts, a different group of researchers studied the nutritional quality of different ploidy rice using their metabolomics and transcriptomics data. In this study Xian et al. tested two different types of rice, autotetraploid rice (AJNT-4x) and diploid rice (AJNT-2x), at various time points

during the seed's endosperm development. Their results established new pathways linking gene expression to key metabolites.

Two genome-wide studies have also been conducted on different plant species. One of the studies, conducted by Chen et al. linked U-box E3 ubiquitin ligase gene family to bacterial wilt resistance in tobacco (*Nicotiana tabacum L.*) and eggplant (*Solanum melongena L.*). A total of 116 NtU-box genes from tobacco and 56 SmU-box genes from eggplant were identified in the genome, which were also categorized into five subfamilies. In addition, phylogenetic analysis also suggested a shared ancestor predating the divergence of six species (tobacco, eggplant, potato, tomato, *Arabidopsis thaliana*, and pepper). In the other study, Zhang et al. employed genomics and genetics approach to identify the genetic regulatory network for lignin content in *Brassica napus* seeds using network-based systems that integrate genotype, phenotype, and molecular phenotypes, four QTLs, eighty-three subnetworks, and three modules with 910 genes have been shown to be associated with lignin content.

Conclusions

Altogether, the studies within this Research Topic have showcased innovative techniques in seed research, analyzed seed omics data, and also reviewed current advancements in the field. The progress in seed research methodologies and the analysis of omics data provide comprehensive insights into seed development, germination, and nutritional content. These advancements will assist researchers to identify key proteins, metabolic pathways, and genes associated with desirable seed traits, paving the way for the development of more nutritious and higher yielding seed varieties. With ongoing research and interdisciplinary collaboration, seed omics datasets hold great potential for driving further future innovations in the field.

Author contributions

DX: Writing – original draft, Writing – review & editing. CY: Writing – original draft, Writing – review & editing. ZH: Writing – original draft, Writing – review & editing. RD: Writing – original draft, Writing – review & editing.

Conflict of interest

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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

Chen, L., Qin, L., Zhang, Y., Xu, H., Bu, Y., Wu, R., et al. (2023). Insights from multi-omics integration into seed germination of Taxus *chinensis* var *mairei*. *Commun. Biol.* 6, 931. doi: 10.1038/s42003-023-05307-x

Klcova, B., Balarynova, J., Trneny, O., Krejci, P., Cechova, M. Z., Leonova, T., et al. (2024). Domestication has altered gene expression and secondary metabolites in pea seed coat. *Plant J.* 118, 2269–2295. doi: 10.1111/tpj.v118.6

Liu, N., Liu, J., Fan, S., Liu, H., Zhou, X., Hua, W., et al. (2022). An integrated omics analysis reveals the gene expression profiles of maize, castor bean, and rapeseed for seed oil biosynthesis. *BMC Plant Biol.* 22, 153. doi: 10.1186/s12870-022-03495-y

Yu, B., Gao, P., Song, J., Yang, H., Qin, L., Yu, X., et al. (2023). Spatiotemporal transcriptomics and metabolic profiling provide insights into gene regulatory networks during lentil seed development. *Plant J.* 115, 253–274. doi: 10.1111/tpj.v115.1

Yuan, X., Jiang, X., Zhang, M., Wang, L., Jiao, W., Chen, H., et al. (2024). Integrative omics analysis elucidates the genetic basis underlying seed weight and oil content in soybean. *Plant Cell* 36, 2160–2175. doi: 10.1093/plcell/koae062

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 June 2023
DOI 10.3389/fpls.2023.1206585

Check for updates

OPEN ACCESS

EDITED BY
Sang-Kyu Lee,
Gyeongsang National University,
Republic of Korea

REVIEWED BY
Lae-Hyeon Cho,
Pusan National University, Miryang,
Republic of Korea
Su-Hyeon Shim,
Kyung Hee University, Republic of Korea

*CORRESPONDENCE
Yuedong Yang
yuedongyang@hotmail.com

RECEIVED 16 April 2023
ACCEPTED 05 June 2023
PUBLISHED 19 June 2023

CITATION
Huang R, Peng F, Wang D, Cao F, Guo C, Yu L, Zhang J and Yang Y (2023) Transcriptome analysis of differential sugar accumulation in the developing embryo of contrasting two *Castanea mollissima* cultivars.
Front. Plant Sci. 14:1206585.
doi: 10.3389/fpls.2023.1206585

COPYRIGHT
© 2023 Huang, Peng, Wang, Cao, Guo, Yu, Zhang and Yang. 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.

Transcriptome analysis of differential sugar accumulation in the developing embryo of contrasting two *Castanea mollissima* cultivars

Ruimin Huang[1,2,3], Fei Peng[2], Dongsheng Wang[1,3], Fei Cao[3,4], Chunlei Guo[1,3], Liyang Yu[1,3], Jingzheng Zhang[1,3,4] and Yuedong Yang[1,2]*

[1]Engineering Research Center of Chestnut Industry Technology, Ministry of Education, Hebei Normal University of Science and Technology, Qinhuangdao, Hebei, China, [2]Hebei Key Laboratory of Active Components and Functions in Natural Products, Hebei Normal University of Science and Technology, Qinhuangdao, Hebei, China, [3]Hebei Collaborative Innovation Center of Chestnut Industry, Qinhuangdao, Hebei, China, [4]Hebei Key Laboratory of Horticultural Germplasm Excavation and Innovative Utilization, College of Horticulture Science and Technology, Hebei Normal University of Science and Technology, Changli, Hebei, China

Chinese chestnut (*Castanea mollissima*) is an important nut tree species, and its embryo is rich in sugar. We combined metabolomic and transcriptomic data to analyze metabolites and genes related to sugar in two Chinese chestnut cultivars at 60, 70, 80, 90 and 100 days after flowering (DAF). The soluble sugar content of high-sugar cultivar at maturity is 1.5 times that of low-sugar cultivar. Thirty sugar metabolites were identified in embryo, with the most dominant being sucrose. Analysis of the gene expression patterns revealed that the high-sugar cultivar promoted the conversion of starch to sucrose by up-regulating genes related to starch degradation and sucrose synthesis at 90-100 DAF. It also strongly increased the enzyme activity of SUS-synthetic, which may promote sucrose synthesis. Gene co-expression network analysis showed that ABA and peroxide were related to starch decomposition during Chinese chestnut ripening. Our study analyzed the composition and molecular synthesis mechanism of sugar in Chinese chestnut embryos, and provided a new insight into the regulation pattern of high sugar accumulation in Chinese chestnut nuts.

KEYWORDS

embryogenesis, sucrose, starch metabolism, cold, ABA

Abbreviations: DAF, days after flowering; DEG, differentially expressed genes; CINV, cytosolic invertase; UDPG, fructose and uridine diphosphate glucose; SUS, sucrose synthase; ADPG, ADP-glucose; GBSS, granule-bound starch synthase; SS, soluble starch synthetase; SBE, starch branching enzyme; ISA, isoamylase type starch debranching enzyme; GWD, glucan water dikinase; PWD, phosphoglucan water dikinase; SPS, sucrose phosphate synthase; SPP, sucrose phosphate phosphatase; AMY, α-Amylase; BAM, β-Amylase; DBE, starch debranching enzyme; LDA, limit dextrinase; PHS, alpha-glucan phosphorylase; DPE, disproportionating enzyme; MEX, maltose excess; GLT, glucose transporter; FK, fructose kinase; UGP, uridyl diphosphate glucose pyrophosphorylase; HXK, hexokinase; PGI, glucose-6-phosphate isomerase; PGM, phosphoglucomutase; GPT, glucose-6-phosphate transporter; PGM, phosphoglucomutase; AGP, adenosine diphosphate-glucose pyrophosphorylase; COR, cold-regulated.

Introduction

Chestnut belongs to the Fagaceae family, which is mainly cultivated in China, Bolivia, Turkey, the Republic of Korea and Italy (Massantini et al., 2021). Chestnuts are rich in starch, minerals, vitamins and phytonutrients, relatively low in fat, and gluten free, making them a healthy and nutritious food (Warmund, 2011; Haytowitz et al., 2018). The composition of carbohydrates has an important effect on the quality of chestnuts. More than 70 percent of respondents said that taste was the most important factor in their decision to buy chestnuts, and sensory analysis showed that chestnut acceptance was strongly dependent on sweetness and related to sucrose content (Künsch et al., 2001; Santos et al., 2022).

Starch is the main form of carbohydrates in chestnuts, and it can be transformed into each other with sucrose (Huang et al., 2021). Sugar and starch metabolism depends on the correct spatial and temporal activity of many enzymes (Chen et al., 2021; Samkumar et al., 2022). During nut development, sucrose is transported to the embryos as a product of photosynthesis. Sucrose can also be converted to glucose and fructose by cytosolic invertase (CINV), or to fructose and uridine diphosphate glucose (UDPG) by sucrose synthase (SUS) (Wan et al., 2018; Moshchenskaya et al., 2019). Then, ADP-glucose (ADPG) is formed through a multi-part reaction and is the common precursor for the synthesis of amylose and amylopectin (Streb and Zeeman, 2012). The synthesis of amylose is catalyzed by granule-bound starch synthase (GBSS). In addition, the synthesis of amylopectin is catalyzed by soluble starch synthetase (SS), starch branching enzyme (SBE) and Isoamylase type starch debranching enzyme (ISA) (Yang et al., 2022).

On the other hand, starch stored in plants will be decomposed into soluble sugar for life activities. Glucan water dikinase (GWD) and phosphoglucan water dikinase (PWD) can phosphate starch and loosen the structure of starch granules in plastid (Mahlow et al., 2016; Xu et al., 2017). Phosphorylated starch can be degraded by a set of glucan-hydrolyzing enzymes (i.e., AMY and BAM) to produce maltose and glucose. Both maltose and glucose can be transported into cytosol to synthesize UDPG (Zhang et al., 2014; Schreier et al., 2019). Some important enzymes such as sucrose phosphate synthase (SPS), sucrose phosphate phosphatase (SPP) and SUS can catalyze UDPG to generate sucrose (Maloney et al., 2015).

Sugar metabolism is regulated by many factors and is easily affected by environment. Low temperatures can destroy the balance of starch and sucrose metabolism, and promote the starch degradation into soluble sugar (Zhang et al., 2014). The contents of glucose, fructose, sucrose, raffinose and galactitol in *Arabidopsis* leaves increased after cold acclimation (Bieniawska, 2007). ABA may be one of the important hormones regulating fruit ripening, which may accelerate sucrose accumulation and promote fruit ripening quality by increasing the enzyme activity of SUS (Wang et al., 2017). Exogenous ABA can increase soluble sugar content in apple fruits at mature stage, and the appropriate concentration of ABA can accelerate glucose conversion and promote sucrose accumulation, which is consistent with the trend of sugar conversion at mature stage (Xiao et al., 2018).

Chinese chestnut is a traditional nut of China and is widely cultivated in East Asia (LaBonte et al., 2018). Hebei Province is the main producing area of Chinese chestnut in China. The four seasons of Hebei Province are clear, and the unique climate may affect the quality of Chinese chestnut (He et al., 2021). In this study, we studied the changes in gene expression related to sugar in the embryos of two Chinese chestnut cultivars, and compared the composition of sugar in the five developmental stages. As a result, the mechanism of high sugar accumulation of Chinese chestnuts was analyzed, and it provided prospects for improving the sugar content of Chinese chestnut cultivars.

Materials and methods

Plant material

Two Chinese chestnut cultivars, high-sugar cultivar 'Yanlong' (HS) and relatively low-sugar cultivar 'Yanshanzaofeng' (LS), were used as plant materials. The nuts of two Chinese chestnut cultivars were collected from the six 11-year-old trees at the Chinese chestnut base in Hebei, China (118°80′E, 40°13′N). Each tree received standard agronomic practices. 'Yanshanzaofeng' (LS) has female flowers in full bloom on June 5, and 'Yanlong' (HS) has female flowers in full bloom on June 25. Flowers that were at the anthesis stage simultaneously were marked, and fruits were collected at 60, 70, 80, 90, and 100 DAF (fruit maturity) in 2021. At each stage of development, nine nuts were mixed into a biological repetition from the selected three trees, with three repetitions per each stage, and then samples were quickly frozen with liquid nitrogen and stored in a refrigerator at -80°C.

Analysis of sugar content

The freeze-dried samples were ground into powder with a mixer mill (MM 400, Retsch) by filtration through a 0.5 mm filter. The sucrose content (about 50 mg of sample powder) was determined using the anthrone colorimetric method (Roggo et al., 2002). The contents of starch components (amylose and amylopectin) were determined through dual-wavelength spectroscopy, with amylopectin measured at 550 nm and 695 nm and amylose measured at 617 nm and 475 nm, using a spectrophotometer (UV-2102C) (Juliano et al., 1981). All the determinations were performed in triplicate.

GC-MS analysis of sugar metabolites

Agilent 8890 gas chromatograph coupled to a 5977B mass spectrometer with a DB-5MS column (30 m length × 0.25 mm i.d. × 0.25 μm film thickness, J&W Scientific, USA) was employed for GC-MS analysis of sugars by MetWare (http://www.metware.cn/). Helium was used as carrier gas, at a flow rate of 1 mL/min. Injections were made in the split mode with a split ratio 5:1 and the intection volume was 1 μL. The oven temperature was held at 160°C for 1 min, and then raised to 200°C at 6°C/min,

raised to 270°C at 10°C/min, raised to 300°C at 5°C/min, raised to 320°C at 20°C/min and held at the temperature for 5.5 min. All samples were analyzed in selective ion monitoring mode. The ion source and transfer line temperature were 230°C and 280°C, respectively (Lowell et al., 1989; Medeiros and Simoneit, 2007; Gomez-Gonzalez et al., 2010).

Transcriptome sequencing

Total RNA was extracted from 0.5 g sample using RNAprep Pure Plant Kit (Tiengen, Beijing, China), and RNA purity was detected using a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). cDNA libraries were prepared using the NEBNext®Ultra™RNA Library Prep Kit (Illumina, San Diego, CA, USA). The NovaSeq 6000 sequencing system (Illumina) was used for sequencing with 150 bp paired-ends by Novogene Company (Beijing, China) (Modi et al., 2021). The raw data of each sample was more than 6 GB. RNA-seq data were uploaded to NCBI and can be accessed through BioProject accession number PRJNA883560.

RNA-Seq data analysis

The Chinese chestnut reference genome and gene model annotation files were downloaded from the website (https://github.com/yongshuai-sun/hhs-omei) (Sun et al., 2020). Clean reads were obtained by filtering raw reads using Perl scripts. Then, the clean reads were compared with the reference genome using Hisat2 v2.0.5 software, and the number of reads mapped to each gene was counted by features v1.5.0-p3 in the subread software to obtain the FPKM value. The differentially expressed genes (DEGs) were analyzed by OmicShare tool (www.omicshare.com/tools) (|log2 (fold change)|≥1, FDR ≤ 0.05). KEGG enrichment analysis was performed using STRING database (http://www.string-db.org). Gene expression data were normalized and plotted using Tbtools V1.09876 software.

The enzyme activity of SUS

SUS was assayed in both the synthetic and cleavage directions. One gram of the frozen powder was resuspended at 4°C in 5 ml of 100 mM HEPES (pH 7.5), 2 mM EDTA and 5 mM dithiothreitol. The medium for SUS synthesis contained in a 0.5 mL volume: extract, 80 mM Hepes (pH 8.5), 5 mm KCN, 5 mm NaF, 100 mM fructose, and 15 mM UDPG. The medium for SUS cleavage in a similar volume consisted of extract, 80 mM MES (pH 5.5), 5 mM NaF, 100 mM sucrose, and 5 mM UDP. Reactions proceeded for 15 min at 30°C and were terminated by boiling for 1 min. UDP production was quantified by measurement of pyruvate kinase-specific loss of NADH in the presence of lactic dehydrogenase, and UDPG production was quantified by measuring UDPG dehydrogenase-specific synthesis of NADH (Baroja-Fernández et al., 2009).

Co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) (V1.69) in R software package was used to construct the gene co-expression network, using the signed-hybrid network type. The co-expression network was mapped using Cytoscape V3.7.1 (https://cytoscape.org/) software. The description of gene function comes from the STRING database (Szklarczyk et al., 2019).

qRT-PCR analysis

Real-time quantitative PCR experiments were performed on ABI 7500 Real-Time PCR system (Applied Biosystems Inc., Foster City, CA, USA) with TB Green Premix Ex Taq (Takara). The instrument settings were: 95°C for 300 s; 40 PCR cycles, with each cycle set at 95°C for 10 s and 60°C for 30 s. The specific primer information was shown in Table S1, in which the *18S* gene of Chinese chestnut was used as the reference gene. The relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method. Three biological replicates were performed.

Results and discussion

Morphological characteristics and sugar content

We have sampled five nut development stages of the two cultivars under the same site conditions. The mature stage of LS was 20 days earlier than HS, and the development temperature of HS is lower than that of LS (Figure 1A). The accumulation of nutrients in two cultivars was mainly carried out from 60 to 100 DAF. At 100 DAF, the dry weight of HS was 6.77g, which was 0.49g heavier than in LS (Figure 1B). The changes of amylose (Figure 1C), amylopectin (Figure 1D), ratio of amylopectin to amylose (Figure 1E) and total starch (Figure 1F) contents in the two cultivars were similar. It is worth noting that the soluble sugar of HS is 9%, which was 1.5 times higher than LS at the mature stage (Figure 1G). The soluble sugar in HS increased rapidly from 6.46% to 6.84% at 90-100 DAF, and the amylopectin decreased from 56.03% to 44.77%. The increase of soluble sugar in HS from 90 to 100 days was due to the decomposition of amylopectin. The difference of temperature and daylength during nut development may be one of the reasons for the difference in soluble sugar content in two cultivars.

Changes in the soluble sugar composition

We tested a total of 32 kinds of sugar metabolites, of which 30 kinds of metabolites were detected, and the two metabolites (D-Arabinitol and L-Rhamnose) were not detected (Table 1). The content of the total sugar metabolites was related to soluble sugar content (Figure 2B), indicating that the experimental data was reliable. The results showed that the soluble sugar composition changed obviously with embryo development. The soluble sugar was mainly

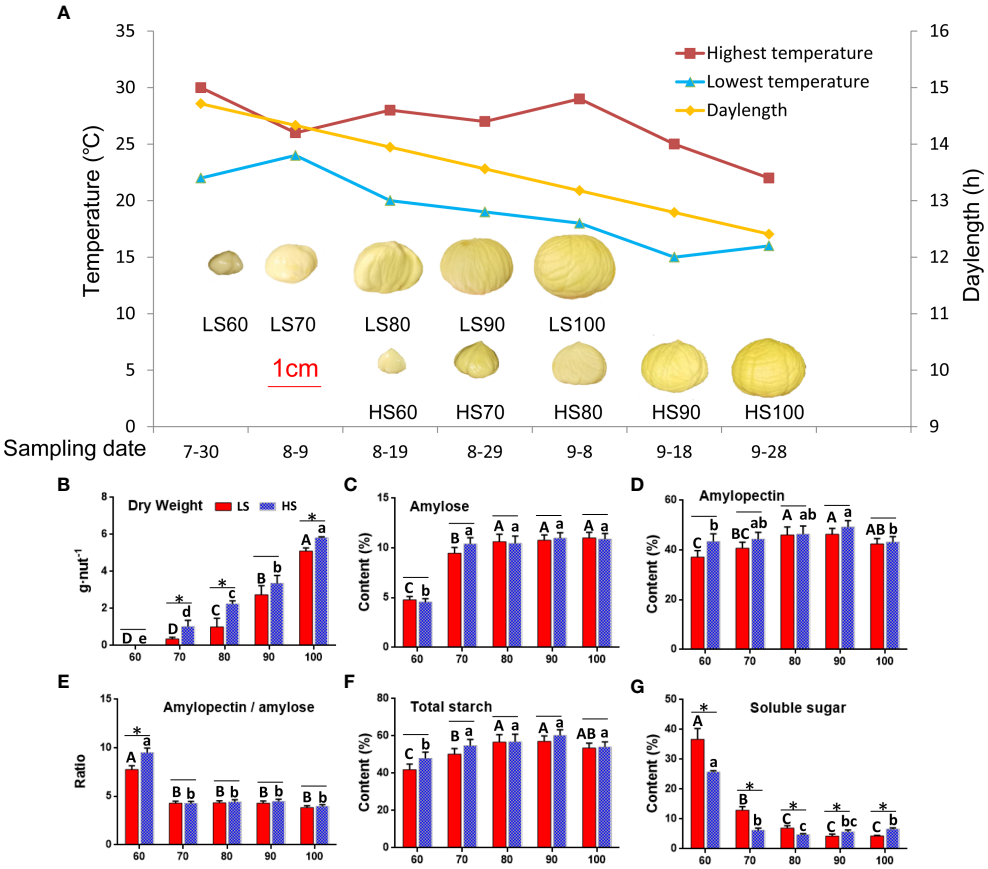


FIGURE 1
Morphological characteristics and sugar contents in two Chinese chestnut cultivars during embryo development: **(A)** morphological characteristics at 60–100 days after flowering (DAF). 'Yanshanzaofeng' (LS) has female flowers in full bloom on June 5, and 'Yanlong' (HS) has female flowers in full bloom on June 25. The ruler is 1 cm. **(B–G)** Dry weight and sugar contents in the developing embryo. Values are means ± standard deviation (SD), $n = 3$. The statistical significance between cultivars was determined by the Student's t test (*$p < 0.05$). The statistical significance between stages of the same cultivar was evaluated by one-way analysis of variance (ANOVA) with Duncan's multiple comparison test ($p < 0.05$), and capital letters indicate LS and Lowercase letters indicate HS. The same letters used within the same cultivar at different stages indicate no significant difference at the $p \geq 0.05$ level.

sucrose, glucose and fructose at 60-70 DAF (Figure 2A). In addition, the content of glucose (Figure 2D) and fructose (Figure 2E) was high (> 87.01 mg/g) at 60 DAF but low (<3.68 mg/g) at 80-100 DAF. In mature chestnuts, the soluble sugar was mainly sucrose (> 38.67 mg/g), and the content of other sugar metabolites was very low (<3.30 mg/g). At 90-100 DAF, the sucrose content in HS was 1.5 times higher than that in LS (Figure 2C). In addition, the content of raffinose (Figure 2F), 1,5-Anhydroglucitol (Figure 2H) and maltose (Figure 2I) was higher in HS than that in LS at mature stage. The content of inositol (Figure 2G) was higher in LS than that in HS at mature stage. During ripening, the increased content of these soluble sugar metabolites will make the chestnuts sweeter.

Transcriptome sequencing and clustering of DEGs

Thirty libraries from Chinese chestnut embryos were sequenced. We obtained a total of 1.41 billion base pairs, with an average of 47,036,059 raw reads and 45,312,612 clean reads per sample. The average ratio of clean reads to raw reads was 96.33% (Table S2). The clean reads are made freely available in the NCBI (accession number: PRJNA883560).

A total of 42,740 unigenes were identified from transcriptome data. Principal components analysis (PCA) of all 30 samples was conducted based on RNAseq FPKM, and two principal components were found to explain 39.9% of the overall variance (27.2% and 12.7% for PC1 and PC2, respectively). The three samples at the same stage were close to each other, indicating that there was a high consistency between the three biological replicates (Figure 3A). A total of 14,430 DEGs were identified by pair comparison between two cultivars at each stage (Figure 3B). The cluster analysis of DEGs was shown in Figure 3C. The results showed that the 30 samples could be divided into three groups: LS60 and HS60 constituted the first group, LS70, LS80, LS90, HS70 and HS80 constituted the second group, and LS100, HS90 and HS100 formed the third group.

Trend analysis of DEGs in embryo

In order to further understand the mechanism of sugar synthesis in Chinese chestnut embryos, we focused on the expression trend of

10.3389/fpls.2023.1206585

TABLE 1 Identification of sugar-related metabolites in the developing Chinese chestnut embryo.

Compounds	LS60	LS70	LS80	LS90	LS100	HS60	HS70	HS80	HS90	HS100
Sucrose	**160.41**	**116.97**	**44.06**	**34.92**	**38.67**	**149.94**	**47.94**	**38.53**	**41.57**	**57.44**
D-Fructose	**111.78**	**42.96**	3.68	0.24	0.07	**92.69**	3.68	0.43	0.07	0.06
Glucose	**106.99**	**47.53**	3.52	0.20	0.08	**87.01**	3.40	0.28	0.08	0.08
Inositol	6.19	4.41	2.20	1.47	2.22	6.36	2.12	1.44	1.35	1.22
Raffinose	2.93	0.67	0.42	0.73	2.29	2.48	0.23	0.58	4.03	3.30
D-Galactose	0.44	0.20	0.09	0.03	0.07	0.50	0.15	0.05	0.05	0.07
1,5-Anhydroglucitol	0.27	0.12	0.08	0.04	0.07	0.52	0.08	0.05	0.04	0.12
Maltose	0.11	0.10	0.20	0.02	0.08	0.14	0.03	0.03	0.03	0.11
D-Mannose	0.17	0.09	0.05	0.02	0.02	0.19	0.07	0.03	0.02	0.02
D-Mannose-6-phosphate sodium salt	0.05	0.03	0.03	0.03	0.03	0.05	0.03	0.03	0.03	0.03
D-Xylulose	0.09	0.05	0.01	0.01	0.01	0.11	0.01	0.01	0.01	0.01
Methyl beta-D-galactopyranoside	0.05	0.03	0.02	0.01	0.02	0.11	0.02	0.01	0.01	0.03
Barium D-ribose-5-phosphate	0.03	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02
Cellobiose	0.06	0.04	0.02	0.01	0.01	0.06	0.01	0.01	0.01	0.01
D-Sorbitol	0.03	0.06	0.02	0.00	0.00	0.09	0.01	0.00	0.00	0.00
2-Acetamido-2-deoxy-D-glucopyranose	0.03	0.02	0.02	0.02	0.01	0.04	0.02	0.02	0.02	0.02
D-Xylose	0.04	0.03	0.02	0.00	0.00	0.07	0.02	0.01	0.00	0.00
Lactose	0.03	0.02	0.01	0.01	0.02	0.02	0.00	0.00	0.00	0.00
L-Fucose	0.02	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01
D-Arabinose	0.03	0.01	0.01	0.00	0.00	0.04	0.01	0.00	0.00	0.00
Levoglucosan	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Trehalose	0.03	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
D-Glucuronic acid	0.01	0.01	0.01	0.00	0.00	0.02	0.01	0.00	0.00	0.00
Deoxyglucose	0.01	0.01	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01
D-Ribose	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00
D-Ribono-1,4-lactone	0.01	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.01
Xylitol	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.01
Phenylglucoside	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D-Galacturonic acid	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
2-Deoxy-D-ribose	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D-Arabinitol	–	–	–	–	–	–	–	–	–	–
L-Rhamnose	–	–	–	–	–	–	–	–	–	–

The metabolites with values > 100 are highlighted with bold font. A total of 32 kinds of sugar metabolites were tested, and two of them were not detected.

DEGs. The 14,430 DEGs were divided into five modules using WGCNA (Figure 4A), and the number of DEGs and KEGG pathways of each module were listed in each module. The largest module (turquoise) contained 6489 DEGs whose expression was highest in the LS60 sample, which included genes related to glycolysis/gluconeogenesis, brassinosteroid biosynthesis, and flavonoid biosynthesis. The second largest module (blue) contained 3484 DEGs whose expression was the highest expression in the LS100 sample, which included genes related to glycolysis/gluconeogenesis, glycerolipid metabolism, limonene and pinene degradation, and flavone and flavonol biosynthesis. The third largest module (brown) contained 1985 DEGs whose expression was the highest expression in the HS60 sample, which included genes related to photosynthesis, oxidative phosphorylation, nitrogen metabolism, tyrosine metabolism.

In order to understand the relationship between the gene expression patterns and metabolites, association analysis was

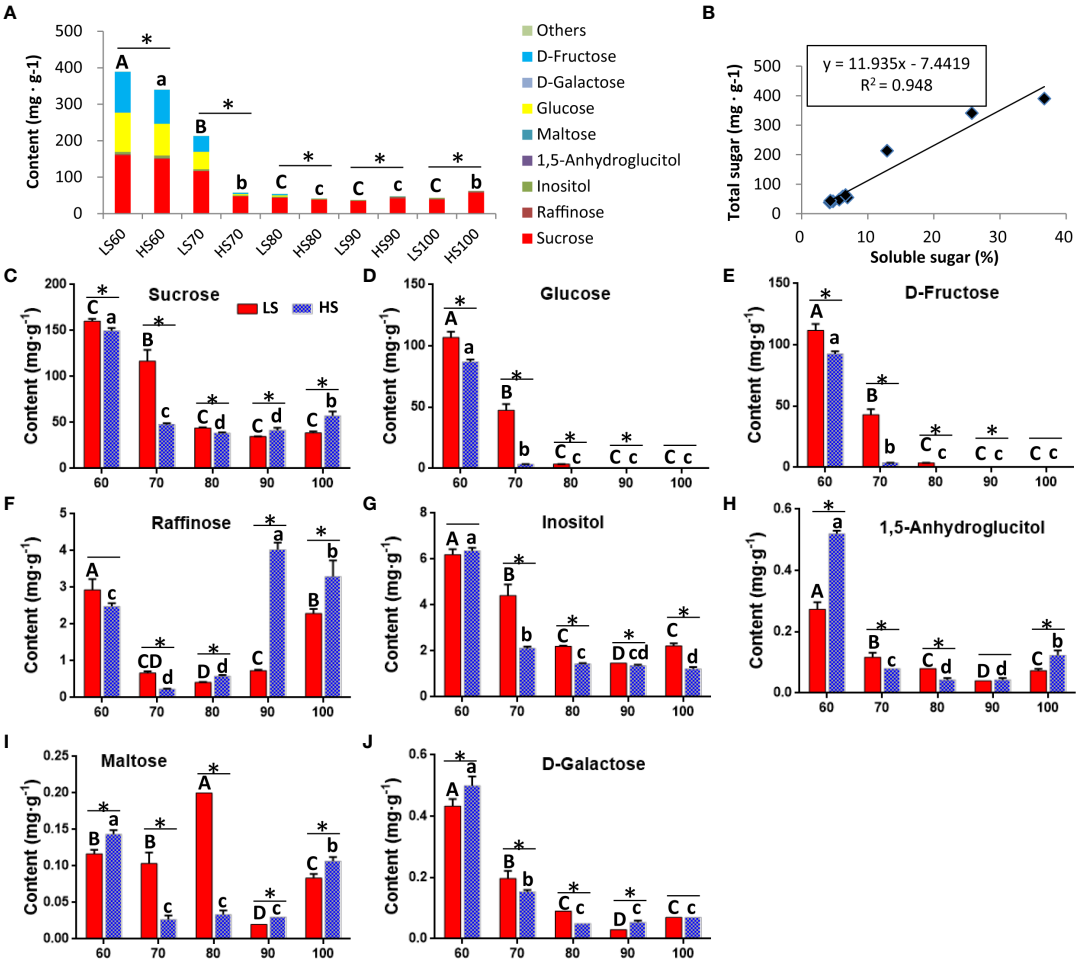


FIGURE 2
The changes of soluble sugar composition in two Chinese chestnut cultivars during embryo development: **(A)** The total content of all soluble sugar metabolites. **(B)** Correlation analysis between total content of all soluble sugar metabolites and soluble sugar content (Figure 1G). **(C–J)** The changes of the soluble sugar metabolites in the developing embryo. Values are means ± standard deviation (SD), $n = 3$. The statistical significance between cultivars was determined by the Student's t test (*$p < 0.05$). The statistical significance between stages of the same cultivar was evaluated by ANOVA with Duncan's multiple comparison test ($p < 0.05$), and capital letters indicate LS and Lowercase letters indicate HS. The same letters used within the same cultivar at different stages indicate no significant difference at the $p \geq 0.05$ level.

performed using WGCNA (Figure 4B). Sugar content was highly correlated with turquoise and brown modules. Combined with KEGG pathway and module-trait correlation analysis, turquoise, blue and brown modules were associated with sugar biosynthesis.

Identification of genes related to sucrose degradation and starch synthesis

The sucrose produced by photosynthesis will be stored in the chestnut as starch. As shown in Table S3, a total of 66 unigenes related to starch synthesis were identified, of which 15 belong to the turquoise module, 9 belong to the blue module, and 4 belongs to the other module. Most of the genes related to starch synthesis were highly expressed at 70 DAF. Some genes such as *SUS*, *GPT* and *GBSS* had FPKM values >1000 in our transcriptome data. These highly expressed genes may play an important role in starch synthesis in Chinese chestnut embryos. The variation trend of these genes related to starch synthesis was similar to that reported in previous articles (Zhang et al., 2015; Shi et al., 2021), indicating that the regulation of these genes in chestnut cultivars was similar.

At 70-80 DAF, the expression level of genes related to starch synthesis (i.e., *SUS*, *PGI*, *UGP*, *PGM* and *AGP*) was higher in LS than that in HS (Figure 5). At 100 DAF, the expression level of genes related to starch synthesis (i.e., *HXK*, *PGI*, *PGM*, *AGP*, *SS*, *SBE*, *ISA*) was higher in HS than that in LS (Figure 6). This indicates that the sugar metabolism of HS was more active than that of LS at mature stage.

Identification of genes related to starch degradation and sucrose synthesis

Starch degradation during ripening is a key additional process for sugar accumulation and sweetness in fruit (Liu et al., 2021). As shown in Table S4, a total of 33 unigenes related to starch degradation were identified, of which 6 belong to the turquoise module, 6 belong to the blue module, and 3 belongs to the other

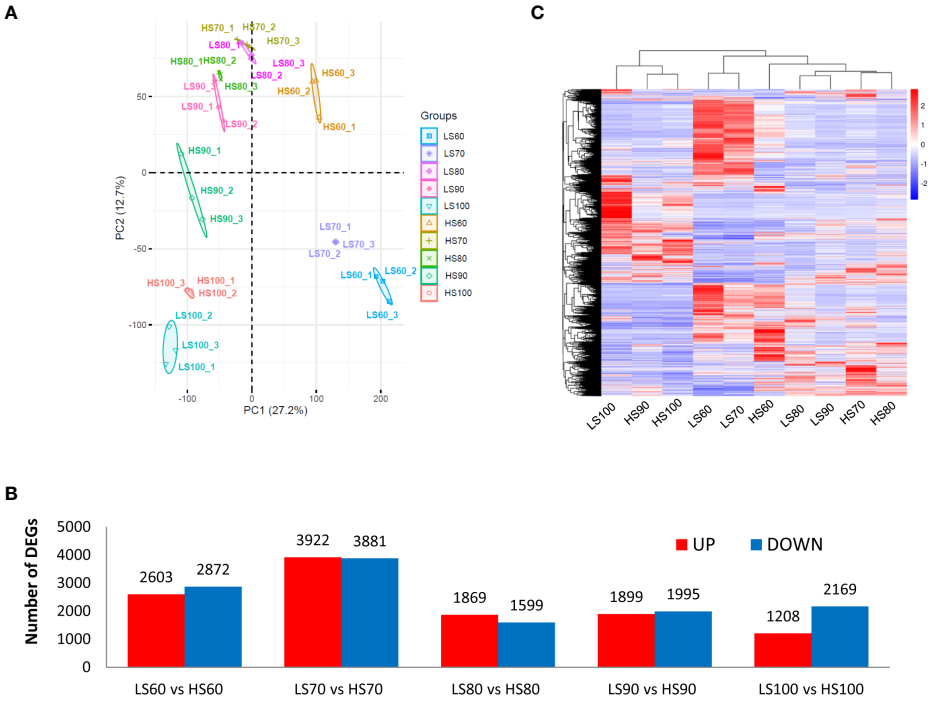


FIGURE 3
PCA of samples and cluster analysis of DEGs: **(A)** PCA of all 30 samples was conducted based on RNA-seq FPKM. **(B)** Pairwise comparison between two cultivars at same stage. We identified a total of 14,430 DEGs in Table S3. **(C)** Cluster analysis of DEGs based on mean FPKM. The color scale (–3 to 3) represents the calculated Z-score.

module. Some genes such as *PWD*, *BAM*, *PHS1*, *GLT*, *PHS2* had FPKM values >100 in our transcriptome data. These highly expressed genes may play an important role in starch degradation in Chinese chestnut embryos. At 100 DAF, the expression level of genes related to starch degradation (i.e., *GWD*, PWD, *LDA*, *ISA*, *PHS1*, *MEX1*) was higher in HS than that in LS (Figure 6). This indicated that starch degradation of HS was more intense than that of LS at mature stage, which was consistent with soluble sugar content.

It is worth noting, the expression level of *GWD* genes was higher in HS than that in LS at 70-100 DAF. GWD catalyzes the phosphorylation of amylopectin to form phosphoglucan, which is a key enzyme in the degradation of plant starch (You et al., 2020). In rice research, GWD can improve the yield and quality characteristics, which has great application potential (Wang et al., 2021). Therefore, GWD may play an important role in starch degradation of chestnut at mature stage.

Analysis of *SUS* gene

Sucrose synthase (SUS) is the key enzyme of sucrose metabolism, which has the function of decomposition and synthetic sucrose (Stein and Granot, 2019). In this study, we identified a total of 8 *SUSs* in Chinese chestnut, including three *SUS2*, one *SUS3*, two *SUS4*, and two *SUS6* genes. Among these 8 *SUSs*, 2, 4, and 2 genes belong to the SUSI subfamily, SUSII subfamily, SUSIII subfamily, respectively (Figure 7A). The *SUS2* (bl_022590, bl_022593 and bl_022594), *SUS3* (bl_024378), and *SUS4* (bl_040354) genes were highly expressed (FPKM> 100) (Table S3), which may play an important role in sucrose content during Chinese chestnut embryo development.

The *SUS6* gene were lowly expressed, but the *SUS2* were highly expressed in developing Chinese chestnut embryos. The expression heat map revealed that three *SUS2* genes (bl_022590, bl_022593 and bl_022594) could be grouped into a category, with the high expression levels observed at 60-90 DAF (Figure 7B). It is noteworthy that the expression level of *SUS3* (bl_024378) was higher in HS than that in LS at 90-100 DAF.

In addition, we measure the enzyme activity of SUS. The enzyme activity of SUS-cleavage continues to decrease during embryo development, and there is no significant difference between the two cultivars (Figure 7C). However, the enzyme activity of SUS-synthetic was significantly different between the two cultivars. The enzyme activity of SUS-synthetic was higher in LS than that in HS at 70-80 DAF, while was higher in HS than that in LS at 90-100 DAF. The enzyme activity of SUS changes were consistent with changes of sucrose content in the two cultivars.

One study showed that after 48 h cold stress treatment of *Arabidopsis thaliana* (Baslam et al., 2017), the expression levels of *AtSus2*, *AtSus4*, *AtSus5* and *AtSus6* were not affected, while the expression levels of *AtSus1* and *AtSus3* were 24.5 times and 5.5 times of those in the untreated group, respectively. The changes of these genes were consistent with our results, and cold stress may affect *SUS3* (bl_024378) gene expression, and then change the enzyme activity of SUS-synthetic in Chinese chestnut.

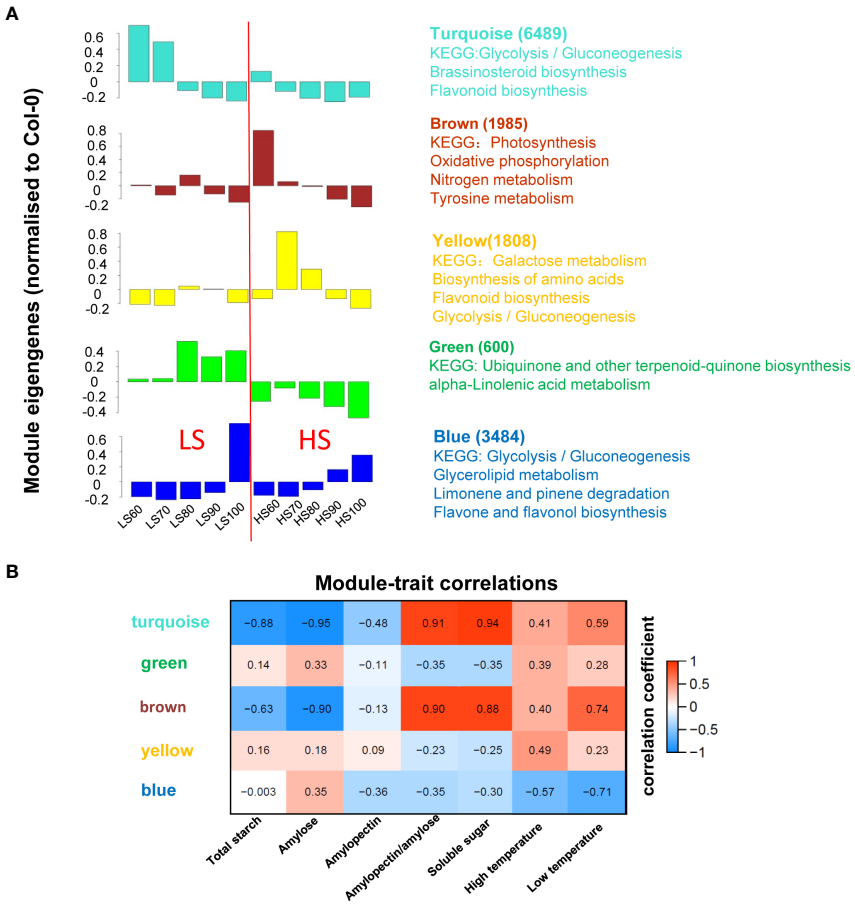


FIGURE 4
WGCNA of RNA-seq and traits: **(A)** WGCNA was calculated by 30 samples and the 14,430 DEGs were classified into 5 modules. Columns represent module eigengene of mean values. The number of DEGs and KEGG pathways for each module were listed; **(B)** expression patterns of the modules were correlated to physiological data. The numbers in each colored box give the values for correlation coefficient.

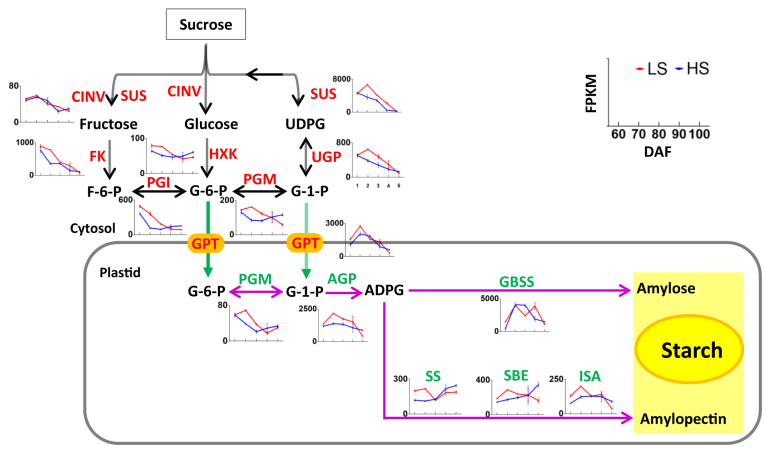


FIGURE 5
Nonpartitioned expression pattern of genes related to sucrose degradation and starch synthesis. Symbols represent mean FPKM values and the horizontal axis represents 60, 70, 80, 90, and 100 DAF. Red line represents LS, and blue line represents HS (high sugar cultivar). The two-way arrow indicates a reversible reaction, and the one-way arrow indicates an irreversible reaction. The information of genes are listed in Table S3.

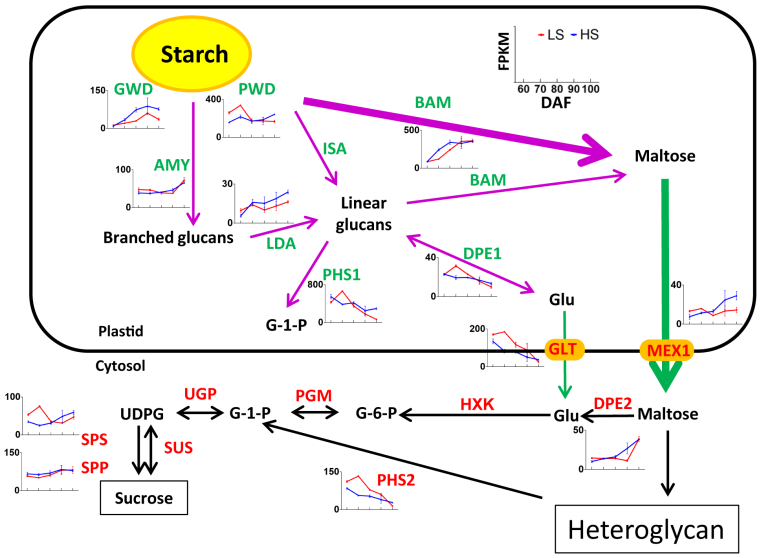


FIGURE 6
Nonpartitioned expression pattern of genes related to sucrose degradation and starch synthesis. Symbols represent mean FPKM values and the horizontal axis represents 60, 70, 80, 90, and 100 DAF. Red line represents LS, and blue line represents HS (high sugar cultivar). The two-way arrow indicates a reversible reaction, and the one-way arrow indicates an irreversible reaction. The information of genes are listed in Table S4.

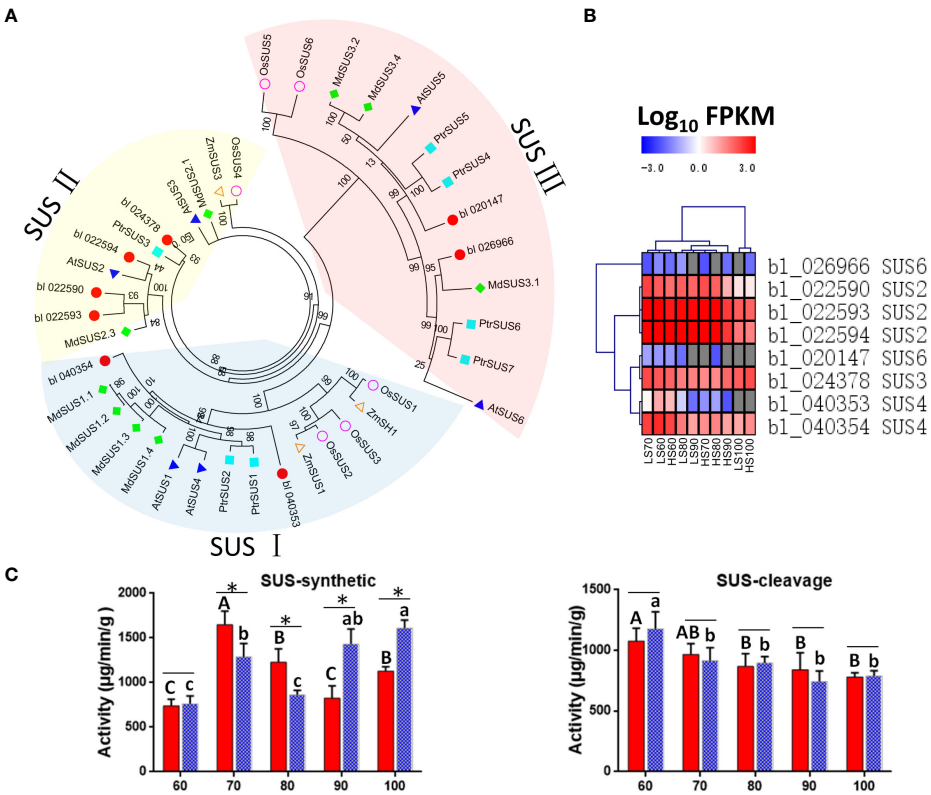


FIGURE 7
Putative sucrose synthase (SUS) unigenes identified in Chinese chestnut: **(A)** phylogenetic analysis of CmSUSs (*Castanea mollissima*) and AtSUSs (Arabidopsis thaliana) with Protein sequences. The Protein sequences were aligned using ClustalX2, and the neighbor-joining (NJ) tree was constructed with the program MEGA 6.0. A total of 1000 bootstrap replications were carried out to indicate reliability. **(B)** Heat map showing expression levels of *CmSUSs*. **(C)** Enzyme activity of SUS in two Chinese chestnut cultivars during embryo development. Values are means ± standard deviation (SD), n = 3. The statistical significance between cultivars was determined by the Student's t test (*$p < 0.05$). The statistical significance between stages of the same cultivar was evaluated by ANOVA with Duncan's multiple comparison test ($p < 0.05$), and capital letters indicate LS and Lowercase letters indicate HS. The same letters used within the same cultivar at different stages indicate no significant difference at the $p \geq 0.05$ level.

ABA promote the increase of soluble sugar

Some studies have shown that cold stress can promote the increase of soluble sugar in plants (Zhang et al., 2011; Ding et al., 2019) (Figure 8A). In addition, COR is a key gene in response to cold stress, which encoded hydrophilic peptide can protect the cells from freezing damage. Our data showed that the expression level of COR gene was higher in HS than that in LS at 60-100 DAF. Notably, *COR* (bl_030433) and *GWD* (bl_017818) were co-expressed, which may reflect the correlation between starch degradation and low temperature. Then, we analyzed the 412 genes co-expressed with *COR* (bl_030433) by string analysis (Figure 8B). There were 9 genes associated with ABA signaling pathway and 8 genes associated with peroxisome. ABA and peroxide may play an important role in starch decomposition during Chinese chestnut ripening.

Therefore, we also analyzed ABA content in two Chinese chestnut cultivars during embryo development (Figure 8C). At 90 DAF, ABA content in HS was higher than that in LS. Correspondingly, the expression level of genes related to ABA (i.e., *PYL*, *PP2C* and *SnRK2*) was also higher in HS than in LS (Figure 8A). In addition, the expression of *CSLA* gene related to polysaccharide synthesis was higher in HS than in LS at 80-100 DAF. Some studies have also shown that ABA plays an important role in promoting starch degradation and sucrose synthesis in fruits. Exogenous ABA treatment could promote mango ripening and increase soluble sugar (Wu et al., 2022).

qPCR analysis of sugar-related genes

In order to verify the relative expression pattern of unigenes in RNA-seq analysis, twelve key genes related to sugar metabolism were analyzed by qPCR (Panels A-L) (Figure 9). Panel M shows that RNA-seq is highly correlated with qPCR data (R^2 = 0.74, p <0.01), indicating that the expression data obtained by RNA-seq is reliable. Among all 12 genes, the expression level of 10 genes (*HXK*, *PGI*, *PGM*, *GPT*, *GBSS*, *SS*, *SBE*, *ISA*, *GWD* and *CLSC20*) was higher in HS than that in LS at 100 DAF, indicating that the active sugar metabolism was related to the increase in the content of soluble sugar in LS at the mature stage.

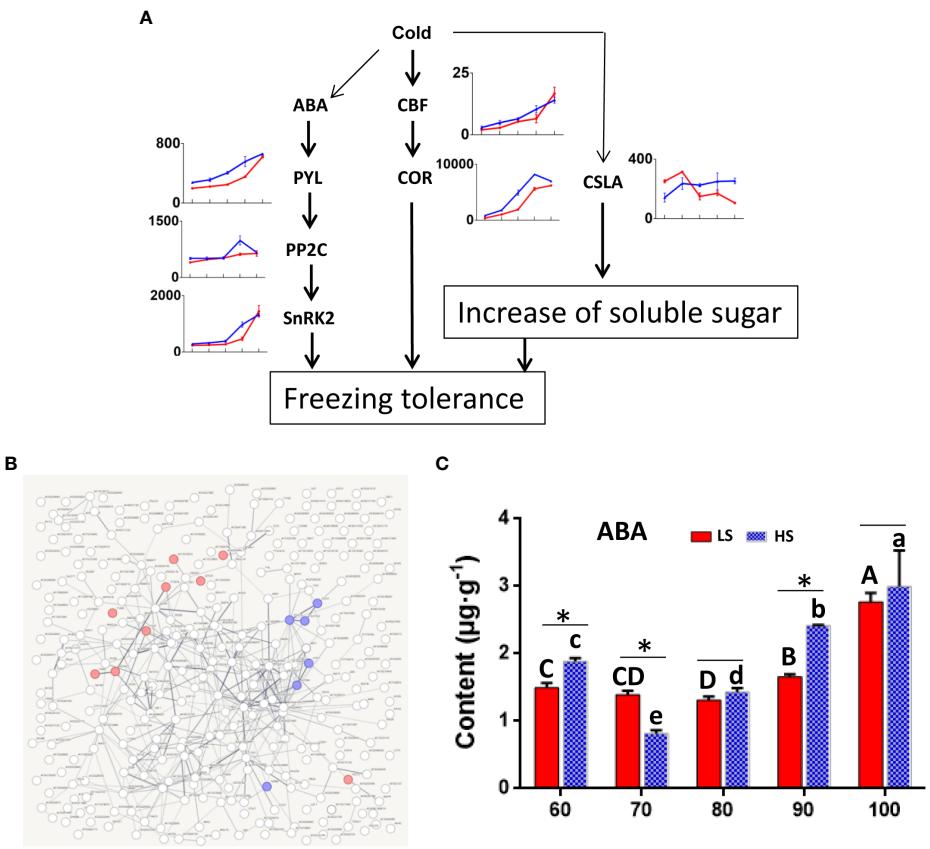


FIGURE 8
Effect of low temperature on soluble sugar in Chinese chestnut. **(A)** Identification of genes related to cold stress. Red line represents LS, and blue line represents HS (high sugar cultivar). The information of genes are listed in Table S5. **(B)** Co-expression network for *COR* (bl_030433). The information of genes are listed in Table S6. Red genes were associated with ABA signaling pathway and blue genes were associated with peroxisome. **(C)** ABA content in two Chinese chestnut cultivars during embryo development. ABA content was carried out by ELISA (Liu et al., 2008). Values are means ± standard deviation (SD), n = 3. The statistical significance between cultivars was determined by the Student's t test (*p < 0.05). The statistical significance between stages of the same cultivar was evaluated by ANOVA with Duncan's multiple comparison test (p < 0.05), and capital letters indicate LS and Lowercase letters indicate HS. The same letters used within the same cultivar at different stages indicate no significant difference at the p≥ 0.05 level.

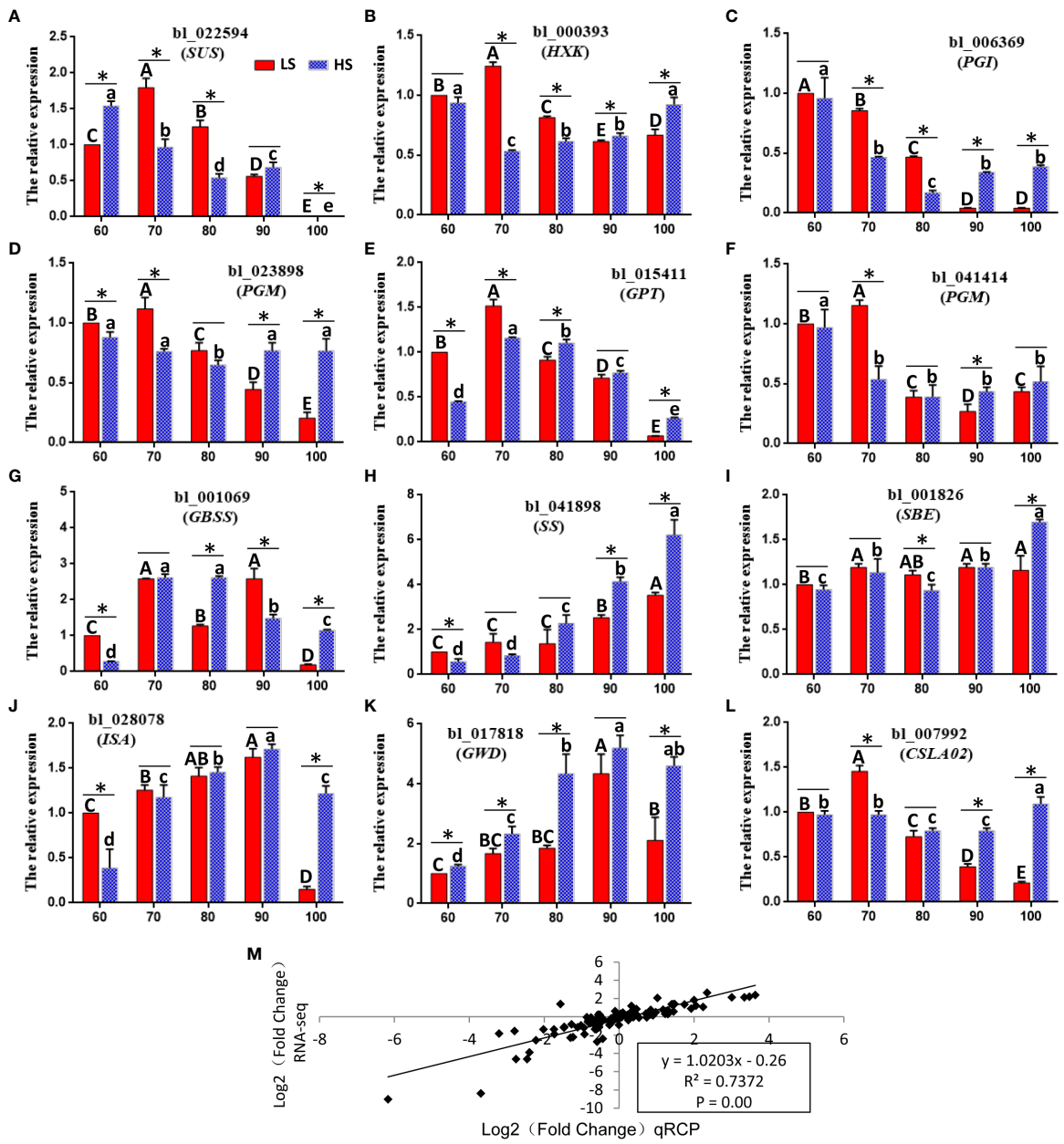


FIGURE 9
qPCR validation of differential expression. **(A–L)** qPCR of 12 important genes associated with sugar biosynthesis. Symbols represent mean values and short vertical lines indicate SE (n = 3). Same letters stand for statistically insignificance ($p > 0.05$). The unit of measure on the x-axis is sample. **(M)** A comparison of the gene expression ratios obtained from RNA-seq data and q-PCR. The statistical significance between cultivars was determined by the Student's t test (*$p < 0.05$). The statistical significance between treatments was evaluated by ANOVA with Duncan's multiple comparison test ($p < 0.05$). Capital letters indicate LS. Lowercase letters indicate HS. The same letters used within the same cultivar at different stages indicate no significant difference at the $p \geq 0.05$ level.

Conclusions

In this study, we investigated the molecular mechanism of the difference in sugar accumulation between two Chinese chestnut cultivars. Metabolome and transcriptome analysis showed that starch degradation was an important pathway for the increase of soluble sugar content in chestnut at mature stage. Furthermore, the enzyme activity of SUS-synthetic and ABA content were higher in HS than those in LS at 90 DAF. Based on our findings, we preliminarily established the regulation model of sucrose and starch conversion in chestnut (Figure 10). During chestnut development, sucrose is synthesized by photosynthesis and transported to the embryo, and sucrose is decomposed into fructose and glucose to synthesize starch. At the mature stage, ABA content in chestnut was increased. ABA further stimulated the expression of genes (i.e., *SUS* and *GWD*), and then promoted the decomposition of starch into sucrose, which increased the sweetness of chestnut kernels. Our study analyzed the composition and molecular synthesis mechanism of sugar in Chinese chestnut embryos, and provided a new insight into the regulation pattern

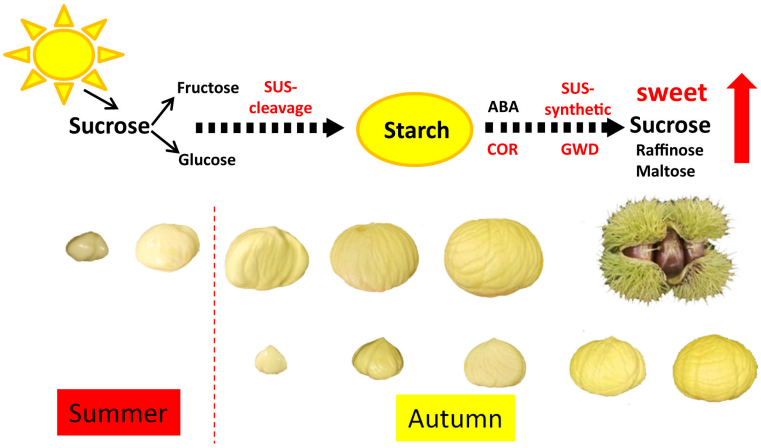


FIGURE 10
Regulation model of sucrose and starch conversion in Chinese chestnut. During Chinese chestnut development, sucrose is synthesized by photosynthesis and transported to the embryo, and sucrose is decomposed into fructose and glucose to synthesize starch. At the mature stage, ABA content in chestnut was increased. ABA further stimulated the expression of genes (i.e., *SUS* and *GWD*), and then promoted the decomposition of starch into sucrose, which increased the sweetness of chestnut kernels.

of high sugar accumulation in Chinese chestnut nuts. It is noteworthy that HS has typical late ripening behavior, while LS has late ripening behavior in this study. Whether mature behavior and low temperature are related to the sugar content of chestnuts deserve further research.

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

RH and YY designed the experiments. RH, FP, DW, FC, and CG carried out the experiments. LY, JZ, and YY analyzed the data. RH and FP drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Hebei Collaborative Innovation Center of Chestnut Industry (202202) and Scientific Research Foundation of Hebei Normal University of Science and Technology (2022YB024).

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.2023.1206585/full#supplementary-material

References

Baroja-Fernández, E., Muñoz, F. J., Montero, M., Etxeberria, E., Sesma, M. T., Ovecka, M., et al. (2009). Enhancing sucrose synthase activity in transgenic potato (Solanum tuberosum l.) tubers results in increased levels of starch, ADPglucose and UDPglucose and total yield. *Plant Cell Physiol.* 50 (9), 1651–1662. doi: 10.1093/pcp/pcp108

Baslam, M., Baroja-Fernández, E., Ricarte-Bermejo, A., Sánchez-López, Á.M., Aranjuelo, I., Bahaji, A., et al. (2017). Genetic and isotope ratio mass spectrometric evidence for the occurrence of starch degradation and cycling in illuminated arabidopsis leaves. *PloS One* 12 (2), e0171245. doi: 10.1371/journal.pone.0171245

Bieniawska, Z. (2007). Functional analysis of the sucrose synthase gene family in arabidopsis thaliana. *universität potsdam.*

Chen, T., Zhang, Z., Li, B., Qin, G., and Tian, S. (2021). Molecular basis for optimizing sugar metabolism and transport during fruit development. *Abiotech* 2 (3), 330–340. doi: 10.1007/s42994-021-00061-2

Ding, Y., Shi, Y., and Yang, S. (2019). Advances and challenges in uncovering cold tolerance regulatory mechanisms in plants. *New Phytol.* 222 (4), 1690–1704. doi: 10.1111/nph.15696

Gomez-Gonzalez, S., Ruiz-Jimenez, J., Priego-Capote, F., and Luque de Castro, M. (2010). Qualitative and quantitative sugar profiling in olive fruits, leaves, and stems by gas chromatography– tandem mass spectrometry (GC-MS/MS) after ultrasound-assisted leaching. *J. Agric. Food Chem.* 58 (23), 12292–12299. doi: 10.1021/jf102350s

Haytowitz, D., Ahuja, J., Wu, X., Khan, M., Somanchi, M., Nickle, M., et al. (2018). USDA National nutrient database for standard reference, legacy. *USDA Natl. Nutrient Database Standard Reference.*

He, L., Min, Q., Hong, C., and Zhang, Y. (2021). Features and socio-economic sustainability of traditional chestnut forestry landscape in China: a case of kuancheng county, hebei province. *Land* 10 (9), 952. doi: 10.3390/land10090952

Huang, L., Tan, H., Zhang, C., Li, Q., and Liu, Q. (2021). Starch biosynthesis in cereal endosperms: an updated review over the last decade. *Plant Commun.* 2 (5), 100237. doi: 10.1016/j.xplc.2021.100237

Juliano, B., Perez, C., Blakeney, A., Castillo, T., Kongseree, N., Laignelet, B., et al. (1981). International cooperative testing on the amylose content of milled rice. *Starch? St rke* 33 (5), 157–162. doi: 10.1002/star.19810330504

Künsch, U., Schärer, H., Patrian, B., Höhn, E., Conedera, M., Sassella, A., et al. (2001). Effects of roasting on chemical composition and quality of different chestnut (Castanea sativa mill) varieties. *J. Sci. Food Agric.* 81 (11), 1106–1112. doi: 10.1002/jsfa.916

LaBonte, N. R., Zhao, P., and Woeste, K. (2018). Signatures of selection in the genomes of Chinese chestnut (Castanea mollissima blume): the roots of nut tree domestication. *Front. Plant Sci.* 9, 810. doi: 10.3389/fpls.2018.00810

Liu, T., Hu, Y., and Li, X. (2008). Comparison of dynamic changes in endogenous hormones and sugars between abnormal and normal castanea mollissima. *Prog. Natural Sci.* 18 (6), 685–690. doi: 10.1016/j.pnsc.2008.03.001

Liu, J., Liu, M., Jia, C., Zhang, J., Miao, H., Wang, J., et al. (2021). Elucidating the mechanism of MaGWD1-mediated starch degradation cooperatively regulated by MaMADS36 and MaMADS55 in banana. *Postharvest Biol. Technol.* 179, 111587. doi: 10.1016/j.postharvbio.2021.111587

Lowell, C. A., Tomlinson, P. T., and Koch, K. E. (1989). Sucrose-metabolizing enzymes in transport tissues and adjacent sink structures in developing citrus fruit. *Plant Physiol.* 90 (4), 1394–1402. doi: 10.1104/pp.90.4.1394

Mahlow, S., Orzechowski, S., and Fettke, J. (2016). Starch phosphorylation: insights and perspectives. *Cell. Mol. Life Sci.* 73, 2753–2764. doi: 10.1007/s00018-016-2248-4

Maloney, V. J., Park, J.-Y., Unda, F., and Mansfield, S. D. (2015). Sucrose phosphate synthase and sucrose phosphate phosphatase interact in planta and promote plant growth and biomass accumulation. *J. Exp. Bot.* 66 (14), 4383–4394. doi: 10.1093/jxb/erv101

Massantini, R., Moscetti, R., and Frangipane, M. T. (2021). Evaluating progress of chestnut quality: a review of recent developments. *Trends Food Sci. Technol.* 113, 245–254. doi: 10.1016/j.tifs.2021.04.036

Medeiros, P. M., and Simoneit, B. R. (2007). Analysis of sugars in environmental samples by gas chromatography–mass spectrometry. *J. Chromatogr. A* 1141 (2), 271–278. doi: 10.1016/j.chroma.2006.12.017

Modi, A., Vai, S., Caramelli, D., and Lari, M. (2021). The illumina sequencing protocol and the NovaSeq 6000 system. *Methods Mol Biol.* 2242, 15–42. doi: 10.1007/978-1-0716-1099-2_2

Moshchenskaya, Y. L., Galibina, N., Novitskaya, L., and Nikerova, K. (2019). The role of sucrose synthase in sink organs of woody plants. *Russian J. Plant Physiol.* 66, 10–21. doi: 10.1134/S1021443719010114

Roggo, Y., Duponchel, L., Noe, B., and Huvenne, J.-P. (2002). Sucrose content determination of sugar beets by near infrared reflectance spectroscopy. comparison of calibration methods and calibration transfer. *J. Near Infrared Spectrosc.* 10 (2), 137–150. doi: 10.1255/jnirs.330

Samkumar, A., Karppinen, K., Dhakal, B., Martinussen, I., and Jaakola, L. (2022). Insights into sugar metabolism during bilberry (Vaccinium myrtillus l.) fruit development. *Physiologia Plantarum* 174 (2), e13657. doi: 10.1111/ppl.13657

Santos, M. J., Pinto, T., and Vilela, A. (2022). Sweet chestnut (Castanea sativa mill.) nutritional and phenolic composition interactions with chestnut flavor physiology. *Foods* 11 (24), 4052. doi: 10.3390/foods11244052

Schreier, T. B., Umhang, M., Lee, S.-K., Lue, W.-L., Shen, Z., Silver, D., et al. (2019). LIKE SEX4 1 acts as a β-amylase-binding scaffold on starch granules during starch degradation. *Plant Cell* 31 (9), 2169–2186. doi: 10.1105/tpc.19.00089

Shi, L., Wang, J., Liu, Y., Ma, C., Guo, S., Lin, S., et al. (2021). Transcriptome analysis of genes involved in starch biosynthesis in developing Chinese chestnut (Castanea mollissima blume) seed kernels. *Sci. Rep.* 11 (1), 1–13. doi: 10.1038/s41598-021-82130-6

Stein, O., and Granot, D. (2019). An overview of sucrose synthases in plants. *Front. Plant Sci.* 10, 95. doi: 10.3389/fpls.2019.00095

Streb, S., and Zeeman, S. C. (2012). Starch metabolism in arabidopsis. *Arabidopsis book/American Soc. Plant Biologists* 10. doi: 10.1199/tab.0160

Sun, Y., Lu, Z., Zhu, X., and Ma, H. (2020). Genomic basis of homoploid hybrid speciation within chestnut trees. *Nat. Commun.* 11 (1). doi: 10.1038/s41467-020-17111-w

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47 (D1), D607–D613. doi: 10.1093/nar/gky1131

Wan, H., Wu, L., Yang, Y., Zhou, G., and Ruan, Y.-L. (2018). Evolution of sucrose metabolism: the dichotomy of invertases and beyond. *Trends Plant Sci.* 23 (2), 163–177. doi: 10.1016/j.tplants.2017.11.001

Wang, Z., Wei, K., Xiong, M., Wang, J. D., Zhang, C. Q., Fan, X. L., et al. (2021). Glucan, water-dikinase 1 (GWD1), an ideal biotechnological target for potential improving yield and quality in rice. *Plant Biotechnol. J.* 19 (12), 2606–2618. doi: 10.1111/pbi.13686

Wang, X.-Q., Zheng, L.-L., Lin, H., Yu, F., Sun, L.-H., and Li, L.-M. (2017). Grape hexokinases are involved in the expression regulation of sucrose synthase-and cell wall invertase-encoding genes by glucose and ABA. *Plant Mol. Biol.* 94, 61–78. doi: 10.1007/s11103-017-0593-9

Warmund, M. R. (2011). Chinese Chestnut (Castanea mollissima) as a niche crop in the central region of the united states. *HortScience* 46 (3), 345–347. doi: 10.21273/HORTSCI.46.3.345

Wu, S., Wu, D., Song, J., Zhang, Y., Tan, Q., Yang, T., et al. (2022). Metabolomic and transcriptomic analyses reveal new insights into the role of abscisic acid in modulating mango fruit ripening. *Horticulture Res.* 9, uhac102. doi: 10.1093/hr/uhac102

Xiao, Y.y., Kuang, J.f., Qi, X.n., Ye, Y.j., Wu, Z. X., Chen, J.y., et al. (2018). A comprehensive investigation of starch degradation process and identification of a transcriptional activator mab HLH 6 during banana fruit ripening. *Plant Biotechnol. J.* 16 (1), 151–164. doi: 10.1111/pbi.12756

Xu, X., Dees, D., Dechesne, A., Huang, X.-F., Visser, R. G., and Trindade, L. M. (2017). Starch phosphorylation plays an important role in starch biosynthesis. *Carbohydr. polymers* 157, 1628–1637. doi: 10.1016/j.carbpol.2016.11.043

Yang, Q., Ding, J., Feng, X., Zhong, X., Lan, J., Tang, H., et al. (2022). Editing of the starch synthase IIa gene led to transcriptomic and metabolomic changes and high amylose starch in barley. *Carbohydr. Polymers* 285, 119238. doi: 10.1016/j.carbpol.2022.119238

You, Y., Zhang, M., Yang, W., Li, C., Liu, Y., Li, C., et al. (2020). Starch phosphorylation and the *in vivo* regulation of starch metabolism and characteristics. *Int. J. Biol. Macromolecules* 159, 823–831. doi: 10.1016/j.ijbiomac.2020.05.156

Zhang, H., Hou, J., Liu, J., Xie, C., and Song, B. (2014). Amylase analysis in potato starch degradation during cold storage and sprouting. *Potato Res.* 57, 47–58. doi: 10.1007/s11540-014-9252-6

Zhang, L., Lin, Q., Feng, Y., Fan, X., Zou, F., Yuan, D.-Y., et al. (2015). Transcriptomic identification and expression of starch and sucrose metabolism genes in the seeds of Chinese chestnut (Castanea mollissima). *J. Agric. Food Chem.* 63 (3), 929–942. doi: 10.1021/jf505247d

Zhang, Y., Xie, Z., Wang, Y., and An, L. (2011). Changes in carbohydrate metabolism and bulb growth as induced by low-temperature release of dormancy in lily bulbs. *Philippine Agric. Scientist* 94 (2), 149–154. doi: 10.1016/j.njas.2011.03.001

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 June 2023
DOI 10.3389/fpls.2023.1210134

Check for updates

OPEN ACCESS

EDITED BY
Qian-Hao Zhu,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

REVIEWED BY
Yheni Dwiningsih,
University of Arkansas, United States
Guangyan Feng,
Sichuan Agricultural University, China

*CORRESPONDENCE
Lan Wang
wanglan@scau.edu.cn
Guoying Yin
yinyinghygy6239@126.com
Xiangdong Liu
xdliu@scau.edu.cn

†These authors have contributed equally to this work

RECEIVED 21 April 2023
ACCEPTED 30 May 2023
PUBLISHED 20 June 2023

CITATION
Xian L, Tian J, Long Y, Ma H, Tian M, Liu X, Yin G and Wang L (2023) Metabolomics and transcriptomics analyses provide new insights into the nutritional quality during the endosperm development of different ploidy rice.
Front. Plant Sci. 14:1210134.
doi: 10.3389/fpls.2023.1210134

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

Metabolomics and transcriptomics analyses provide new insights into the nutritional quality during the endosperm development of different ploidy rice

Lin Xian[1,2,3†], Jiaqi Tian[1†], Yanxi Long[1], Huijin Ma[1], Min Tian[1], Xiangdong Liu[1,4,5*], Guoying Yin[1,2*] and Lan Wang[1,4*]

[1]College of Agriculture, South China Agricultural University, Guangzhou, China, [2]Guizhou Academy of Tobacco Science, Guiyang, China, [3]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, China, [4]Guangdong Provincial Key Laboratory of Plant Molecular Breeding, College of Agriculture, South China Agricultural University, Guangzhou, China, [5]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China

Autotetraploid rice is developed from diploid rice by doubling the chromosomes, leading to higher nutritional quality. Nevertheless, there is little information about the abundances of different metabolites and their changes during endosperm development in autotetraploid rice. In this research, two different kinds of rice, autotetraploid rice (AJNT-4x) and diploid rice (AJNT-2x), were subjected to experiments at various time points during endosperm development. A total of 422 differential metabolites, were identified by applying a widely used metabolomics technique based on LC-MS/MS. KEGG classification and enrichment analysis showed the differences in metabolites were primarily related to biosynthesis of secondary metabolites, microbial metabolism in diverse environments, biosynthesis of cofactors, and so on. Twenty common differential metabolites were found at three developmental stages of 10, 15 and 20 DAFs, which were considered the key metabolites. To identify the regulatory genes of metabolites, the experimental material was subjected to transcriptome sequencing. The DEGs were mainly enriched in starch and sucrose metabolism at 10 DAF, and in ribosome and biosynthesis of amino acids at 15 DAF, and in biosynthesis of secondary metabolites at 20 DAF. The numbers of enriched pathways and the DEGs gradually increased with endosperm development of rice. The related metabolic pathways of rice nutritional quality are cysteine and methionine metabolism, tryptophan metabolism, lysine biosynthesis and histidine metabolism, and so on. The expression level of the genes regulating lysine content was higher in AJNT-4x than in AJNT-2x. By applying CRISPR/Cas9 gene-editing technology, we identified two novel genes, *OsLC4* and *OsLC3*, negatively regulated lysine content. These findings offer novel insight into dynamic metabolites and genes expression variations during endosperm development of different ploidy rice, which will aid in the creation of rice varieties with better grain nutritional quality.

KEYWORDS

rice, different ploidy, metabolome, transcriptome, nutritional quality, lysine level

Introduction

Rice (*Oryza sativa* L.) is a food crop that is commonly cultivated worldwide, and China is the world's largest rice producer, with a total output of nearly 212 million tons in 2022. The current situation of stable rice yield and increased rice yield has resulted from three revolutions, namely dwarf rice breeding, hybrid rice breeding and breeding rice with ultra-high yield. While exploring the improvement of rice yield, the nutritional quality and stress resistance of rice must also be increased. In nature, polyploids are extensively present (Van de Peer et al., 2017; Qu et al., 2023). Most angiosperms, most eukaryotes, and crops such as rice have all undergone polyploidization during the course of evolution (Xiong et al., 2022; He et al., 2023). Some beneficial agronomic characteristics of polyploid rice include large grain size, high weight per thousand grains, a strong stem, and high stress tolerance (Guo et al., 2017; He et al., 2022; Liu et al., 2022; Wang et al., 2022c). Unfortunately, various constraints, including low seed-setting rates, have limited the production and use of such rice (Lu et al., 2020). To overcome this limitation, neo-tetraploid rice lines with greater than 80% seed-setting rates and highly fertile tetraploid rice lines with polyploid meiotic stability (PMeS) genes have been developed (Guo et al., 2017; Ku et al., 2022; Wang et al., 2022c). In addition, our research team developed a new type of tetraploid rice (known as 'neo-tetraploid rice'), including Huaduo1 to Huaduo 5 and Huaduo 8, which showed normal fertility and strong yield heterosis when crossed with low fertility autotetraploid rice lines (Kamara et al., 2021; Yu et al., 2021a; Chen et al., 2022). Following that, a number of academic and practical studies on high-fertility tetraploid rice were conducted, which encouraged the use of polyploid rice (Ghaleb et al., 2020; Yu et al., 2020; Chen et al., 2021; Ku et al., 2022). In addition to causing gigantism, which increases biomass output, polyploidization also alters the nutritional quality of the organism; for instance, the levels of carbohydrates, proteins, vitamins, and alkaloids typically increase (Sabooni and Gharaghani, 2022). However, rice loses some of its nutritious value as a result of polyploidization. When compared to diploid rice, autotetraploid rice has a protein content that is approximately 30% higher, an amino acid content that is 20–30% higher, and an amylose content that is approximately 12% lower (Gan et al., 2021; Singh et al., 2021). According to a recent study, polyploidization affects the biosynthesis, transport, and deposition of glutelin, which increase the glutelin concentration in rice seeds (Gan et al., 2021). These modifications have made rice more nutrient and flavor dense. In the past, conventional chemical techniques have typically been used to examine the changes in nutrients following polyploidization. Nonetheless, metabolomics is quickly becoming recognized as a crucial analytical approach in nutritional studies due to the essential role that metabolism plays in nutrition (Ulaszewska et al., 2019; Shen et al., 2023). By qualitatively and quantitatively examining all of an organism's metabolites, a technique known as "metabolomics" can be used to investigate dynamic changes in metabolites as well as accumulation patterns and the genetic bases of plant metabolites (Shen et al., 2023; Wang et al., 2023). A revolutionary method called widely targeted metabolomics analysis allows for the simultaneous quantification of over 1000 known and unknown compounds in addition to hundreds of recognized metabolites (Wang et al., 2018; Li and Song, 2019; Waris et al., 2022; Xue et al., 2022). The nutrient analysis of rice is now being performed increasingly with metabonomics. There were 121 metabolites in mature seeds of *indica* and *japonica* rice, and there were significant differences in their relative quantities (Hu et al., 2014). These findings close the gap between the genome and phenome and make it easier to identify the genetic controls over metabolic traits that can be used as a foundation for the future improvement of rice quality through metabolic engineering. They also provide significant insights into metabolic adaptation in rice subgroups. In addition, Wang et al. (2021) used liquid chromatography-tandem mass spectrometry to compare the metabolites of a group of diploid-tetraploid japonica brown rice and a group of diploid-tetraploid indica brown rice. A total of 401 metabolites were found to differ between the two diploid-tetraploid groups, 180 of which showed opposing expression trends while 221 displayed the same trends. This research offers a foundation for employing polyploidization to alter the nutritional content of rice and serves as a novel theoretical guide for developing nutrient-rich rice cultivars.

However, there have been few studies comparing the metabolites and gene expression of diploid and autotetraploid rice, which has led to limited knowledge about endosperm development after flowering. In our previous, we developed an autotetraploid rice variety, AJNT-4x, with an obviously high seed setting rate of 49.94% by self-crossing for 52 generations. We used iTRAQ-based quantitative glutelin proteomic analysis to identify the physiological metabolism process during endosperm development in terms of protein (Xian et al., 2021). Nevertheless, metabolomics studies on autotetraploid rice are not well developed. AJNT-4x with higher seed setting was studied further at 10, 15, and 20 DAF to reveal the molecular mechanism by coexpression analysis of metabolomics and transcriptomics data. In this research, AJNT-4x and its diploid counterpart AJNT-2x were further researched to identify differentially expressed genes (DEGs) and differential metabolites in different developmental stages. The combination of these phenotypic and molecular findings advances knowledge of biological processes, particularly the control of metabolism and gene expression during the growth and maturation of rice endosperm. Additionally, more studies will be performed on DEGs to clarify the molecular mechanism underlying metabolite accumulation in autotetraploid rice.

Materials and methods

Plant materials

On the farm of South China Agricultural University, rice lines (*O. sativa* L. sp. Aijiaonante (AJNT)-2x and AJNT-4x after self-crossing for 54 generations) were produced in the early and late seasons of 2021. Each of the test plots had ten rows planted with a row spacing of 18 to 21 cm. The grains were marked at the beginning of blooming, collected 10, 15, and 20 days after flowering (DAF), and then kept in a freezer at -80°C.

Metabolite extraction

A total of 100 mg of the sample were weighed into a 2 mL EP tube, 1 mL of 70% methanol aqueous solution and two small steel balls were added, the sample was processed in tissue grinder (50 Hz, 5 min), and the product was refrigerated at 4 degrees overnight (more than 8 h). The samples were removed the following day and centrifuged for 10 minutes at 4°C and 13000 rpm with the supernatant collected, and then the filtered samples were put in a sample vial for LC-MS analysis.

Ultra-performance liquid chromatography-multi reaction monitor quantitative analysis

The separation and quantitative measurement of metabolites were carried out using a tandem QTRAP6500 Plus high-sensitivity mass spectrometer (SCIEX, USA) and UPLC instrument (Waters, USA). A Waters UPLC HSS T3 (model: 1.8 um*2.1*100 mm) is the chromatography column was utilized. The mobile phases were an aqueous solution containing 0.1% formic acid (solution A) and 100% acetonitrile containing 0.1% formic acid (solution B). The following gradient was used for elution: 0-2 min, 5% solution B; 2-22 min, 5%-95% solution B; 22-27 min, 95% solution B; and 27.1-30 min, 5% solution B. The flow rate was 0.3 mL/min, and the column temperature was 40°C. For the QTRAP 6500 Plus equipped with the EST Turbo Metabolite Spray Interface, the metabolite source parameters were set as follows: metabolite source temperature, 500°C; metabolite spray voltage (IS), 5500 V (positive mode) or -5500 V (negative mode)); and metabolite source gas I (GS1), gas II (GS 2) and curtain gas (CUR) set to 40, 40 and 25 psi, respectively. The MRM technique was carried out in MRM mode, which included the MRM parent-daughter metabolite pair information, collision energy (CE) and declustering potential (DP) and RT (Retention time) of the target metabolite.

According to the RT of metabolite detection, the metabolic logistics intensity of metabolite target separation detection. Each color-coded peak of the mass spectrum represents one metabolite detected. The characteristic metabolites of each substance were screened by the triple four-pole, and the signal response intensity of the characteristic metabolites was obtained in the mass spectrometer. The peak Area of each chromatographic peak was calculated by integrating the mass spectrometer data with MultiQuant software. The peak area of each chromatographic peak represented the relative content of the corresponding substance. Finally, the peak areas of all chromatographic peaks were derived for subsequent statistical analysis.

Transcriptome sequencing and analysis

A TRIzol Reagent kit (Invitrogen, Carlsbad, CA) was used to extract total RNA. The concentration and purity were determined by both spectrophotometry (A260, A260/280 respectively. The manufacturer's protocols were followed while creating the cDNA libraries, which constructed from 500 ng of HMW DNA and fragmented using Covaris sonicator E220, and sheared DNA underwent End repair. The adapters kit instructions to make the adapters ligation and clean it with the supplied DNA Clean Beads. PCR amplification was used for purified adapter-ligated DNA and cleaned-up again by magnetic beads. Then, quality control is carried out, purified PCR products were denaturated and ligated to generate single-strand circular DNA libraries and the DNBSEQ system was utilized to sequence the transcriptomes. Adaptor sequences, low-quality reads, and possible contaminants from chloroplast, mitochondrion, and ribosomal DNA were initially removed from the raw reads.

To identify exons and splice junctions, the clean reads were next compared with the *Oryza sativa* L.spp. *japonica* genomic sequence using TopHat. The expression levels of matching genes in each cDNA collection were calculated and normalized to fragments per kilobase of exon per million fragments mapped. For the analysis of hierarchical gene clusters, Cluster 3.0 was used (de Hoon et al., 2004). Using EBSeq, DEGs were detected in several samples (Leng et al., 2013).

Bioinformatics analysis

Multivariate statistical analysis and univariate analysis were used to screen different metabolites between groups. Change multiple analysis and T test were performed on the data. Fold change (FC) was obtained by variance analysis, and p value was obtained by T-test, fold change>1.2 or<0.83, p value<0.05.

The three biological replicates of the 6 groups of rice samples were analyzed with the DESeq R program (1.18.0) in transcriptome analysis (Anders and Huber, 2010). DEGs were defined as genes with a p <0.05 and an absolute value of log2 (fold change) >1. To find enriched Gene Ontology (GO) categories, the Blast2GO package was applied (Conesa et al., 2005). DEGs were considered substantially enriched for GO keywords with a corrected p <0.05. KOBAS software was used to test for the statistical enrichment of DEGs among the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Mao et al., 2005). Using the R package, a coexpression network based on weighted gene coexpression network analysis (WGCNA) was produced (Langfelder and Horvath, 2008), namely, there were four steps: calculation of correlation coefficient between genes, determination of gene modules, co-expression network and correlation between modules and traits, median absolute deviation (MAD) was screened by FPKM value of all genes and the top 75%.

Gene function analysis by clustered, regularly interspaced, short palindromic repeat gene editing

A CRISPR/Cas9 gene editing construct of *OsLC3* and *OsLC4* was designed as previously described (Wang et al., 2022a). Two targets were designed for *OsLC3* and *OsLC4 via* the CRISPR-GE

website (http://skl.scau.edu.cn/). Several rounds of PCR were used to create the expression cassette vector. By means of overlapping PCR, the target sequences were added to the sgRNA expression cassettes. The pYLCRISPR/Cas9 vector based on the Golden Gate system was used to insert the purified PCR products (sgRNA expression cassettes). *Escherichia coli* DH5a competent cells were transformed directly using the ligated products containing the sgRNA expression cassettes. *Agrobacterium tumefaciens* strain EHA105 was used to successfully create CRISPR/Cas9 constructs. Thereafter, investigations using A. tumefaciens-mediated gene transfer were conducted on AJNT-2x backgrounds.

Lysine level analysis

The content of 13 amino acids (including lysine) of grains in AJNT-2x and AJNT-4x was determined using the fully automatic amino acid analyzer of German Sykam (S-433D), referring to the method of Zhou et al. (2016).

The shells and embryos of rice seeds are removed, and ground into powder in a mortar, and then collected through a pore size of 250 mesh. 0.1g of the sample was added in a hydrolysis tube, then added 12 mL HCl (6M), connecting to a vacuum pump and vacuum it. Hydrolyze at 110°C for 22 hours. After cooling, the hydrolysate was poured into a volumetric flask, and finally was diluted to 30 mL with ddH_2O. 1 mL of the fixed solution was taken and evaporated to dryness in a 60°C water bath, and then 1 mL of the fixed solution was taken again and evaporated to dryness in a 60°C water bath. Add 3 mL of diluent to dissolve, and then which was filtered through a tetrafluoroethylene membrane, and finally the filter liquor was placed in a brown vial for analysis.

Results

Detection of metabolites according to LC–MS/MS

The endosperm of different ploidy rice at 10, 15, and 20 DAF development stages was analyzed by using metabolomics and transcriptomics data (Figure 1A). The MRM quantitative software MultiQuant (SCIEX, USA) was used in combination with the extensive targeted metabolic standard database (BGI-Wide Target-Library) independently established by BGI to identify and quantitatively analyze metabolites. The compounds that were found in the samples are represented by the mass spectrometry peaks, each of which represents one metabolite.

QC sample detection curve overlap was extremely high, demonstrating the strong repeatability and dependability of the mass spectrometry data (Figures S1A, B). A total of 422 metabolites were detected in the experiment (Table S1), including 46 different types of class substances, such as flavonoids, terpenoids, amino acids, peptides and analogs, alkaloids, phenols and derivatives and other metabolites. Among them, flavonoids (21.80%), terpenoids (10.66%), amino acids, peptides and analogs (8.29%), and alkaloids (7.82%) were the most abundant (Figure 1B). According to the results of the principal component analysis (PCA), the first principal component (pca1) distinguished between the various treatment groups and explained 37.19% of the variation; the second principal component (pca2) distinguished between the two varieties and explained 21.86% (Figure 1C). Our findings demonstrate that the metabolomic profiles of AJNT-4x and AJNT-2x at various phases of development were distinct from one another, and the triplicates of the same sample converged together, demonstrating the high dependability and good reproducibility of the data acquired.

Metabolic differences between AJNT-4x and AJNT-2x at different developmental stages

Hierarchical cluster analysis (HCA) was used to examine the accumulation of metabolites between AJNT-4x and AJNT-2x at various developmental phases (Figure 2A). There were noticeable differences in abundance between AJNT-4x and AJNT-2x at different developmental stages with regard to the accumulation of these metabolites. Analysis of the differences between AJNT-4x and AJNT-2x showed that the abundances of 189, 206 and 203 metabolites in AJNT-4x were higher than those in AJNT-2x at 10, 15, and 20 DAF, respectively (Figure 2B), and 89 kinds of metabolites in AJNT-4x were more abundant than those in AJNT-2x in the three developmental stages. Similarly, the levels of 233, 216 and 219 metabolites in AJNT-4x were lower than those in AJNT-2x at 10, 15, and 20 DAF, respectively. Moreover, 103 kinds of metabolites in AJNT-4x were lower in abundance than those in AJNT-2x in the three developmental stages (Figure 2C). The differentially expressed metabolites were examined using the criteria of expression fold changes more than 1.2 with a P value<0.05 (fold change>1.2 or<0.83, P value<0.05) to accurately screen differential metabolites. AJNT-4x and AJNT-2x were compared at 10, 15, and 20 DAF, and 113, 111, and 165 metabolites were discovered (fold change>1.2 or<0.83, P value<0.05), respectively, and 20 metabolites were found in the three developmental stages, which can be considered the key metabolites in different developmental stages (Figure 2D).

According to the results of the KEGG classification and enrichment analysis, the differences in metabolites between the compared groups were primarily related to biosynthesis of secondary metabolites, microbial metabolism in diverse environments, biosynthesis of cofactors, ABC transporters, protein digestion and absorption and biosynthesis of alkaloids derived from the shikimate pathway.

Analysis of differentially expressed genes between AJNT-4x and AJNT-2x in different developmental stages

Comparative analysis of AJNT-4x and AJNT-2x at 10, 15 and 20 DAF was conducted, which revealed that 1321, 1369, and 2597 genes were upregulated in AJNT-4x, and 1732, 3273, and 2970

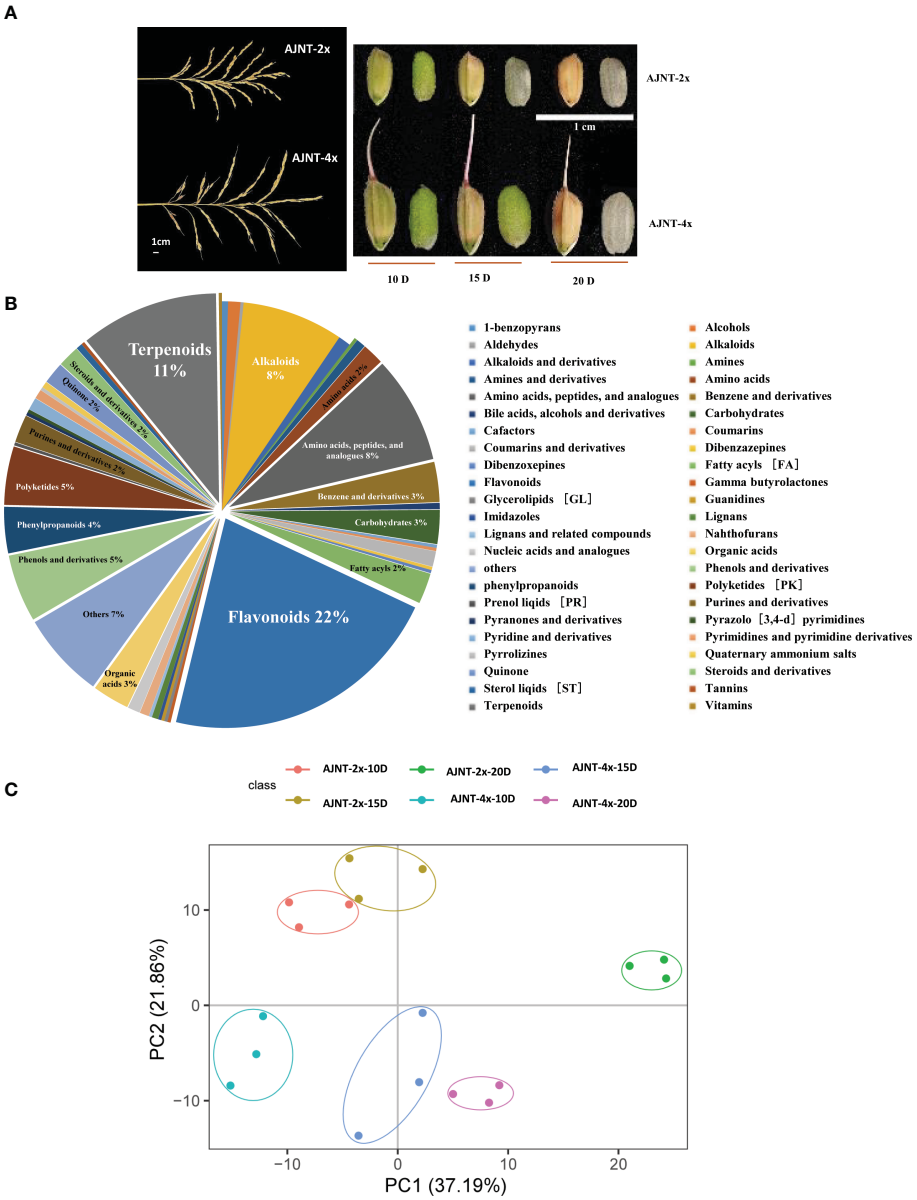


FIGURE 1
Analysis of metabolites according to LC-MS/MS. **(A)** The spike shape and the grain size of 10 DAF, 15 DAF, and 20 DAF for AJNT-2x and AJNT-4x. **(B)** Classification of the 422 metabolites. **(C)** PCA. The abscissa and the ordinate represent the scores of pc1 and pc2, respectively.

genes were upregulated in AJNT-2x (log2-fold change>1 or <-1, *p* value<0.05) (Figure 3). DEGs between AJNT-4x and AJNT-2x at 10, 15 and 20 DAF were compared and annotated by the GO database, and selected for enrichment analysis. At 10 DAF, the main terms were single-organism cellular process in the biological process category, intrinsic to membrane in the cellular component category, and protein binding and hydrolase activity in the molecular function category. At 15 DAF, the enriched biological processes mainly included organic substance metabolic process and primary metabolic process, the main enriched cellular component was intracellular and intracellular part, and the main enriched molecular functions were ion binding and transferase activity. At 20 DAF, the main enriched GO terms cellular metabolic process among biological processes, intracellular part and intracellular for cellular components, and organic cyclic compound binding and heterocyclic compound binding among molecular functions (Figure S2).

Then, to examine metabolic processes and possible signaling pathways, these DEGs were examined by KEGG analysis. At 10 DAF, the DEGs were mainly enriched in metabolic pathways and starch and sucrose metabolism. At 15 DAF, the DEGs were mainly enriched in ribosome and biosynthesis of amino acids. At 20 DAF, the DEGs were mainly enriched in metabolic pathways and biosynthesis of secondary metabolites. Interestingly, they were enriched in plant-pathogen interaction, taurine and hypotaurine metabolism, biosynthesis of amino acids, glyoxylate and dicarboxylate metabolism and carbon metabolism (Figure 4). Therefore, the number of enriched pathways gradually increased

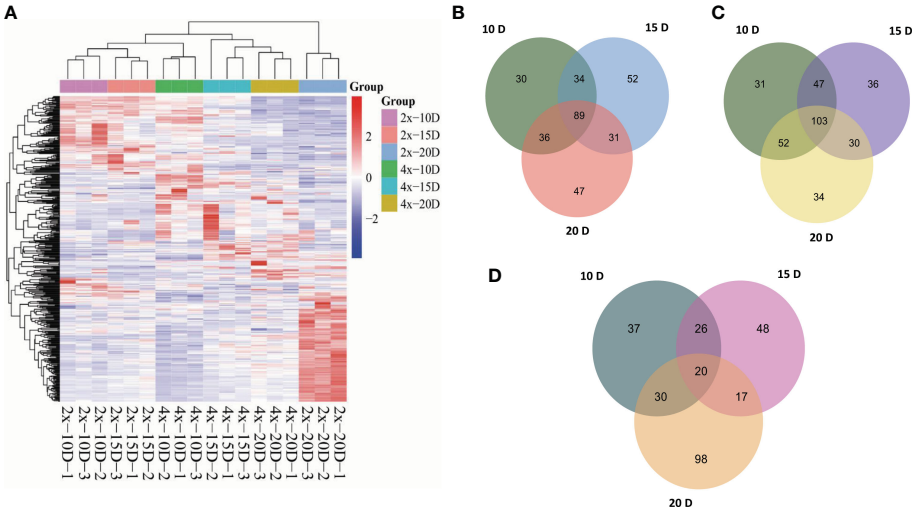


FIGURE 2
Metabolomic profiles and comparative analysis of AJNT-4x and AJNT-2x. **(A)** Hierarchical cluster analysis (HCA). The color scale shows that the abundance of accumulated is different among samples. The abscissa represents the sample name. **(B)** Upregulation of metabolites in AJNT-4x at 10, 15, and 20 DAF. **(C)** Downregulation of metabolites in AJNT-4x at 10, 15, and 20 DAF. **(D)** Comparison of significantly different metabolites (fold change>1.2 or fold change<0.83, *p* value<0.05) between AJNT-4x and AJNT-2x.

with endosperm development, and the number of DGEs also increased with the development of rice endosperm.

Analysis of genes related to nutritional quality between AJNT-4x and AJNT-2x at maturity

The nutritional quality of rice is related to the accumulation and metabolism of amino acids. The related metabolic pathways are cysteine and methionine metabolism; tryptophan metabolism; glycine, serine and threonine metabolism; valine, leucine and isoleucine degradation; phenylalanine metabolism; arginine and proline metabolism; lysine degradation; alanine, aspartate and glutamate metabolism; phenylalanine, tyrosine and tryptophan biosynthesis; arginine biosynthesis; tyrosine metabolism; valine, leucine and isoleucine biosynthesis; lysine biosynthesis and histidine metabolism. In addition, the expression of most of these genes was higher in AJNT-2x than in AJNT-4x, but the levels of many amino acids, such as lysine, aspartic acid, and glycine, were higher in AJNT-4x than in AJNT-2x (Table S2). There may be a negative correlation between gene expression and some compound accumulation.

Weighted correlation network analysis between AJNT-4x and AJNT-2x at different developmental stages

We identified 19 WGCNA modules after performing a weighted gene coexpression network analysis (WGCNA) of the FPKM values of all genes and screening those with the top 75% of the median absolute deviation values; nevertheless, the genes in the "MEgray" module were not separated into other modules (Figure 5A). To understand the mode-trait correlation, differential metabolites from three stages were selected for analysis, including 20 key metabolites selected from different developmental stages, such as succinate, vanillin, spermidine, vicine, oleuropein, 4-guanidinobutyric acid, corticosterone and other metabolites (Figure 5B). The results revealed that spermidine, vicine, 4

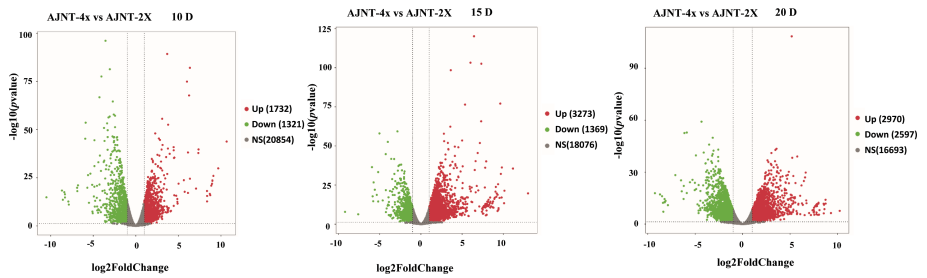


FIGURE 3
Volcano plot of differentially expressed genes (DEGs). The panel shows AJNT-4x compared with AJNT-2x, from left to right, it's 10, 15, 20DAF, respectively, red dots represent upregulation and green dots represent downregulation in AJNT-4x (log2-fold change>1 or <-1, *p* value<0.05).

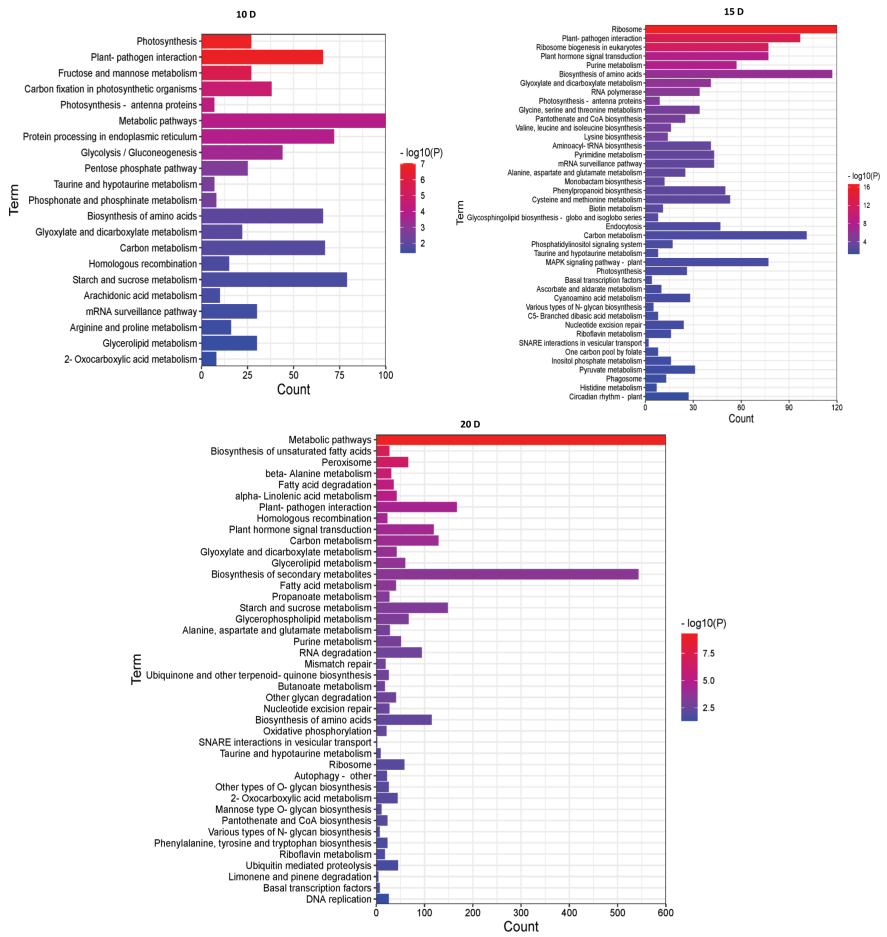


FIGURE 4
KEGG enrichment analysis of DEGs between AJNT-4x and AJNT-2x. The abscissa represents the number of genes, the ordinate represents the enrichment of pathways, the length of the column represents the number of genes, and the color represents the *p* value.

–guanidinobutyric acid and 9 other metabolites had similar expression patterns, while methyl eugenol, oleuropein, eurycomalactone and 15 other metabolites showed the opposite pattern of expression.

Molecular function verification of two lysine-related DEGs by CRISPR/Cas9 gene editing technology

The nutritional quality of rice is not only determined by its protein content, but also related to the composition and abundance of amino acids, especially the abundance of essential amino acids. Lysine is the first restrictive essential amino acid in rice seeds, and is the main source of nutrition for humans (Yang et al., 2018). Surprisingly, 39 genes related to lysine synthesis were found in the transcriptome analysis, and only 35 of the genes were expressed (Figure 6A). Two downregulation genes unreported previously, *XP_015633476.1* and *XP_015631441.1* (namely, *OsLC4* and *OsLC3*, respectively), were selected to verify their molecular function by CRISPR/Cas9 gene editing technology.

In the T_0 transgenic lines of *OsLC4*, the targets of four of 38 plants were edited, and the first target generated a frameshift mutation with a 4-nt (TATT) deletion, while the second target did not generate a mutant. In the T_0 transgenic lines of *OsLC3*, the targets of four of 12 plants were edited, and the first target generated two editing types, *lc3*-1 for a 1-nt (G) insertion, and *lc3*-2 for a 14-nt (CCTGAGGTTTGTTT) deletion, while the second target did not generate a mutant (Figure 6B). The target mutantions of the two genes all resulted in frameshift mutations of amino acids. T_0 mutant plants were then grown in T_1 lines. Homozygous mutants with free T-DNA were examined for lysine content, and their mRNA expression was also checked simultaneously. In the knockout mutants, the mRNA expression level obviously declined (Figure 6C), and the lysine level dramatically increased (Figure 6D). These results showed that both *OsLC4* and *OsLC3* negatively regulated lysine abundance, which further verified to the transcriptome data. In the autotetraploid rice AJNT-4x, the expression levels of the two genes *OsLC4* and *OsLC3* were lower than those in the diploid rice AJNT-2x, but the lysine abundance in AJNT-4x was higher than that in AJNT-2x.

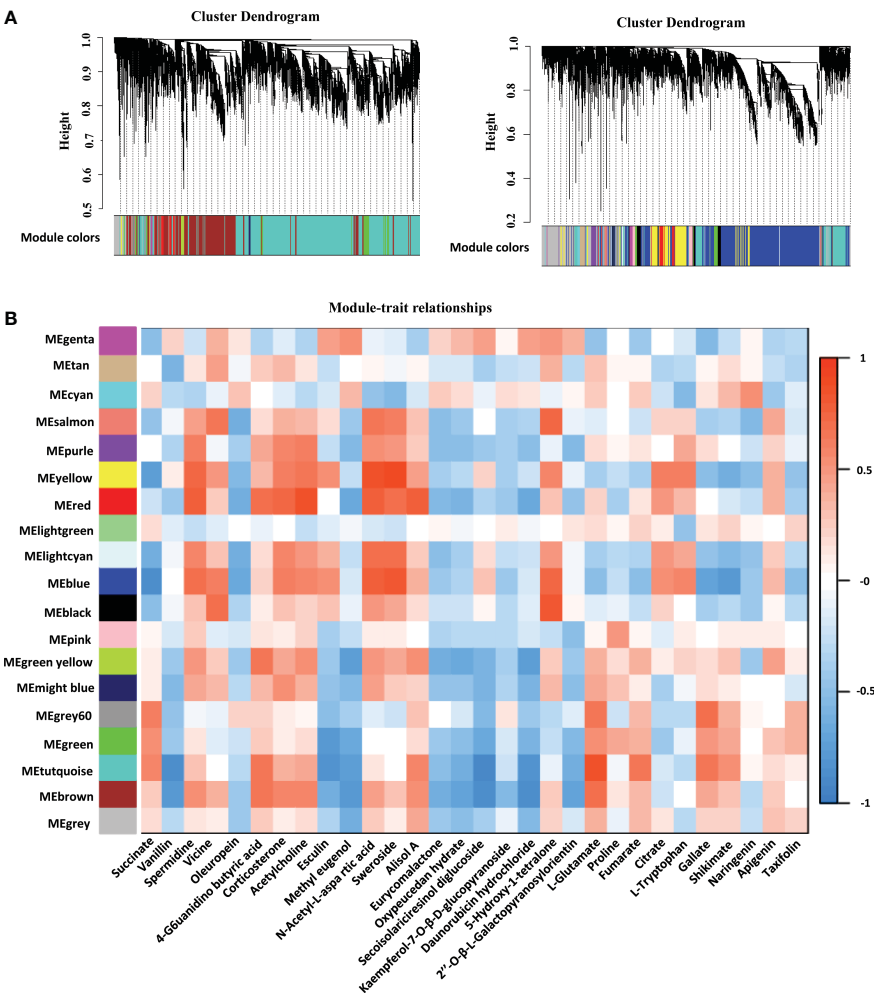


FIGURE 5
Network analysis dendrogram presenting the modules identified by weighted gene coexpression network analysis (WGCNA). **(A)** All genes are divided into different modules by different functions. **(B)** Module-metabolites weight correlations and corresponding *p* values. The left panel displays the 19 modules. The scale bar on the right illustrates the module-trait correlation from -1 (blue) to 1 (red). The text below represents different metabolites.

Discussion

Autotetraploid rice may possess better stress resistance and have greater health and nutritional benefits

Plant phenotypes and nutrition are typically altered as a result of polyploidization. Tetraploid rice seeds frequently expand in size and weight compared with those of diploid rice seeds. The degree, rate, gel consistency, alkali spreading value, and gelatinization temperature are all altered at the same time (Gu et al., 2015; Ku et al., 2022). In tetraploid rice, the protein and amino acid levels rise, but the amylose concentration frequently declines (Wang et al., 2022c). Functional proteins, oils, vitamins, flavonoids, trace minerals, necessary amino acids, and functional and essential proteins are unique nutritional components found in rice (Nam et al., 2018; Yang et al., 2019; Sabooni and Gharaghani, 2022; Zhang et al., 2022). Strong antioxidant properties of flavonoids were demonstrated by their ability to directly scavenge free radicals, inhibit the production of free radicals by the enzyme oxidase, chelate metal ions produced by free radicals, and promote the regeneration of antioxidant enzymes and small-molecular antioxidant factors (Bors and Michel, 2002; Borges-Bubols et al., 2013). In this study, A total of 422 metabolites were detected in the experiment (Table S1), including flavonoids, terpenoids, amino acids, peptides and analogs, alkaloids, phenols and derivatives and other metabolites. It was found that the abundance of most flavonoids in autotetraploid rice was higher than that in autodiploid rice during endosperm development. Therefore, autotetraploid rice has stronger antioxidant activity, which can better resist oxidation. The flavonoids include hesperidin, methyl hesperidin, naringenin, neohesperidin, naringenin chalcone and 42 other flavonoids. Hesperidin has anti-inflammatory, antiviral, antibacterial and other effects (Wilmsen et al., 2005; Xiong et al., 2019). Methyl hesperidin is as powerful as vitamin P, which can boost vitamin C's effects and has potent antiviral and antibacterial properties (Jawien et al., 2017). Naringin can inhibit the infiltration of macrophages into fat cells in a high-fat diet, thus inhibiting obesity (Bharti et al., 2014; Chen et al., 2016). Neohesperidin has anticancer, antioxidant and antiviral effects (Gong

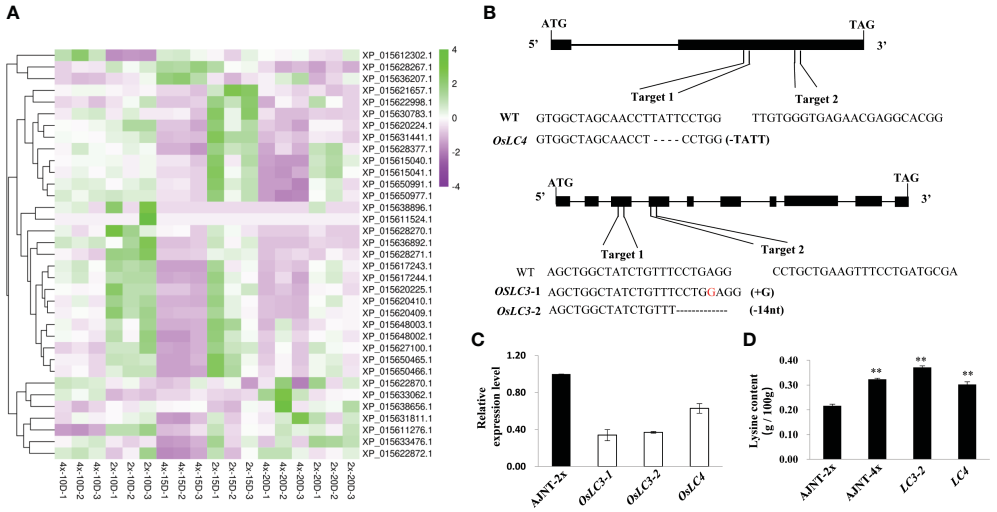


FIGURE 6
DEGs identified by lysine-related analysis and gene function verification by CRISPR/Cas9 technology. **(A)** DEGs of lysine-related analysis. The abscissa represents the sample name, the ordinate represents the gene name, and the color represents the level of expression. **(B)** *OsLC3* and *OsLC4* generated editing types in the target plant. **(C)** The expression level of the knockout mutant compared with wild-type AJNT-2x in the T_2 generation, where the two asterisks show a significant difference based on a t test at $p<0.01$. **(D)** The lysine level of the knockout mutant compared with AJNT-2x and AJNT-4x in the T_2 generation. The two asterisks show a significant difference based on a t test at $p<0.01$.

et al., 2019; Karim et al., 2021). According to the results of the KEGG enrichment analysis, the difference in metabolites were primarily related to biosynthesis of secondary metabolites, microbial metabolism in diverse environments, biosynthesis of cofactors, ABC transporters, and so on. Much of the accumulation of metabolites is generated in the biosynthetic pathway of secondary metabolites and microbial metabolism can affect the accumulation of metabolites (Wei et al., 2019; Su et al., 2023), and polyploid rice include large grain size, high weight per thousand grains (Liu et al., 2022; Wang et al., 2021; He et al., 2022), which may result in that the autotetraploid rice is more nutritious and edible than the diploid rice.

Between the endosperm developmental stages of AJNT-4x and AJNT-2x, metabolites change more

The metabolites in seeds are crucial determinants of their nutritional value (Farooq et al., 2010; Zhao et al., 2020a). Hence, metabonomics has been frequently utilized to examine seed nutrients frequently (Oikawa et al., 2008; Zhao et al., 2020b). Researchers have previously compared nutrients including amino acids, proteins, and carbohydrates between diploid and autotetraploid rice using chemical methods (Gu et al., 2015; Wang et al., 2021). However, there are no publications on secondary metabolomics comparisons between diploid and tetraploid rice. Based on LC-MS/MS, 422 metabolites from six groups of diploid-tetraploid rice were discovered in this work at various endosperm developmental stages (Table S1). The research showed that the metabolite expression abundances were very different between the diploid and tetraploid groups. L-glutamate is a key substance in histidine metabolism, nitrogen metabolism, D-amino acid metabolism and other metabolic pathways and it decreases with the development of rice endosperm. L-glutamate plays an important role in the ornithine cycle (urea synthesis) pathway (Blachier et al., 2009). The amino group can be removed by glutamate dehydrogenase in mitochondria to provide free ammonia for the synthesis of carbamoyl phosphate (Potel et al., 2009; Adeva et al., 2012). In the ornithine cycle (urea synthesis) pathway, glutamate dehydrogenase in mitochondria removes the amino group of glutamate and provides free ammonia for the synthesis of carbamoyl phosphate (Newsholme et al., 2003; Smith and Stanley, 2008). Glutamic oxalate transaminase in the cytoplasm transfers the amino group of glutamic acid to oxaloacetic acid, which then forms aspartic acid through the ornithine cycle, and glutamic acid indirectly provides a second amino group for the cycle (Fernández and Zúñiga, 2006).

Therefore, we speculated that with the development of rice endosperm, L-glutamate could provide increasingly less free ammonia, and the circulation rate of ornithine decreased. L-proline is one of the amino acids that make up human protein, and it is also an ideal osmotic regulating substance that can be used as a membrane and enzyme protective substance and free radical scavenger (Liang et al., 2013; Meena et al., 2019). The growth of plants under osmotic stress plays a protective role, and L-proline also plays a regulatory role in the osmotic balance of the cytoplasm for the accumulation of another important osmotic regulatory substance in the vacuole of potassium ion organisms (Parvaiz and Satyawati, 2008; Takagi, 2008).

Interestingly, the level of L-proline in tetraploid rice decreased with the development of rice endosperm, while the content of L-proline in diploid rice increased first and then decreased, which may indicate that the initial and maximum accumulation times of metabolites differed among materials. The TCA cycle is a ubiquitous metabolic pathway in aerobic organisms (Dijkstra et al., 2011; Wang et al., 2022b). Fumarate and succinate decreased with endosperm development, while citrate increased with endosperm development in tetraploid rice and

decreased first and then increased in diploid rice (Figure 7). According to WGCNA, genes related to succinate and fumarate synthesis were mainly enriched in the 'MEturquoise' module, which contained 11,231 genes (Figure 5B). The main functions were positive regulation of signaling, PAS complex and AP-type membrane coat adaptor complex (Figure S3), which can impact the TCA cycle process.

Higher amino acid levels in AJNT-4x than in AJNT-2x during the late stage of development and two novel genes involved in the regulation of lysine synthesis

Humans are unable to produce the nine necessary amino acids valine, leucine, isoleucine, phenylalanine, tryptophan, threonine, histidine, methionine, and lysine (Galili et al., 2016; Shi et al., 2022). Valine can help the body grow normally, repair tissues, control blood sugar, and meet energy needs (Hao et al., 2020; Yu et al., 2021b). Threonine can help people feel less tired, encourage growth and development, hold water in the skin, bind to oligosaccharide chains, preserve cell membranes, and increase phospholipid production and fatty acid oxidation in living organisms (Cheng et al., 2019; Tang et al., 2021). Methionine can protect the liver, treat depression and lower blood pressure (Sanderson et al., 2019; Lauinger and Kaiser, 2021). Lysine is the first limiting amino acid in the human body and has very important physiological functions in the human body (Shewry, 2007; Liao et al., 2015). It can participate in the synthesis of skeletal muscle, enzymes and peptide hormones and is one of the ketogenic amino acids. When the body lacks available carbohydrates, lysine can participate in ketone body synthesis and glucose metabolism to maintain acid-base balance in the body (Li et al., 2020). Inadequate lysine intake will block the absorption of other amino acids, thus hindering the effective utilization of rice protein, resulting in the imbalance of dietary structure and nutritional composition of food and ultimately the weakening of human metabolic function and metabolic disorders (Yao et al., 2020). The lysine level of tetraploid and diploid rice at 15 and 20 DAF was detected and found to be higher in tetraploid rice (Xian et al., 2021).

In this study, we found that the content of most essential amino acids in AJNT-4x was higher than that in AJNT-2x, so we assume that the nutritional value of tetraploid rice is higher than that of diploid rice. In addition, combined with transcriptome analysis, 39 genes related to the lysine synthesis pathway were identified, and only 35 of the genes were expressed. The expression levels of most genes were lower in tetraploids than in diploids, and these genes have the potential to reverse regulate lysine synthesis. Two novel genes (*OsLC4* and *OsLC3*) were selected to verify their molecular function by CRISPR/Cas 9 gene editing technology. We found that the target mutations of the two genes resulted in frameshift mutations of amino acids, the mRNA expression level obviously declined, and the lysine level dramatically increased in the knockout mutants. Therefore, the *OsLC4* and *OsLC3* genes both negatively regulate lysine abundance, and we will select more genes related to lysine synthesis for verification in the future.

Conclusion

In conclusion, a total of 422 metabolites were observed between AJNT-4x and AJNT-2x, and 3053, 4642, and 5567 DGEs were

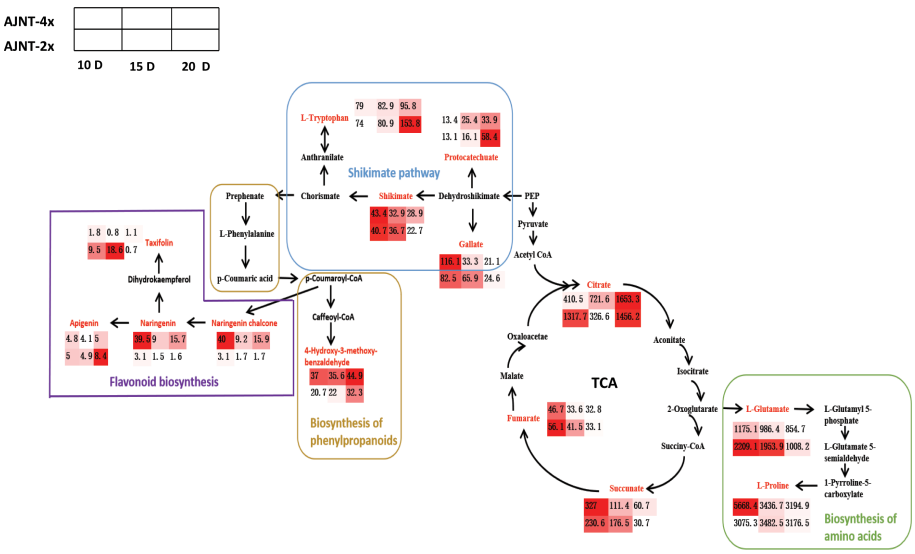


FIGURE 7
Screening for maps of metabolic pathways involving in key differentially expressed metabolites. The pathway map includes mainly flavonoid biosynthesis, shikimate pathway, biosynthesis of phenylpropanoids, biosynthesis of amino acids and TCA pathway; the red color indicates the differentially expressed metabolites screened; the small heat boxes show the changes in the contents of differentially expressed metabolites between AJNT-4x and AJNT-2x at 10, 15, and 20 DAF, respectively.

identified at 10, 15 and 20 DAF, respectively. We identified key metabolites participating in the biosynthesis of amino acids and phenylpropanoids and the metabolism of glutathione and inositol phosphate, which provide energy and raw materials for rice polyploids and performed WGCNA to understand the regulatory genes of metabolites. In addition, by using CRISPR/Cas9 gene-editing technology, we found that two novel genes, *OsLC4* and *OsLC3*, negatively regulated lysine abundance. These results provide new insights into dynamic metabolite and gene expression differences during endosperm development in autotetraploid rice, which will aid in the development of rice cultivars with increased yield and improved grain nutritional quality.

Data availability statement

The datasets generated for this study can be found in the Supplementary material, and China National GeneBank DataBase (CNGBdb) database under the accession number: CNP0004286.

Author contributions

LX: performing the experiments, analyzing the data, writing-original draft, writing-review and editing. JT: writing-review and editing. YL: performing the experiments. HM: performing the experiments. MT: performing the experiments. XL: developing autotetraploid rice AJNT-4x, writing-review and editing. GY: funding acquisition, writing-review and editing. LW: funding acquisition, designing the research, analyzing the data, writing-review and editing. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Guangdong Provincial Natural Science Foundation Program (No. 2019A1515011826), and Guangdong Provincial Special Projects in Key Field of Rural Revitalization (No. 2020ZDZX1038).

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.2023.1210134/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Detection of metabolites according to LC-MS/MS. **(A)** MRM metabolite detection multipeak map (multisubstance extraction ion current spectrum, XIC). Note: The abscissa is the retention time of metabolite detection (retention time, Rt). The ordinate is the ion current intensity of the ion detection (the intensity units are counts per second (cps)). **(B)** Detection of the TIC overlap map by QC sample mass spectrometry.

SUPPLEMENTARY FIGURE 2
GO enrichment analysis of DEGs between AJNT-4x and AJNT-2x. The ordinate represents the functions of biological process (BP), cellular component (CC) and molecular function (MF), and bubble size represents the number of genes.

SUPPLEMENTARY FIGURE 3
Enrichment of graph. Note: Four circles from the outside to the inside. The first circle: enriched classification, outside the circle is the scale of gene number. Different colors represent different categories. Second circle: the number of background genes in this category and Q or P values. The more genes there are, the longer the bar, the smaller the value, and the redder the color. Third circle: the total number of foreground genes. Fourth circle: RichFactor value of each classification (the number of foreground genes divided by the number of background genes in the classification). Each small cell of the background auxiliary line represents 0.1.

References

Adeva, M. M., Souto, G., Blanco, N., and Donapetry, C. (2012). Ammonium metabolism in humans. *Metabolism* 61, 1495–1511. doi: 10.1016/j.metabol.2012.07.007

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, 106. doi: 10.1038/npre.2010.4282.1

Bharti, S., Rani, N., Krishnamurthy, B., and Arya, D. S. (2014). Preclinical evidence for the pharmacological actions of naringin: a review. *Planta Med.* 80, 437–451. doi: 10.1055/s-0034-1368351

Blachier, F., Boutry, C., Bos, C., and Tome, D. (2009). Metabolism and functions of l-glutamate in the epithelial cells of the small and large intestines. *Am. J. Clin. Nutr.* 90, 814S–821S. doi: 10.3945/ajcn.2009.27462S

Borges-Bubols, G., da-Rocha Vianna, D., Medina-Remon, A., von Poser, G., Maria Lamuela-Raventos, R., Lucia Eifler-Lima, V., et al. (2013). The antioxidant activity of coumarins and flavonoids. *Mini Rev. Med. Chem.* 13, 318–334. doi: 10.2174/138955713804999775

Bors, W., and Michel, C. (2002). Chemistry of the antioxidant effect of polyphenols. *Ann. New York Acad. Sci.* 957, 57–69. doi: 10.1111/j.1749-6632.2002.tb02905.x

Chen, R., Feng, Z., Zhang, X., Song, Z., and Cai, D. (2021). A new way of rice breeding: polyploid rice breeding. *Plants* 10, 422. doi: 10.3390/plants10030422

Chen, R., Qi, Q. L., Wang, M. T., and Li, Q. Y. (2016). Therapeutic potential of naringin: an overview. *Pharm. Biol.* 54, 3203–3210. doi: 10.1080/13880209.2016.1216131

Chen, Y., Shahid, M. Q., Wu., J., Deng, R., Chen, Z., Wang, L., et al. (2022). Thermo-sensitive genic Male sterile lines of neo-tetraploid rice developed through gene editing technology revealed high levels of hybrid vigor. *Plants* 11 (11), 1390. doi: 10.3390/plants11111390

Cheng, Z. X., Guo, C., Chen, Z. G., Yang, T. C., Zhang, J. Y., Wang, J., et al. (2019). Glycine, serine and threonine metabolism confounds efficacy of complement-mediated killing. *Nat. Commun.* 10, 1–17. doi: 10.1038/s41467-019-11129-5

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

de Hoon, M. J., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. *Bioinformatics* 20, 1453–1454. doi: 10.1093/bioinformatics/bth078

Dijkstra, P., Thomas, S. C., Heinrich, P. L., Koch, G. W., Schwartz, E., and Hungate, B. A. (2011). Effect of temperature on metabolic activity of intact microbial communities: evidence for altered metabolic pathway activity but not for increased maintenance respiration and reduced carbon use efficiency. *Soil Biol. Biochem.* 43, 2023–2031. doi: 10.1016/j.soilbio.2011.05.018

Farooq, M., Basra, S. M., Wahid, A., and Ahmad, N. (2010). Changes in nutrient-homeostasis and reserves metabolism during rice seed priming: consequences for seedling emergence and growth. *Agric. Sci. China.* 9, 191–198. doi: 10.1016/S1671-2927(09)60083-3

Fernández, M., and Zúñiga, M. (2006). Amino acid catabolic pathways of lactic acid bacteria. *Crit. Rev. Microbiol.* 32, 155–183. doi: 10.1080/10408410600880643

Galili, G., Amir, R., and Fernie, A. R. (2016). The regulation of essential amino acid synthesis and accumulation in plants. annu. *Rev. Plant Biol.* 67, 153–178. doi: 10.1146/annurev-arplant-043015-112213

Gan, L., Huang, B., Song, Z., Zhang, Y., Zhang, Y., Chen, S., et al. (2021). Unique glutelin expression patterns and seed endosperm structure facilitate glutelin accumulation in polyploid rice seed. *Rice* 14, 1–19. doi: 10.1186/s12284-021-00500-0

Ghaleb, M. A. A., Li, C., Shahid, M. Q., Yu, H., Liang, J., Chen, R., et al. (2020). Heterosis analysis and underlying molecular regulatory mechanism in a wide-compatible neo-tetraploid rice line with long panicles. *BMC Plant Biol.* 20, 1–15. doi: 10.1186/s12870-020-2291-z

Gong, Y., Dong, R., Gao, X., Li, J., Jiang, L., Zheng, J., et al. (2019). Neohesperidin prevents colorectal tumorigenesis by altering the gut microbiota. *Pharmacol. Res.* 148, 104460. doi: 10.1016/j.phrs.2019.104460

Gu, Y., Dai, X., and Li, J. (2015). Rice quality analysis of different ploidy rice. *Zhengzhou Univ.* 47, 80–85. doi: 10.3969/j.issn.1671-6841.2015.04.016

Guo, H., Mendrikahy, J. N., Xie, L., Deng, J., Lu, Z., Wu, J., et al. (2017). Transcriptome analysis of neo-tetraploid rice reveals specific differential gene expressions associated with fertility and heterosis. *Sci. Rep.* 7, 1–11. doi: 10.1038/srep40139

Hao, Y., Ma, Q., Liu, X., Fan, X., Men, J., Wu, H., et al. (2020). High-yield production of l-valine in engineered escherichia coli by a novel two-stage fermentation. *Metab. Eng.* 62, 198–206. doi: 10.1016/j.ymben.2020.09.007

He, L., Guo, F. Y., Cai, X. J., Chen, H. P., Lian, C. L., Wang, Y., et al. (2023). Evolutionary origin and establishment of a dioecious diploid-tetraploid complex. *Mol. Ecol.* 00, 1–18. doi: 10.1111/mec.16902

He, W. T., Zhang, X. H., Lv, P. C., Wang, W., Wang, J., He, Y. C., et al. (2022). Full-length transcriptome reconstruction reveals genetic differences in hybrids of *Oryza sativa* and *Oryza punctata* with different ploidy and genome compositions. *BMC Plant Biol.* 22 (1), 131. doi: 10.1186/s12870-022-03502-2

Hu, C., Shi, J., Quan, S., Cui, B., Kleessen, S., Nikoloski, Z., et al. (2014). Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics. *Sci. Rep.* 4, 1–10. doi: 10.1038/srep05067

Jawien, A., Bouskela, E., Allaert, F. A., and Nicolaides, A. N. (2017). The place of ruscus extract, hesperidin methyl chalcone, and vitamin c in the management of chronic venous disease. *Int. Angiology: J. Int. Union Angiology* 36, 31–41. doi: 10.23736/s0392-9590.16.03788-3

Kamara, N., Jiao, Y., Lu, Z., Aloryi, K. D., Wu, J., Liu, X., et al. (2021). Cytological observations and bulked-segregant analysis coupled global genome sequencing reveal two genes associated with pollen fertility in tetraploid rice. *Int. J. Mol. Sci.* 22 (2), 841. doi: 10.3390/ijms22020841

Karim, N., Shishir, M. R. I., Rashwan, A. K., Ke, H., and Chen, W. (2021). Suppression of palmitic acid-induced hepatic oxidative injury by neohesperidin-loaded pectin-chitosan decorated nanoliposomes. *Int. J. Biol. Macromol.* 183, 908–917. doi: 10.1016/j.ijbiomac.2021.05.010

Ku, T., Gu, H., Li, Z., Tian, B., Xie, Z., Shi, G., et al. (2022). Developmental differences between anthers of diploid and autotetraploid rice at meiosis. *Plants* 11 (13), 1647. doi: 10.3390/plants11131647

Langfelder, P., and Horvath, S. (2008). WGCNA: an r package for weighted correlation network analysis. *BMC Bioinf.* 9, 1–13. doi: 10.1186/1471-2105-9-559

Lauinger, L., and Kaiser, P. (2021). Sensing and signaling of methionine metabolism. *Metabolites* 11, 83. doi: 10.3390/metabo11020083

Leng, N., Dawson, J. A., Thomson, J. A., Ruotti, V., Rissman, A. I., Smits, B. M., et al. (2013). EBSeq: an empirical bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29, 1035–1043. doi: 10.1093/bioinformatics/btt087

Li, Q., and Song, J. (2019). Analysis of widely targeted metabolites of the euhalophyte suaeda salsa under saline conditions provides new insights into salt tolerance and nutritional value in halophytic species. *BMC Plant Biol.* 19, 1–11. doi: 10.1186/s12870-019-2006-5

Li, X., Zheng, S., and Wu, G. (2020). Amino acid metabolism in the kidneys: nutritional and physiological significance. *Amino Acids Nutr. Health* 1265, 71–95. doi: 10.1007/978-3-030-45328-2_5

Liang, X., Zhang, L., Natarajan, S. K., and Becker, D. F. (2013). Proline mechanisms of stress survival. *Antioxid. Redox Signal.* 19, 998–1011. doi: 10.1089/ars.2012.5074

Liao, S. F., Wang, T., and Regmi, N. J. S. (2015). Lysine nutrition in swine and the related monogastric animals: muscle protein biosynthesis and beyond. *Springerplus* 4, 1–12. doi: 10.1186/s40064-015-0927-5

Liu, X. D., Wu, J. W., and Shahid, M. Q. (2022). Development of neo-tetraploid rice and research progress on its heterosis mechanism. *Biotechnol. Bull.* 38, 44–50. doi: 10.13560/j.cnki.biotech.bull.1985.2021-0406

Lu, Z., Guo, X., Huang, Z., Xia, J., Li, X., Wu, J., et al. (2020). Transcriptome and gene editing analyses reveal MOF1a defect alters the expression of genes associated with tapetum development and chromosome behavior at meiosis stage resulting in low pollen fertility of tetraploid rice. *Int. J. Mol. Sci.* 21, 7489. doi: 10.3390/ijms21207489

Mao, X., Cai, T., Olyarchuk J, G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430

Meena, M., Divyanshu, K., Kumar, S., Swapnil, P., Zehra, A., Shukla, V., et al. (2019). Regulation of l-proline biosynthesis, signal transduction, transport, accumulation and its vital role in plants during variable environmental conditions. *Heliyon.* 5, e02952. doi: 10.1016/j.heliyon.2019.e02952

Nam, K. H., Kim, D. Y., Shin, H. J., Pack, I. S., and Kim, C. G. (2018). Changes in the metabolic profile and nutritional composition of rice in response to NaCl stress. *Korean J. Agric. Sci.* 45, 154–168. doi: 10.7744/kjoas.20180019

Newsholme, P., Procopio, J., Lima, M. M. R., Pithon-Curi, T. C., and Curi, R. (2003). Glutamine and glutamate - their central role in cell metabolism and function. *Cell Biochem.* 21, 1–9. doi: 10.1002/cbf.1003

Oikawa, A., Matsuda, F., Kusano, M., Okazaki, Y., and Saito, K. (2008). Rice metabolomics. *Rice* 1, 63–71. doi: 10.1007/s12284-008-9009-4

Parvaiz, A., and Satyawati, S. (2008). Salt stress and phyto-biochemical responses of plants-a review. *Plant Soil Environ.* 54, 89–99. doi: 10.17221/2774-PSE

Potel, F., Valadier, M. H., Ferrario-Méry, S., Grandjean, O., Morin, H., Gaufichon, L., et al. (2009). Assimilation of excess ammonium into amino acids and nitrogen translocation in arabidopsis thaliana - roles of glutamate synthases and carbamoylphosphate synthetase in leaves. *FEBS J.* 276, 4061–4076. doi: 10.1111/j.1742-4658.2009.07114.x

Qu, Y., Shang, X., Zeng, Z., Yu, Y., Bian, G., Wang, W., et al. (2023). Whole-genome duplication reshaped adaptive evolution in a relict plant species, cyclocarya paliurus. *Genomics Proteomics Bioinf.* S1672-0229 (23), 00033-5. doi: 10.1016/j.gpb.2023.02.001

Sabooni, N., and Gharaghani, A. (2022). Induced polyploidy deeply influences reproductive life cycles, related phytochemical features, and phytohormonal activities in blackberry species. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.938284

Sanderson, S. M., Gao, X., Dai, Z., and Locasale, J. W. (2019). Methionine metabolism in health and cancer: a nexus of diet and precision medicine. *Nat. Rev. Cancer* 19, 625–637. doi: 10.1038/s41568-019-0187-8

Shen, S. Q., Zhan, C. S., Yang, C. K., Fernie, A. R., and Luo, J. (2023). Metabolomics-centered mining of plant metabolic diversity and function: past decade and future perspectives. *Mol. Plant* 16 (1), 43–63. doi: 10.1016/j.molp.2022.09.007

Shewry, P. R. (2007). Improving the protein content and composition of cereal grain. *J. Cereal Sci.* 46, 239–250. doi: 10.1016/j.jcs.2007.06.006

Shi, Y., Zhang, Y., Sun, Y., Xie, Z., Luo, Y., Long, Q., et al. (2022). Natural variations of OsAUX5, a target gene of *OsWRKY78*, control the contents of neutral essential amino acids in rice grains. *Mol. Plant* 12, 13. doi: 10.1016/j.molp.2022.12.013

Singh, A., Kumar, Y., and Arora, B. (2021). Storage proteins profile in diploid and tetraploid seeds of *oryza sativa*. *Cutting-Edge Res. Agric. Sci.* 8, 41–52. doi: 10.9734/bpi/cras/v8/8102D

Smith, T. J., and Stanley, C. A. (2008). Untangling the glutamate dehydrogenase allosteric nightmare. *Trends Biochem. Sci.* 33, 557–564. doi: 10.1016/j.tibs.2008.07.007

Su, J., Wang, Y., Bai, M., Peng, T., Li, H., Xu, H. J., et al. (2023). Soil conditions and the plant microbiome boost the accumulation of monoterpenes in the fruit of citrus reticulata 'Chachi'. *Microbiome* 11 (1), 61. doi: 10.1186/s40168-023-01504-2

Takagi, H. (2008). Proline as a stress protectant in yeast: physiological functions, metabolic regulations, and biotechnological applications. *Appl. Microbiol. Biotechnol.* 81, 211–223. doi: 10.1007/s00253-008-1698-5

Tang, Q., Tan, P., Ma, N., and Ma, X. (2021). Physiological functions of threonine in animals: beyond nutrition metabolism. *Nutrients* 13, 2592. doi: 10.3390/nu13082592

Ulaszewska, M. M., Weinert, C. H., Trimigno, A., Portmann, R., Andres Lacueva, C., Badertscher, R., et al. (2019). Nutrimetabolomics: an integrative action for metabolomic analyses in human nutritional studies. *Mol. Nutr. Food Res.* 63, 1800384. doi: 10.1002/mnfr.201800384

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26

Wang, N. N., Fan, X. H., Lin, Y. J., Li, Z., Wang, Y. K., Zhou, Y. M., et al. (2022c). Alkaline stress induces different physiological, hormonal and gene expression responses in diploid and autotetraploid rice. *Int. J. Mol. Sci.* 23 (10), 5561. doi: 10.3390/ijms23105561

Wang, Y., Hao, R., Guo, R., Nong, H., Qin, Y., and Dong, N. (2023). Integrative analysis of metabolome and transcriptome reveals molecular insight into metabolomic variations during hawthorn fruit development. *Metabolites* 13 (3), 423. doi: 10.3390/metabo13030423

Wang, L., Liu, Y., Zhao, H., Zheng, Y., Bai, F., Deng, S., et al. (2022a). Identification of *qGL3. 5*, a novel locus controlling grain length in rice through bulked segregant analysis and fine mapping. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.921029

Wang, L., Qu, F., Zhu, Z., Zhao, Y., Chen, X., Shi, M., et al. (2022b). The important role of tricarboxylic acid cycle metabolism pathways and core bacterial communities in carbon sequestration during chicken manure composting. *Waste Manage.* 150, 20–29. doi: 10.1016/j.wasman.2022.06.034

Wang, W., Tu, Q., Chen, R., Lv, P., Xu, Y., Xie, Q., et al. (2021). Polyploidization increases the lipid content and improves the nutritional quality of rice. *Plants* 11, 132. doi: 10.3390/plants11010132

Wang, D., Zhang, L., Huang, X., Wang, X., Yang, R., Mao, J., et al. (2018). Identification of nutritional components in black sesame determined by widely targeted metabolomics and traditional Chinese medicines. *Molecules* 23, 1180. doi: 10.3390/molecules23051180

Waris, M., Koçak, E., Gonulalan, E. M., Demirezer, L. O., Kir, S., and Nemutlu, E. (2022). Metabolomics analysis insight into medicinal plant science. *TrAC Trends Anal. Chem.* 157, 116795. doi: 10.1016/j.trac.2022.116795

Wei, Z., Gu, Y., Friman, V. P., Kowalchuk, G. A., Xu, Y., Shen, Q., et al. (2019). Initial soil microbiome composition and functioning predetermine future plant health. *Sci. Adv.* 5 (9), eaaw0759. doi: 10.1126/sciadv.aaw0759

Wilmsen, P. K., Spada, D. S., and Salvador, M. (2005). Antioxidant activity of the flavonoid hesperidin in chemical and biological systems. *J. Agric. Food Chem.* 53, 4757–4761. doi: 10.1021/jf0502000

Xian, L., Long, Y., Yang, M., Chen, Z., Wu, J., Liu, X., et al. (2021). iTRAQ-based quantitative glutelin proteomic analysis reveals differentially expressed proteins in the physiological metabolism process during endosperm development and their impacts on yield and quality in autotetraploid rice. *Plant Sci.* 306, 110859. doi: 10.1016/j.plantsci.2021.110859

Xiong, H. J., Wang, J., Ran, Q., Lou, G. H., Peng, C. Y., Gan, Q. X., et al. (2019). Hesperidin: a therapeutic agent for obesity. drug design. *Dev. Ther.* 13, 3855. doi: 10.2147/DDDT.S227499

Xiong, Z., Xiong, D. L., Cai, D. T., Wang, W., Cui, K. H., Peng, S. B., et al. (2022). Effect of stomatal morphology on leaf photosynthetic induction under fluctuating light across diploid and tetraploid rice. *Environ. Exp. Bot.* 194, 104757. doi: 10.1016/j.envexpbot.2021.104757

Xue, G. R., Su, S. S., Yan, P. F., Shang, J. W., Wang, J. X., Yan, C. Y., et al. (2022). Integrative analyses of widely targeted metabolomic profiling and derivatization-based LC-MS/MS reveals metabolic changes of *Zingiberis rhizoma* and its processed products. *Food Chem.* 389, 133068. doi: 10.1016/j.foodchem.2022.133068

Yang, M., Yang, J., Su, L., Sun, K., Li, D., Liu, Y., et al. (2019). Metabolic profile analysis and identification of key metabolites during rice seed germination under low-temperature stress. *Plant Sci.* 289, 110282. doi: 10.1016/j.plantsci.2019.110282

Yang, Q. Q., Zhao, D. S., Zhang, C. Q., Wu, H. Y., Li, Q. F., Gu, M. H., et al. (2018). A connection between lysine and serotonin metabolism in rice endosperm. *Plant Physiol.* 176, 1965–1980. doi: 10.1104/pp.17.01283

Yao, Z. D., Cao, Y. N., Peng, L. X., Yan, Z. Y., and Zhao, G. (2020). Coarse cereals and legume grains exert beneficial effects through their interaction with gut microbiota: a review. *J. Agric. Food Chem.* 69, 861–877. doi: 10.1021/acs.jafc.0c05691

Yu, H., Li, Q. H., Li, Y. D., Yang, H. J., Lu, Z. J., Wu, J. W., et al. (2021a). Genomics analyses reveal unique classification, population structure and novel allele of neo-tetraploid rice. *Rice* 14, 16. doi: 10.1186/s12284-021-00459-y

Yu, D., Richardson, N. E., Green, C. L., Spicer, A. B., Murphy, M. E., Flores, V., et al. (2021b). The adverse metabolic effects of branched-chain amino acids are mediated by isoleucine and valine. *Cell Metab.* 33, 905–922.e906. doi: 10.1016/j.cmet.2021.03.025

Yu, H., Shahid, M. Q., Li, Q., Li, Y., Li, C., Lu, Z., et al. (2020). Production assessment and genome comparison revealed high yield potential and novel specific alleles associated with fertility and yield in neo-tetraploid rice. *Rice* 13, 1–13. doi: 10.1186/s12284-020-00387-3

Zhang, F., Yang, L., Huang, W., Luo, X., Xie, J., Hu, B., et al. (2022). Flavonoid metabolic profiles and gene mapping of rice (Oryza sativa l.) purple gradient grain hulls. *Rice* 15, 1–12. doi: 10.1186/s12284-022-00589-x

Zhao, X., Chen, L., Wu, J., He, Y., and Yang, H. (2020b). Elucidating antimicrobial mechanism of nisin and grape seed extract against listeria monocytogenes in broth and on shrimp through NMR-based metabolomics approach. *Int. J. Food Microbiol.* 319, 108494. doi: 10.1016/j.ijfoodmicro.2019.108494

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

Zhou, M. Y., Yong, Q., and Zhang, Q. (2016). Experimental teaching design on amino acid detection by automatic amino acid analyzer. *Lab. Sci.* 19-22, 26. doi: 10 3969/j.issn.1672-4305.2016.05.006

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 03 January 2024
DOI 10.3389/fpls.2023.1322728

Check for updates

OPEN ACCESS

EDITED BY
Raju Datla,
Global Institute for Food Security (GIFS), Canada

REVIEWED BY
Jennifer C. Fletcher,
Agricultural Research Service (USDA), United States
Vijay Kumar,
Lovely Professional University, India

*CORRESPONDENCE
Alison M. R. Ferrie
Alison.Ferrie@nrc-cnrc.gc.ca

RECEIVED 16 October 2023
ACCEPTED 11 December 2023
PUBLISHED 03 January 2024

CITATION
Yuan HY, Kagale S and Ferrie AMR (2024) Multifaceted roles of transcription factors during plant embryogenesis.
Front. Plant Sci. 14:1322728.
doi: 10.3389/fpls.2023.1322728

COPYRIGHT
© 2024 His Majesty the King in Right of Canada. 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.

Multifaceted roles of transcription factors during plant embryogenesis

Hai Ying Yuan, Sateesh Kagale and Alison M. R. Ferrie*

Aquatic and Crop Resource Development Research Center, National Research Council Canada, Saskatoon, SK, Canada

Transcription factors (TFs) are diverse groups of regulatory proteins. Through their specific binding domains, TFs bind to their target genes and regulate their expression, therefore TFs play important roles in various growth and developmental processes. Plant embryogenesis is a highly regulated and intricate process during which embryos arise from various sources and undergo development; it can be further divided into zygotic embryogenesis (ZE) and somatic embryogenesis (SE). TFs play a crucial role in the process of plant embryogenesis with a number of them acting as master regulators in both ZE and SE. In this review, we focus on the master TFs involved in embryogenesis such as BABY BOOM (BBM) from the APETALA2/Ethylene-Responsive Factor (AP2/ERF) family, WUSCHEL and WUSCHEL-related homeobox (WOX) from the homeobox family, LEAFY COTYLEDON 2 (LEC2) from the B3 family, AGAMOUS-Like 15 (AGL15) from the MADS family and LEAFY COTYLEDON 1 (LEC1) from the Nuclear Factor Y (NF-Y) family. We aim to present the recent progress pertaining to the diverse roles these master TFs play in both ZE and SE in Arabidopsis, as well as other plant species including crops. We also discuss future perspectives in this context.

KEYWORDS

AP2/ERF, B3, homeobox, somatic embryogenesis, transcription factor, zygotic embryogenesis

Abbreviations: ABI3, abscisic acid insensitive3; AFL, Abi3/Fus3/Lec2; AGL, agamous-Like; ARF, auxin response factor; BBM, baby boom; FUS3, fusca3; GFR, growth-regulating factor; GIF, grf-interacting factor; LEC1, leafy cotyledon1; LEC2, leafy cotyledon2; LIL, lec1-like; PLT, plethora; RKD, rwp-rk domain containing protein; SEP3, sepallata3; VAL, viviparous/abi3-like; VP1, viviparous1; WUS, wuschel; WOX, wuschel-like homeobox; YUC, yucca.

1 Overview of transcription factors

Transcription factors (TFs) are regulatory proteins that participate in the regulation of gene transcription. TFs have DNA-binding domains that bind to specific DNA regulatory sequences. TFs exhibit dual capabilities depending on their binding sites: they can facilitate transcription initiation when binding to DNA promoter sequences, or alternatively, they can activate or repress gene transcription when binding to enhancer sequences (Wang et al., 2015). TFs can be ubiquitous and exist in all cell types, or they can be specialized and only exist in certain cell types or certain developmental stages. There are a large number of transcription factors in living organisms including plants. Arabidopsis has close to 2300 TFs based on the recent classification, which corresponds to ~8.3% of its total genes (Hong, 2016). Crop species have a similar percentage of TFs in their genomes with 5.7% in wheat (*Triticum aestivum*), 6.5% in rice (*Oryza sativa*), and 6.1% in canola (*Brassica napus*) (Priya and Jain, 2013; Zheng et al., 2016; Evans et al., 2022). Fruit fly (*Drosophila melanogaster*) has a similar genome size as Arabidopsis, but the recent classification has revealed it has only ~5.5% TFs (Pfreundt et al., 2010; Shokri et al., 2019). Similarly, mouse (*Mus musculus*) and maize (*Zea mays*) have comparable genome sizes, while the former has ~6.8% TFs as compared to ~8.3% in maize (Jin et al., 2017; Zhou et al., 2017). The abundance, variety, and remarkable diversity of DNA-binding specificities exhibited by plant TFs, when compared to their counterparts in animals with similar genome sizes, suggest a potentially more important role for TFs and their transcriptional regulations in plants (Shiu et al., 2005; Mitsuda and Ohme-Takagi, 2009).

As the initial step in governing gene expression, transcriptional regulation has a direct impact on proteome, metabolome, and phenome. Because of the diverse roles of TFs, cells possessing the same genome within an organism can have different functions. TFs have been shown to play important roles in various growth and developmental processes. Morphogenesis-related processes such as light-controlled seedling morphology, the formation of floral traits, fruit morphology, as well as thermomorphogenesis, where the morphology changes under high temperatures, all have TFs' involvement (Sasaki, 2018; Shi et al., 2018; Han et al., 2019; Zhang et al., 2020a; Chopy et al., 2023). TFs also modulate organogenesis such as leaf, shoot and root development, or even nodule development in the symbiotic relationship between legume and rhizobium (Sluis and Hake, 2015; Tu et al., 2021; Chakraborty et al., 2022; Liu et al., 2022). Gene transcriptional regulation, facilitated by TFs, is essential in controlling numerous biological processes in plants, including signal transduction, stress and defense responses, as well as carbohydrate metabolism (Seo and Choi, 2015; Hoang et al., 2017; Shahzad et al., 2021; Strader et al., 2022; Zou and Sun, 2023).

2 Plant embryogenesis

Embryogenesis is a process where embryos form and develop. Plant embryogenesis starts from non-embryogenic cells. These non-embryogenic cells can be unfertilized egg cells for most flowering plants including crop species, but they can also be any cells that eventually develop into embryo-like structures and are capable of further developing into plants (de Vries and Weijers, 2017). The process of fertilizing egg cells to form zygotes and further develop into embryos under natural conditions is usually referred to as zygotic embryogenesis (ZE), while the process of embryos developing from any other cells without fertilization is usually referred to as somatic embryogenesis (SE). SE can occur naturally within an organism, as seen in cases like apomixis where embryos develop *in vivo* from unfertilized ovules, and parthenogenesis, where embryos develop from unfertilized egg cells. It is also not uncommon in plants that cells of different origins such as somatic cells or even microspores develop into embryos under inducive *in vitro* culture conditions.

Embryogenesis is a multi-stage process irrespective of whether it is ZE or SE. ZE involves a zygote developing through stages including 1-cell, 2-cell to octant, globular, heart, and finally a mature cotyledonary embryo for dicotyledon species (dicots) such as Arabidopsis (Figure 1). There are classic reviews regarding Arabidopsis ZE that readers can refer to including Capron et al., Wendrich et al., and ten Hove et al. (Capron et al., 2009; Wendrich and Weijers, 2013; Ten Hove et al., 2015). Zygotic embryos developed from monocots such as cereal crops wheat, rice and maize are morphologically different from those developed in dicots. However, developmental processes such as pattern formation and transcriptional regulation of ZE are conserved to a remarkable extent between monocots and dicots (Nardmann et al., 2007; Zhao et al., 2017). For readers interested in ZE in monocot species, we recommend the following comprehensive reviews of Vernoud et al., and Kruglova et al.(Vernoud et al., 2005; Kruglova et al., 2022). Despite originating differently, SE shares high similarities with ZE at morphological, physiological and molecular levels, and both processes share developmental stages like globular, heart, torpedo, and cotyledonary stages (Ikeda et al., 2006; Winkelmann, 2016). During SE, sometimes an intermediate stage involving embryogenic callus occurs, and this process is referred to as indirect SE (Figure 1). For readers interested in SE, the following comprehensive reviews of Smertenko and Winkelmann can be a good starting point (Smertenko and Bozhkov, 2014; Winkelmann, 2016).

Whether it is ZE or SE, the diverse origin of plant embryos signifies substantial changes occurring during embryogenesis. The acquirement of embryogenic competence and continued development requires the regulation and coordination of a myriad of genes, various gene networks and factors. Epigenetic control such as chromosome remodeling, transcriptional gene regulation and hormonal regulation have been associated with plant embryogenesis (Gulzar et al., 2020; Armenta-Medina et al., 2021); however, there is still a need for more work to fully understand the molecular mechanisms involved in plant embryogenesis. Such insights could open doors to opportunities like engineering crop embryos with desired traits/characteristics.

3 Transcription factors in plant embryogenesis

TFs play key roles during plant embryogenesis. From cell fate determination and apical-basal patterning initiation to embryonic

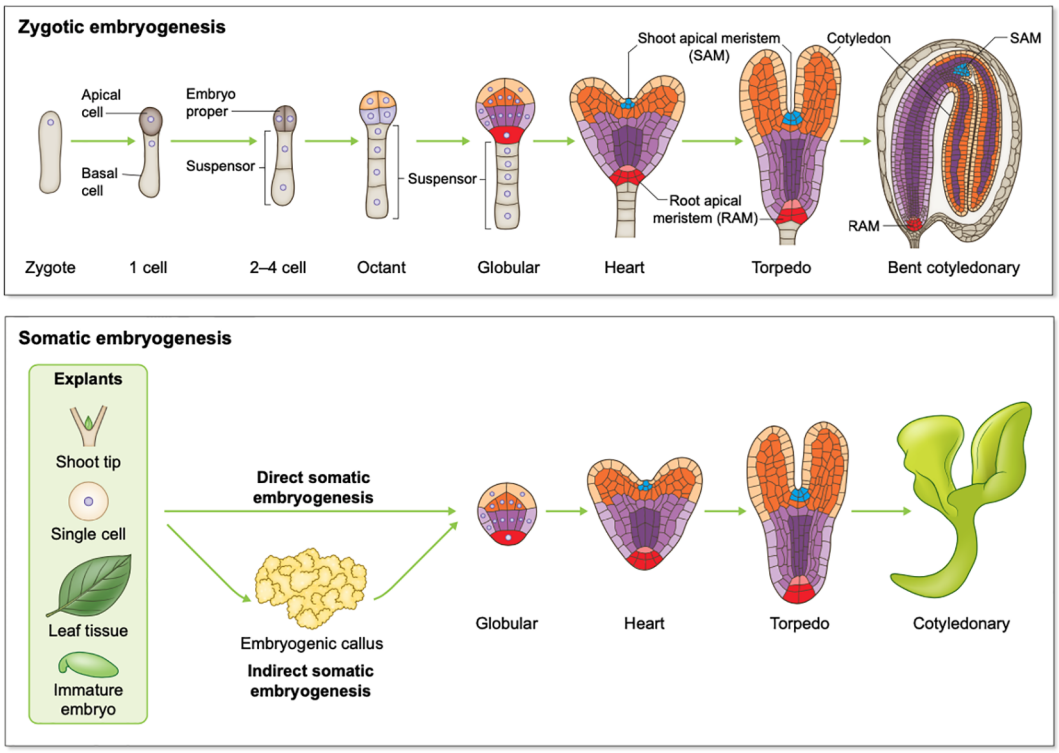


FIGURE 1
Schematic overview of plant embryogenesis. Top panel: Arabidopsis zygotic embryogenesis from the zygote to the bent cotyledonary stage; Bottom panel: somatic embryogenesis in dicots.

shoot, and root formation, various members of different TF families have been shown to be essential for these processes (Le et al., 2010; Horstman et al., 2017a; Méndez-Hernández et al., 2019; Gulzar et al., 2020; Gundu et al., 2020; Tian et al., 2020; Verma et al., 2022). TFs such as BABY BOOM (BBM) from the APETALA2/Ethylene-Responsive Factor (AP2/ERF) family, WUSCHEL and WUSCHEL-related homeobox (WOX) from the homeobox family, LEAFY COTYLEDON 2 (LEC2) from the B3 family are some of the master regulators that perform essential roles during plant embryogenesis. In this concise review, we present the latest developments in understanding the varied roles and functions of these master TFs in plant embryogenesis, along with a list of TFs that are capable of inducing somatic embryogenesis in different plant species (Table 1).

3.1 AP2/ERF family

The AP2/ERF family is one of the largest TF families in plants and plays an essential role in development processes and stress responses. Members within the AP2/ERF family share a common DNA-binding domain – AP2 domain and they are further categorized into 4 sub-families based on the copy number difference and the sequence variation of the domain (Gu et al., 2017). Within the AP2/ERF family, BABY BOOM/PLETHORA4 (BBM/PLT4) is the key player in plant embryogenesis. In addition, BBM-like genes including other PLT genes are increasingly recognized as important players in plant embryogenesis from recent studies.

The expression of Arabidopsis BBM is detected not only in the embryo starting as early as at the zygote stage, but also in the chalazal region of the ovule and the endosperm cells at the early initiating stage, suggesting a broader range of functions for BBM in both embryo and endosperm development in Arabidopsis (Chen et al., 2022a). BBM gene in Arabidopsis didn't show parent-of-origin expression patterns during embryo development and the analyses of single, and double mutants of BBM and PLT2 created using CRISPR-Cas9 showed that both genes coordinately work together in maintaining embryo development beyond the 4-cell stage, as well as regulating cell division planes and cell shapes (Chen et al., 2022a). Maize (*Zea mays*) BBM-like genes *ZmBBML1* and *ZmBBML2* showed induced expression 12 hours after pollination with the increased expression of *ZmBBML3* at 24 hours after pollination and after zygote division (Chen et al., 2017). The expression of rice BBM gene *OsBBM1* was highly induced during the initiation of embryogenesis (Anderson et al., 2017). In addition, rice BBM1 from the male parent initiated embryo development in the fertilized egg cell and subsequently stimulated BBM1 from the female parent to work together on embryo patterning, which is different from Arabidopsis (Khanday et al., 2018; Chen et al., 2022a). Wheat BBM homologs *TaBBM-gA* and *TaBBM-gD* were also induced in an embryogenic microspore population (Bilichak et al., 2018).

Earlier studies have shown that BBM overexpression induces somatic embryo development in many species including

TABLE 1 Transcription factors used to induce somatic embryogenesis in plants.

Transcription Factor	TF family	Plant Species	Approach	References
AGL15	MADS family	*Arabidopsis thaliana, Glycine max*	Genetic transformation	Thakare et al., 2008; Zheng and Perry, 2014,
AGL18	MADS family	*Glycine max*	Genetic transformation	Zheng and Perry, 2014,
BBM	AP2/ ERF family	*Arabidopsis thaliana, Brassica napus, Theobroma cacao, Capsicum annuum, Populus tomentosa, Nicotiana tabacum, Sorghum bicolor, Zea mays, Malus domestica, Oryza sativa, Gossypium hirsutum, Solanum Lycopersicon*	Genetic transformation, Gene editing	Boutilier et al., 2002; Srinivasan et al., 2007; Deng et al., 2009; Heidmann et al., 2011; Florez et al., 2015; Mookkan et al., 2017; Mookkan et al., 2017; Yavuz et al., 2020; Nelson-Vasilchik et al., 2022; Qi et al., 2022; Khanday et al., 2023; Xiao et al., 2023, Chen et al., 2022a
BBM/WUS2	AP2/ERF family/ Homeobox family	*Zea mays, Sorghum bicolor, Eragrostis tef, Panicum virgatum, Cenchrus americanus, Setaria italica, Triticum aestivum, Secale cereale, Hordeum vulgare, Saccharum officinarum, Oryza sativa*	Genetic transformation, gene editing	Lowe et al., 2018; Aregawi et al., 2020; Hoerster et al., 2020; Peterson et al., 2021; Aregawi et al., 2022; Beyene et al., 2022, Xu et al., 2022b, Nelson-Vasilchik et al., 2022; Johnson et al., 2023; Wang et al., 2023,
FUS3	B3 family	*Citrus unshiu, Arabidopsis thaliana*	Genetic transformation	Liu et al., 2018,
GRF	GRF family	*Zea mays*	Genetic transformation	Kong et al., 2020,
GRF-GIF	GRF family/ GIF family	*Triticum aestivum*	Genetic transformation, gene editing	Debernardi et al., 2020,
LEC1	NF-Y family	*Arabidopsis thaliana, Nicotiana tabacum, Manihot esculenta, Oryza sativa*	Genetic transformation, gene editing	Guo et al., 2013; Pelletier et al., 2017; Brand et al., 2019; Niu et al., 2021,
LEC2	B3 family	*Nicotiana tabacum, Manihot esculenta, Theobroma cacao*	Genetic transformation	Guo et al., 2013; Shires et al., 2017; Brand et al., 2019; Li et al., 2019,
RKD	RKD family	*Arabidopsis thaliana, Oryza sativa, Citrus sinensis*	Genetic transformation	Waki et al., 2011; Shimada et al., 2018; Purwestri et al., 2023,
WOX9	Homeobox family	*Medicago truncatula*	Genetic transformation	Tvorogova et al., 2019,
WOX2a	Homeobox family	*Zea mays*	Genetic transformation	Mcfarland et al., 2023,
WUS	Homeobox family	*Coffea canephora, Gossypium hirsutum, Medicago truncatula, Zea mays*	Genetic transformation	Arroyo-Herrera et al., 2008; Bouchabké-Coussa et al., 2013; Mookkan et al., 2017; Kadri et al., 2021,

Arabidopsis thaliana, canola (*Brassica napus*), cocoa (*Theobroma cacao*), sweet pepper (*Capsicum annuum*), poplar (*Populus tomentosa*), and tobacco (*Nicotiana tabacum*) as shown in Table 1 (Boutilier et al., 2002; Srinivasan et al., 2007; Deng et al., 2009; Heidmann et al., 2011; Florez et al., 2015). Overexpression of rice *OsBBM1* in the egg cell has been demonstrated to induce parthenogenesis, an embryo development process without fertilization. In contrast, the generation of triple knock-out mutants involving rice BBM1, BBM2, and BBM3 through gene editing resulted in the abortion of developing embryos during the early stage at 5 days after pollination (Khanday et al., 2018). CRISPR activation system targeting maize BBM2 (*ZmBBM2*) in egg cells also induced parthenogenesis (Qi et al., 2022). In a similar study, ectopic expression of *B. napus* BBM gene *BnBBM1* in the egg cells of Arabidopsis, canola, and tomato was able to induce haploid embryos in all three species (Chen et al., 2022a). *PsASGR-BABY BOOM-like* (*PsASGR-BBML*) gene, a member of the BBM-like AP2/ERF transcription factors, originates from an apomictic species of pearl millet, *Pennisetum squamulatum* (Conner et al., 2015). When introduced as a transgene, the *PsASGR-BBML* has been shown to induce parthenogenesis in sexual tetraploid pearl millet, maize, rice, and even tobacco, a dicot species (Conner et al., 2015; Conner et al., 2017; Zhang et al., 2020b). Most recently, overexpression of any of the three BBM genes from foxtail millet (*Setaria italica*): *SiBBM1, SiBBM2* and *SiBBM3* in egg cells induced parthenogenesis in rice (Chahal et al., 2022). Overexpression of *MdBBM* promoted somatic embryogenesis in cultured young leaves of apple (*Malus domestica*) (Xiao et al., 2023). This further confirms the conservation of BBM functions among monocot and dicot species.

Elevated expression of BBM-like PLT gene members, such as *PLT1, PLT2, PLT3*, and *PLT7*, has been demonstrated to induce somatic embryogenesis. Notably, *PLT1* and *PLT3* were found to recover early-stage embryo defects of *plt2 bbm* mutants, suggesting a redundant and overlapping role of BBM and BBM-like genes in

embryogenesis (Horstman et al., 2017b; Kerstens et al., 2022). In addition, BBM and PLT2-induced somatic embryogenesis is dosage-dependent with higher levels of BBM/PLT2 resulting in the initiation of embryogenesis (Horstman et al., 2017b). This also aligns with previous research indicating that transcription factors have maintained a dosage-dependent pattern following historical polyploidization events (Birchler and Veitia, 2007). The broad applications and evidence of BBM and BBM-like genes in both zygotic and somatic embryogenesis show that they play important roles in plant embryogenesis.

3.2 Homeobox family

The homeobox family is a large family of TFs with a DNA-binding domain called homeodomain (HD) and plays important roles in various development processes including organism differentiation as well as increasing developmental complexity (Mukherjee et al., 2009; Lutova et al., 2015; Jha et al., 2020). Within the family, WUSCHEL (WUS) and WUSCHEL-LIKE HOMEOBOX (WOX) are key players in plant embryogenesis.

WUS has been reported to regulate stem cell fate in Arabidopsis as early as the late twentieth century (Mayer et al., 1998). Similarly, studies have shown that after zygote division, embryo development is directed by the apical-cell-expressed WOX2, while basal-cell-expressed WOX8 governs suspensor development and root initiation, as well as regulating WOX2 expression during early embryogenesis (Breuninger et al., 2008). The enrichment of WOX8/WOX9 in the basal cells was further confirmed from Arabidopsis early embryos at the single-cell level (Kao et al., 2021). Members of the WOX family play specific roles during embryogenesis such as embryonic development, preservation of meristematic stem cells, formation of lateral organs, seed production, and regeneration of separated tissues and organs (Zhang et al., 2016; Jha et al., 2020). Recent work shows that WOX8 expression is further controlled by both paternal and maternal regulators (Ueda et al., 2017). As a master regulator of embryogenesis, WOX8 integrates the signals from both maternal and paternal factors to regulate embryo patterning through the initiation of the asymmetric division of the zygote.

The expression patterns of WUS and WOX5 are conserved in shoot and root apical meristem during early embryogenesis as demonstrated through the analysis of 13 WOX family members from the tobacco (*Nicotiana tabacum*) genome as compared to their counterparts in Arabidopsis (Zhou et al., 2018). The stage-specific expression pattern was identified in tobacco with the expression of five WOXs (WOX2, WOX9, WOX11, WOX13a, and WOX13b) started as early as 2-cell proembryo stage, while WUS and WOX5 only started at 8-cell embryo stage. In addition, the analysis revealed that WOX genes in tobacco displayed parent-of-origin effects, and the formation of embryo pattern is established post-fertilization involving the expression of WOX2 and WOX9 in the zygote, which differs from the situation in Arabidopsis. However, cell-type specific expression patterns of WOXs in the apical/basal cells were conserved between maize, tobacco and Arabidopsis (Chen et al., 2017; Zhou et al., 2018). In addition, WOX8/9 was identified as the most prevalent suspensor-specific TF at globular embryos and had a similar role in suspensor development in scarlet runner bean (*Phaseolus coccineus*), common bean (*Phaseolus vulgaris*), and soybean (*Glycine max*), similar to Arabidopsis (Chen et al., 2021). The rice orthologs of Arabidopsis WOX8/9 and WOX2 show increased expression at 2.5 and 5 hours after pollination, indicating their roles during the early stage of embryogenesis (Anderson et al., 2017). In addition, the parent of origin for WOX8/9 expression was solely paternal (Anderson et al., 2017). Similarly, Maize WOX genes *ZmWOX9A* and *ZmWOX9B*, homologs of Arabidopsis WOX8 and WOX9, showed increased expression at 12 hours after pollination and had higher expression in basal cells as in Arabidopsis (Chen et al., 2017).

Subsequently, it has been shown that embryogenic stem cell regeneration relies on WUS during somatic embryogenesis in Arabidopsis (Su et al., 2009). WUS expression was upregulated in embryogenic calli before somatic embryos developed. *HvWUS* exhibits higher expression when immature embryos are used as the explants compared to mature embryos, thus facilitating the induction of embryogenic callus formation in barley (*Hordeum vulgare*) (Suo et al., 2021). Overexpression of the Arabidopsis WUS gene promotes somatic embryogenesis in various plant species including coffee (*Coffea canephora*), cotton (*Gossypium hirsutum*), and barrel clover (*Medicago truncatula*) (Arroyo-Herrera et al., 2008; Bouchabké-Coussa et al., 2013; Kadri et al., 2021). The inclusion of the WUS gene within the gene transformation cassette promotes somatic embryogenesis of both leaf and hairy root explants in barrel clover without the need for plant growth regulators. Additionally, it enables the transformation of the historically recalcitrant maize and sorghum (*Sorghum bicolor*) genotypes using immature embryos as explants (Mookkan et al., 2017; Aregawi et al., 2020; Hoerster et al., 2020; Kadri et al., 2021). Furthermore, when under the control of tissue-specific, and auxin-inducible promoters, the BBM/WUS2 transgene triggers swift and direct somatic embryogenesis in maize in a genotype-independent manner (Lowe et al., 2018). The co-expression of maize BBM and WUS genes also improves somatic embryogenesis in recalcitrant maize and sorghum genotypes (Mookkan et al., 2017; Nelson-Vasilchik et al., 2022).

LdWOX2, a homolog of Arabidopsis WOX2 in European larch (*Larix decidua*), was highly expressed during early embryogenesis in somatic embryos (Rupps et al., 2016). Three different barrel clover lines were used to test the effects of overexpression of *MtWOX9-1* using leaves as the explants and the results showed *MtWOX9-1* promoted somatic embryogenesis as well as led to the expression changes of two embryogenesis-associated MADS-box genes (Tvorogova et al., 2019). Similarly, WOX2a promotes somatic embryogenesis in recalcitrant maize genotypes (Mcfarland et al., 2023). Overexpression of the WOX2a gene from an embryogenic maize genotype A188 produced somatic embryos from a recalcitrant genotype B73. In addition, the overexpression of the WOX2a gene from B73 had a similar effect on somatic embryo production. The key roles of both WUS and WOXs in somatic embryogenesis underscores their significant contributions to plant embryogenesis.

3.3 B3 family

The B3 family is among the most extensive plant-specific transcription factor families. Its members may feature a single B3 domain as in ARF (AUXIN RESPONSE FACTOR) and LAV (LEAFY COTYLEDON2-ABSCISIC ACID INSENSITIVE3-VAL) subfamilies, or they can possess as many as six B3 domains as observed in REM (REPRODUCTIVE MERISTEM) subfamily (Swaminathan et al., 2008). RAV (RELATED TO ABI3/VP1) subfamily is unique in that some members within the subfamily contain both a B3 domain and an AP2 domain, and the binding sites need to have sequences for both domains.

Members within the LAV subfamily of B3 transcription factors such as ABA INSENSITIVE3 (ABI3), FUSCA3 (FUS3), LEAFY COTYLEDON 2 (LEC2), VIVIPAROUS1/ABI3-LIKE1(VAL1), VAL2 and VAL3 function as regulators of embryogenesis and are part of the embryogenesis-related genes in Arabidopsis (Swaminathan et al., 2008; Carbonero et al., 2017; Kumar et al., 2020). An earlier investigation has demonstrated that Arabidopsis LEC2 is involved in embryonic cell fate maintenance during early embryogenesis, as well as the initiation of embryo maturation during the late embryogenesis stage (Stone et al., 2008). Within the LAV subfamily, AFL (ABI3/FUS3/LEC2) plays an important role in embryogenesis and the studies on the effects of individual AFL loss-of-function mutants revealed the redundancy of their function on the establishment of embryo morphology, more specifically, the cotyledon shape and bending (Devic and Roscoe, 2016). Spatiotemporal regulation of FUS3 expression in ovule integuments and endosperm ensures a coordinated embryo and endosperm growth (Wu et al., 2020). Through a single-nucleus RNA-seq study of Arabidopsis early embryos, FUS3 was shown to be highly expressed in the basal cell controlling suspensor development and root initiation among other TFs (Kao et al., 2021).

Five AFL orthologs were identified in maize with *ZmAFL2*, a FUS ortholog, *ZmAFL3/ZmVP1*, an ABI3 ortholog and *ZmAFL4, ZmAFL5, ZmAFL6*, orthologs of LEC2 (Grimault et al., 2015). *ZmAFL* genes consecutively expressed at different stages, with the peak expression of *ZmAFL2, ZmAFL5*, and *ZmAFL6* as early as 3 days after pollination (DAP), while *ZmVP1* reached peak expression at 35 DAP during zygotic embryogenesis. Maize AFLs also showed distinct spatial expression patterns with *ZmAFL2* and *ZmVP1* specifically expressed in the embryos, while *ZmAFL4* was mostly expressed in the endosperm. Though there is a difference in spatial gene expression of AFL genes between Arabidopsis and Maize, the sequential expression of FUS3 and ABI3 is maintained. Barley ABI3 ortholog *HvVP1* expressed in both embryos and endosperms and showed higher expression at the immature embryo stage (20 DAP) (Abraham et al., 2016). Both *MtFUS3-like* and *MtABI3-like* showed the highest expression at the torpedo stage embryos during zygotic embryogenesis in barrel clover, and *MtABI3-like* expressed throughout the embryo proper, but not in the suspensor (Kurdyukov et al., 2014).

As for Arabidopsis embryogenesis *in vitro*, FUS3 and ABI3 were significantly upregulated in early-stage somatic embryos (Day 5 and Day 10 after the induction), and ABI3, FUS3 and LEC2 were highly expressed in somatic embryos as compared to leaf tissues in a global scale transcriptomic study (Wickramasuriya and Dunwell, 2015). In addition, ectopic expression of LEC2 induces somatic embryogenesis and the LEC2-overexpressing explants show higher somatic embryo induction under low auxin concentration (Wójcikowska et al., 2013). Three AFL genes from barrel clover *MtLEC2, MtFUS3*, and *MtABI3* show higher expression 10 days after initiating the culture in an embryogenic genotype as compared to no or low expression in a non-embryogenic genotype (Barreto et al., 2019). *CaABI3* is highly expressed in embryogenic cells/calli and could be used as a biomarker for somatic embryogenesis in coffee, while *CaVAL2* shows higher expression in cotyledonary embryos, a later stage in embryogenesis (Freitas et al., 2019). In addition, FUS3 is expressed all through somatic embryogenesis in coffee with the highest expression in globular embryos (Awada et al., 2023). Overexpression of the *CsFUS3* gene from sweet orange (*Citrus sinensis*) promotes somatic embryogenesis in recalcitrant Satsuma mandarin (*Citrus unshiu*) and restores embryogenesis in Arabidopsis *fus3* mutants (Liu et al., 2018). In addition, three other B3 TFs (*CsABI3, CsABI5*, and *CsVAL1*) show significantly higher expression in the *CsFUS3*-overexpression lines as compared to the control in sweet orange. Overexpression of LEC2 ortholog induced somatic embryogenesis from leaf explants in cocoa (Shires et al., 2017), while a single *MeLEC2* overexpression was able to induce somatic embryogenesis in cassava (*Manihot esculenta*) (Brand et al., 2019). Though copy numbers of AFL genes vary among monocot and dicot species, their significance in plant embryogenesis remains consistent.

3.4 MADS family

The MADS transcription factor family is an ancient group of transcription factors with a considerably expanded number of family members in plants, which share a core DNA-binding domain: MADS domain (Gramzow and Theissen, 2010). Initially identified as a major factor in controlling flower development, the MADS TF family has been shown to play important roles in various developmental processes including pollen and embryo sac development, seed development and fruit development (Theißen and Gramzow, 2016). Within MADS TFs, AGAMOUS-Like 15 (AGL15) and AGL18 are two key members in plant embryogenesis.

Previous immunohistochemical studies have shown that AGL15 is highly expressed in embryogenic tissues derived from various sources, including Arabidopsis and maize zygotic embryos, dandelion (*Taraxacum officinale*) apomixis embryos, canola microspore-developed embryos, as well as alfalfa (*Medicago sativa*) somatic embryos (Perry et al., 1999). This indicates a much-conserved role for AGL15 in embryo development across plant species. AGL15 has been shown to directly target AFL (ABI3/FUS3/LEC2) genes, the key regulators of embryogenesis, through a genome-wide binding-site identification study (Zheng et al., 2009). Among the MADS genes, 23 were specifically expressed in embryogenic tissues of Arabidopsis, with several of them displaying differential expression across various stages of somatic embryo development (Wickramasuriya and Dunwell, 2015). Among these genes, AGL15 exhibited higher expression levels in

Arabidopsis somatic embryos compared to leaf tissues. Mutations in the rice MADS gene FEMALE-STERILE (FST) lead to a complete embryo abortion (Lee et al., 2013). Analyses of gene expression changes in *fst* mutants show that genes involved in auxin transportation, cell differentiation, and embryogenic development were down-regulated.

Arabidopsis AGL18 and AGL15 have redundant functions in promoting somatic embryogenesis (Paul et al., 2022). They interact in somatic embryo tissues and the phosphorylation of both AGL15 and AGL18 is pivotal to the process through combined chromatin immunoprecipitation (ChIP-seq) and RNA-seq studies. In addition, AGL15 transgenic seedlings showed increased somatic embryo induction at 64.4%, and AGL18 transgenic seedlings had a 40.8% induction rate as compared to a 19.8% induction rate in wild-type seedlings. In a more recent study, MADS gene AGL62 was identified as the gene responsible for activating small invertase inhibitors in the syncytial endosperm and controlling the rate of embryo growth in Arabidopsis (Hoffmann et al., 2022).

MtAGL15 expression increased in *MtWOX9-1* overexpressing calli which showed increased somatic embryogenesis capacity (Tvorogova et al., 2019). Coffee AGL15 is expressed during the early stages of somatic embryogenesis, albeit at a relatively modest level (Awada et al., 2023). An earlier study shows that overexpression of Arabidopsis AGL15 promotes somatic embryo development in both Arabidopsis and soybean (Thakare et al., 2008). Constitutive expression of soybean *GmAGL15* or *GmAGL18* sped up and increased somatic embryogenesis in soybeans, with ABI3 and FUS3 directly upregulated by *GmAGL15* during the process (Zheng and Perry, 2014). Three AGL15 homologs *GhAGL15-1, GhAGL15-3* and *GhAGL15-4* were isolated from cotton and their expressions were found to be increased during the somatic embryogenesis process (Yang et al., 2014). Overexpression of any of the three *GhAGL15s* in hypocotyls dramatically increased embryogenic callus formation, with *GhAGL15-4* having the highest increase in formation rate from 38.1 to 65.2%. The AGAMOUS subfamily of MADS TFs has long been shown to be a key player in embryogenesis, however, there is a lack of recent progress, particularly in understanding the functions of their corresponding homologs in monocot species.

3.5 NF-Y family

The Nuclear Factor Y (NF-Y) family is yet another transcription factor family that has much-expanded family members in plants. NF-Y TFs are also called Heme-associated proteins (HAPs) or CCAAT box binding factors (CBFs) and they form a heterotrimeric complex with one single subunit each of NF-YA, NF-YB and NF-YC to bind at CCAAT sites to regulate various developmental processes (Petroni et al., 2012). Initial research has indicated the significance of LEAFY COTYLEDON gene (LEC1, NF-YB9) in both early and late embryo development regulating processes such as embryonic cell fate, cotyledon identity and embryo maturation (Lotan et al., 1998). However, a growing body of evidence suggests that additional members within the family including LEC1– like (LIL, NF-YB6) also play crucial roles (Kwong et al., 2003; Fornari et al., 2013; Mu et al., 2013).

Eight of ten Arabidopsis NF-YA genes show expression in embryos (Siriwardana et al., 2014). Arabidopsis NF-YA3 and NF-YA8 are functionally redundant and are required in early embryogenesis (Fornari et al., 2013). *In situ* hybridization showed that both NF-YA3 and NF-YA8 have the highest expression during the early embryo development stages. Embryos from *nf-ya3 nf-ya8* double mutants or RNAi suppression exhibit defects and do not progress to the heart stage. Arabidopsis NF-YA1, 5, 6, and 9 have redundant roles in multiple developmental processes including male gametophyte development, embryogenesis, and seed development (Mu et al., 2013). NF-YA1, 5, 6, and 9 co-express with LEC1 in similar developmental windows from the early heart stage to the late torpedo stage and nf-ya1 mutants show embryo development defects at the early heart stage. Among the ten Arabidopsis NF-YA genes, NF-YA1 and NF-YA9 are most similar, as well as NF-YA5 to NF-YA6. The similarity of functions among LEC1 and LIL (NF-YB subunit), and NF-YA1, 5, 6, and 9 suggests they may be part of the NF-Y heterotrimeric complex involved in the regulation of embryogenesis. A recent study demonstrated that Arabidopsis LIL/NF-YC3/NF-YA6 trimers specifically bind to the CCAAT motif, but not the separate subunit (Gnesutta et al., 2017). This further shows that LEC1 and LIL, both NF-YB subunits, need to form a heterotrimeric complex with the other two NF-Y subunits (NF-YA and NF-YC) to function properly.

RNA-seq study of developing seeds from Arabidopsis *lec1-1* mutant shows that genes affected by *lec1-1* mutation are mostly expressed in embryo and endosperm, and expressions of TFs such as BBM, WOX2, FUS3, and ABI3 were down-regulated in *lec1-1* mutants (Pelletier et al., 2017). Soybean has four LEC1 paralogs: *GmLEC1-1, -2, -3*, and *-4*, with the first two more closely resembling Arabidopsis LEC1 in terms of their expression patterns. Chip-seq study using developing soybean embryos from various development stages including morphogenesis, transition and maturation stages suggested a functional conservation of LEC1 between Arabidopsis and soybean (Pelletier et al., 2017).

Overexpression of NF-YA1, 5, 6, and 9 can induce somatic embryogenesis from Arabidopsis seedlings, while overexpression of LEC1 and LIL leads to induced expression of NF-YA1, 5, and 9 (Mu et al., 2013). Arabidopsis plants overexpressing canola *BnLEC1* and *BnLIL* genes showed similar phenotypes as to Arabidopsis *AtLEC1* overexpression plants, suggesting the similarity of their functions (Mu et al., 2008). Overexpression of *AtLEC1* can induce embryonic transition in tobacco seedlings (Guo et al., 2013). In both zygotic and somatic embryos of the European larch, *LdLEC1* was highly expressed during early embryogenesis (Rupps et al., 2016). LEC1 was expressed during the induction phase of somatic embryogenesis and was linked to the embryogenic development in somatic cells in barrel clover (Orłowska et al., 2017). In callus tissue overexpressing *MtWOX9-1*, which exhibits enhanced somatic embryogenic potential, *MtNF-YB10*, a barrel clover homolog of Arabidopsis LEC1, displayed elevated and co-related expression with *MtWOX9-1* (Tvorogova et al., 2019). Rice LEC1 homologs *OsNF-YB9* and *OsNF-YB7* show different expression patterns with the

former mainly expressed in endosperm, while the latter in embryo (Niu et al., 2021). However, overexpression of either *OsNF-YB9* or *OsNF-YB7* can restore Arabidopsis *lec1-1* mutants with more than 69.4% of seeds produced showing normal embryo morphology, suggesting a functional conservation of LEC1 between monocots and dicots. In addition, CRISPR/Cas9 knock-out of *OsNF-YB7* is fatal due to abnormal embryo development, providing further confirmation of its importance in rice embryogenesis. Though more research is needed to understand the organization of heterotrimeric complexes within the NF-Y subunits, it is evident that the NF-Y TF family plays a crucial role in plant embryogenesis.

3.6 Other TF families

In addition to the above-mentioned TFs, several others have been identified to have significant roles in plant embryogenesis. RWP-RK DOMAIN–CONTAINING PROTEIN (RKD) TFs are plant-specific transcription factors and five RKD genes were identified in Arabidopsis, namely *AtRKD1, AtRKD2, AtRKD3, AtRKD4*, and *AtRKD5* (Köszegi et al., 2011). Gene expression analysis shows that four of the Arabidopsis RKD genes *AtRKD1* to *AtRKD4* are mostly expressed in the reproductive tissues, with *AtRKD1* and *AtRKD2* mostly in egg cells, while *AtRKD4* in early embryos. Loss of function *rkd4* mutants fail to develop embryos properly and overexpression of RKD4 induces somatic embryogenesis (Waki et al., 2011). In addition, overexpression of *AtRKD1* and *AtRKD4* was able to induce somatic embryogenesis from suspensor cells (Radoeva et al., 2020). Maize RKD gene *Shohai1(Shai1)*, a close ortholog to *AtRKD5*, was shown to play an important role in embryo and endosperm development (Mimura et al., 2018). Loss of *Shai1* function leads to embryo defects, and the *shai1* mutant embryos can be partially rescued by overexpressing *Shai1* in the endosperm. *OsRKD3*, one of the rice RKD genes and the one that is closely related to *AtRKD4*, induces somatic embryogenesis in black rice (*Oryza sativa*) (Purwestri et al., 2023). A RKD homolog from Indonesian local pigmented rice (*Oryza sativa*), which is similar to *OsRKD*, is induced during the early stage of microspore embryogenesis (Nurbaiti et al., 2021). *CitRKD1* from satsuma mandarin was confirmed to be the candidate gene leading to somatic embryogenesis in citrus and transgenic sweet oranges with *CitRKD1* loss-of-function did not succeed in generating somatic embryos (Shimada et al., 2018). These studies show a conserved function of RKD genes in embryogenesis across various plant species.

GROWTH-REGULATING FACTORs (GRFs) are plant-specific transcription factors as well and multiple GRF genes have been identified in various species, with 9 GRFs in Arabidopsis, 13 in rice, 17 in maize and 26 in soybean (Omidbakhshfard et al., 2015). GRFs play important roles in various developmental processes including seed development with GRF1 or GRF5 overexpressing lines producing larger seeds in Arabidopsis (Van Daele et al., 2012). Overexpression of *AtGRF5* or respective GRF5 orthologs improved regeneration and transformation efficiency in both monocots such as maize and dicot species such as sugar beet (*Beta vulgaris ssp. Vulgaris*) including some recalcitrant sugar beet genotypes, canola, and soybean (Kong et al., 2020). Ectopic expression of *ZmGRF5-like1* and *ZmGRF5-like2* improved embryogenic callus growth as well as transformation efficiency in maize. GRFs and GRF-INTERACTING FACTORs (GIFs), a group of transcription cofactors, form a transcriptional complex to function and the complex is required for the development of meristematic and pluripotent cells (Lee et al., 2018). Overexpression of a wheat *TaGRF4* and *TaGIF1* chimeric protein improves embryogenesis and transformation efficiency in wheat, triticale and rice (Debernardi et al., 2020). The homologs of wheat *TaGRF4* and *TaGIF1* in citrus and grape (*Vitis vinifera*) were also used to generate a citrus GRF4-GIF1 chimera and a grape GRF4-GIF1 chimera. Both chimeras also increased transformation efficiency in citrus. Most recently, a new transformation system named GGB (GRF-GIF-BBM) utilizing *TaGRF4-GIF1* and *ZmBBM* was shown to increase the transformation efficiency in genome-edited maize genotypes with different genetic backgrounds (Chen et al., 2022b). It appears that the function of the GRF-GIF complex is conserved across monocot and dicot species, though the GRF-GIF complex mostly improves pluripotency (organogenesis), while it has a limited impact on totipotency (embryogenesis).

4 Transcription factor network during plant embryogenesis

Plant embryogenesis is a complex developmental process characterized by significant transformations, which require coordinated regulations of various genes, and transcription factors. It is evident that the key TFs we have emphasized in the preceding sections do not function in isolation during this process. Instead, they engage in active interactions with one another, as well as with their target genes, to establish complex transcriptional networks. These networks of TFs ensure the sequential and orderly progression of every stage of embryogenesis encompassing cellular reprogramming, patterning formation, and differentiation of meristematic tissues.

An *OsBBM1-OsYUC* module has been shown to play an important role in both zygotic embryogenesis and somatic embryogenesis in rice (Khanday et al., 2023). Paternal-genome-originated *OsBBM1* directly triggers maternal-genome-originated auxin biosynthesis gene *OsYUCCA* (*OsYUC*) to initiate zygotic embryo development; while under the culture conditions, exogenous auxin induces *OsBBM1*, which then activates endogenous *OsYUC* genes (*OsYUC6*, *OsYUC7* and *OsYUC9*) to promote somatic embryogenesis (Khanday et al., 2018; Khanday et al., 2023). A two-step model has also been proposed that cell totipotency is established through BBM-induced gene expression such as LEC1, LEC2, and FUS3, and then the induction of auxin biosynthesis is required for the maintenance of embryo identity and embryo development (Li et al., 2022). Auxin plays a fundamental role in pattern formation during embryogenesis. The above-mentioned master TFs such as BBM, LEC1, LEC2, and FUS3 have been reported to activate auxin biosynthesis genes such as

YUCCA genes (YUCs) and there are complex feedback loops among these master TFs during embryogenesis (Kagaya et al., 2005; To et al., 2006; Yamamoto et al., 2010; Junker et al., 2012; Wójcikowska et al., 2013; Tian et al., 2020). A similar scenario of a coordinated TF regulatory network among these master TFs occurs in the context of Arabidopsis somatic embryogenesis. BBM and PLT2 activate the LAFL (LEC1-ABI3-FUS3-LEC2) network to induce somatic embryogenesis from Arabidopsis seedlings (Horstman et al., 2017b), while LEC2 promotes somatic embryogenesis through direct activation of WOX2 and WOX3 (Wang et al., 2020).

The regulatory network of MADS-domain TFs showed that AGL15 suppresses SEPALLATA3 (SEP3), but induces several important embryogenesis-related genes including FUS3, ABI3, as well as a gibberellin (GA) oxidase gene GA2ox6 while interacting with BBM and LEC1 (Wang et al., 2004; Zheng et al., 2009; Paul et al., 2022). Moreover, AGL18 induces AGL16, LEC1, PLETHORA2 (PLT2), and ABI4, while both AGL15 and AGL18 suppress GA3ox2, a gibberellin (GA) biosynthetic gene, to promote somatic embryogenesis (Paul et al., 2022). Earlier studies have shown that LEC2 and FUS3 negatively affect GA biosynthesis by repressing GA3ox2 expression, thus regulating the embryonic development (Curaba et al., 2004; Gazzarrini et al., 2004). The activation of auxin biosynthesis genes and the suppression of GA biosynthesis during embryogenesis further implies the importance of hormonal regulation during the process, however, delving into this aspect is beyond the scope of this review. Through the combination of ChIP-seq and RNA-seq studies, 283 genes were found to be induced and 472 genes were repressed in AGL15-overexpressed embryogenic cultures as compared to wild types (Joshi et al., 2022a). The relationship between AGL15 and other transcription factors, hormone genes, and genes involved in epigenetic modification, suggests a more complex network interaction in embryogenesis (Joshi et al., 2022b).

A BPC1 (Basic Pentacysteine 1)-FIS (Fertilization-independent Seed)-PRC2 (Polycomb Repressive Complex 2) network controls the spatiotemporal expression of FUS3 to coordinate embryo and endosperm growth (Wu et al., 2020). A number of embryogenic-related transcription factors have also been shown to be repressed by PRC 1 and 2 including WOX5, WOX8, AGL15, LEC1, LEC2, FUS3, ABI3, and BBM (Makarevich et al., 2006; Chen et al., 2010; Bouyer et al., 2011; Ikeuchi et al., 2015; Duarte-Aké et al., 2019). Polycomb Repressive Complexes (PRC 1 and 2) are one of the key components in epigenetic gene regulation and PRC2 leads to the trimethylation of histone H3 (H3K27me3) which will then activate PRC1-induced ubiquitination of histone H2A (H2Aub), processes that make chromatin transcriptionally inactive therefore repressing gene expression (Mozgova et al., 2015; Baile et al., 2022). As members of the B3 TF family VAL genes were initially identified in sugar signalling as High-level expression of sugar-inducible gene2 (HSI2/VAL1) and HSI2-LIKE1 (HSL1/VAL2) and expression of LEC1, LEC2, and FUS3 were induced in *hsi2 hsl1* mutants (Tsukagoshi et al., 2007). Chromatin immunoprecipitation shows that HSI2/VAL1 binds to AGL15 to repress its expression (Chen et al., 2018). Furthermore, increased evidence suggests that VAL genes interact, and recruit subunits of PRC for gene silencing as transcriptional repressors (Duarte-Aké et al., 2019; Yuan et al., 2021; Baile et al., 2022). However, as histone-combinational-binding effector proteins, GRFs recognize both histone phosphorylation and histone trimethylation at specific sites (H3K28ph and H3K27me3) to reverse the repression effects caused by these histone modifications (Zhao et al., 2018; Xu et al., 2022a). In addition, microRNAs also mediate transcriptional and post-transcriptional silencing of genes involved in plant development and microRNAs are mostly induced in non-embryogenic cultures than the embryogenic ones (Wu et al., 2015). MicroRNAs such as miR172 and miR1160, were shown to repress WUS and AP2/ERF TFs during the embryogenesis as well (Wu et al., 2015; Li et al., 2017; Nowak et al., 2022).

During early embryogenesis in rice, the spatiotemporal expression of multiple TFs including C2C2, homeobox, MADS, bHLH, and NAC showed preferential and divided expression patterns in distinct embryonic organs and domains (Itoh et al., 2016). These differences in the spatial expression of TFs suggested that the roles of TF family members in the initial patterning of the embryo and the arrangement of the embryonic organs were established at the early globular embryo stage (Itoh et al., 2016). The functions of WRKY2, HDG11 (HOMEODOMAIN GLABROUS 11), and WOX8/9 on suspensor development during early embryogenesis are conserved in plants and the activation of WOX8/9 is dependent on WRKY2 and HDG11 in addition to the feedback loop between WOX9 and HDG11 (Chen et al., 2021). RKD genes have been shown to induce a number of genes including AP2/ERFs, and MYBs during embryogenesis (Waki et al., 2011; Purwestri et al., 2023). Four TF classes, namely TEOSINTE BRANCHED1-CYCLOIDEA-PCF transcription factors (TCPs), Auxin response factors (ARFs), MYBs, and WOXs, were identified as playing central roles in the transcriptional regulation network during pattern formation of wheat zygotic embryos and a regulatory module involving LEC1-MYB118-ZHD5-LEC2-BBM was confirmed (Zhao et al., 2023). Briefly, LEC1, MYB118, and ZHD5 were induced first during the pro-embryo stage, followed by LEC2 to reach the peak at the transition stage, and then BBM at the mid-embryo stage. These spatial and sequential events show a coordinated TF regulatory network in charge of early embryogenesis.

A hierarchical transcriptional regulatory network for somatic embryogenesis was also identified through a combinational approach utilizing ATAC-seq, ChIP-seq, RNA-seq, and genetic transformation of immature Arabidopsis embryos (Wang et al., 2020). The sequential TF functions included bHLH and BES1 (BRI1-EMS-SUPPRESSOR1) TFs at the early stage of somatic embryogenesis, WRKY and CAMTA1 (calmodulin binding transcription activator 1) TFs from 0 to 8 hours after the initiation of the process, and then ARF, AP2, B3, and TCP (TEOSINTE BRANCHED1/CYCLOIDEA/PCF) TFs including BBM and LEC2 after 24 hours after the initiation. This sequential action of various TFs indicates a well-organized transcriptional regulatory network during somatic embryogenesis, underscoring

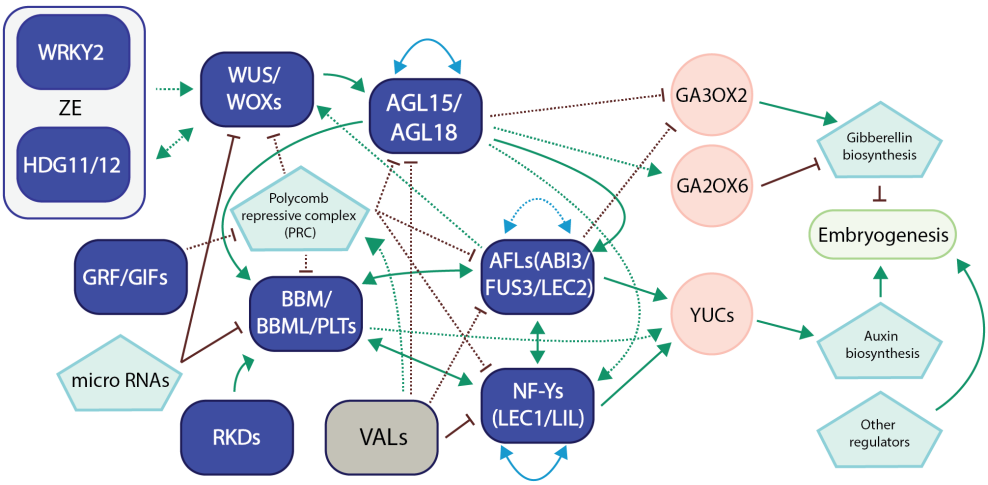


FIGURE 2
A representative model of transcription factor networks during plant embryogenesis. This model summarizes our recent research and the latest developments in the field regarding transcription factor networks in embryogenesis in Arabidopsis and other crop species (Horstman et al., 2017a; Pelletier et al., 2017; Ueda et al., 2017; Chen et al., 2018; Kong et al., 2020; Tian et al., 2020; Wang et al., 2020; Chen et al., 2022a; Kerstens et al., 2022; Paul et al., 2022; Khanday et al., 2023; Purwestri et al., 2023; Zhao et al., 2023). Transcription factors in purple usually promote embryogenesis, while TFs in gray suppress the process. For TF network interaction, solid lines refer to gene regulation, while dotted lines refer to protein-protein interaction or both. Green arrows refer to direct activation, while red bars refer to suppression. Blue lines with arrows on both ends refer to autoregulation/self-regulation or regulation between group members. Gray regions were used to show TF networks that have only been verified/proposed in zygotic embryogenesis (ZE). For the sake of clarity, we have included only the most significant networks and references.

the significance of transcription factor networks during plant embryogenesis. We have proposed a model for the network interactions involving these master transcription factors in plant embryogenesis, drawing upon recent advancements and our own research (Figure 2).

5 Future perspectives and conclusion

Cells undergo a series of intricated processes to develop into embryos with specific spatial organization, therefore the position of each cell within an embryo is critical for its function. Spatial-temporal single-cell transcriptomics can uncover gene expression of individual cells in various locations within an embryo throughout the course of embryogenesis at an unparalleled resolution. Therefore, spatiotemporal single-cell gene expression study during embryogenesis could be a valuable resource for functional analyses of transcription factors and a foundation for comparative studies of plant embryogenesis.

Overexpression of an individual transcription factor can lead to somatic embryogenesis in various plant species. Some of these master TFs include AP2 TF BBM/PLTs, Homeobox TF WUS/WOXs, B3 TF LEC2/FUS3, NF-Y TF LEC1, and RKD4 as described above. The combinational usage of two TFs such as BBM and WUS, or even more, such as the GGB (GRF-GIF-BBM) system, not only stimulates somatic embryogenesis in recalcitrant genotypes under hormone-free culture conditions but also dramatically speeds up the embryogenesis process. Could functional and mechanistic studies reveal additional master TFs? A better understanding of the functions and molecular mechanisms of these master TFs and a diverse selection of them will greatly improve somatic embryogenesis and increase transformation efficiency in a genotype-independent manner in crop species. This will drastically increase the applications and impacts of gene transformation and gene editing technologies for crop improvement through embryogenesis.

Over the past decade, much progress has been made in understanding the functions of TFs in plant embryogenesis and the molecular mechanisms related to these functions. TF regulatory networks involved in plant embryogenesis have gradually come to light, though much more research is needed to better understand the networks and their interactions during the process. The functional redundancy and divergence exhibited by TFs within the same family suggest a much more intricate regulatory process. How do the DNA-binding properties shared by TFs within the same family affect their target genes and the related regulatory networks is an intriguing question. The rapid technology development and applications such as ChIP-seq will provide powerful solutions to identify target genes of the TFs and the related TF regulatory networks during plant embryogenesis.

Author contributions

HY: Conceptualization, Funding acquisition, Writing – original draft. SK: Funding acquisition, Writing – review & editing. AF: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The work

was supported by the Sustainable Food System Initiative and the Sustainable Protein Production Program from the National Research Council of Canada.

Acknowledgments

We gratefully acknowledge Debbie Maizels at Zoobotanica, UK (http://www.scientific-art.com/) for the preparation of Figure 1. We thank members from the Aquatic and Crop Resource Development Research Centre, National Research Council of Canada at Saskatoon for the help related to the published article. We apologize to those whose work we were unable to cite owing to space constraints.

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

Abraham, Z., Iglesias-Fernández, R., Martínez, M., Rubio-Somoza, I., Díaz, I., Carbonero, P., et al. (2016). A developmental switch of gene expression in the barley seed mediated by hvVP1 (Viviparous-1) and hvGAMYB interactions. *Plant Physiol.* 170, 2146–2158. doi: 10.1104/PP.16.00092

Anderson, S. N., Johnson, C. S., Chesnut, J., Jones, D. S., Khanday, I., Woodhouse, M., et al. (2017). The zygotic transition is initiated in unicellular plant zygotes with asymmetric activation of parental genomes. *Dev. Cell* 43, 349–358.e4. doi: 10.1016/J.DEVCEL.2017.10.005

Aregawi, K., Shen, J., Pierroz, G., Bucheli, C., Sharma, M., Dahlberg, J., et al. (2020). Pathway to validate gene function in key bioenergy crop, sorghum bicolor. *bioRxiv*. doi: 10.1101/2020.12.08.416347

Aregawi, K., Shen, J., Pierroz, G., Sharma, M. K., Dahlberg, J., Owiti, J., et al. (2022). Morphogene-assisted transformation of Sorghum bicolor allows more efficient genome editing. *Plant Biotechnol. J.* 20, 748–760. doi: 10.1111/PBI.13754

Armenta-Medina, A., Gillmor, C. S., Gao, P., Mora-Macias, J., Kochian, L. V., Xiang, D., et al. (2021). Developmental and genomic architecture of plant embryogenesis: from model plant to crops. *Plant Commun.* 2, 100136. doi: 10.1016/J.XPLC.2020.100136

Arroyo-Herrera, A., Ku Gonzalez, A., Canche Moo, R., Quiroz-Figueroa, F. R., Loyola-Vargas, V. M., Rodriguez-Zapata, L. C., et al. (2008). Expression of WUSCHEL in Coffea canephora causes ectopic morphogenesis and increases somatic embryogenesis. *Plant Cell. Tissue Organ Cult.* 94, 171–180. doi: 10.1007/S11240-008-9401-1

Awada, R., Lepelley, M., Breton, D., Charpagne, A., Campa, C., Berry, V., et al. (2023). Global transcriptome profiling reveals differential regulatory, metabolic and hormonal networks during somatic embryogenesis in Coffea arabica. *BMC Genomics* 24, 1–19. doi: 10.1186/S12864-022-09098-Z/TABLES/1

Baile, F., Gómez-Zambrano, Á., and Calonje, M. (2022). Roles of Polycomb complexes in regulating gene expression and chromatin structure in plants. *Plant Commun.* 3, 100267. doi: 10.1016/J.XPLC.2021.100267

Barreto, H. G., Ságio, S. A., Chalfun-Júnior, A., Fevereiro, P., and Benedito, V. A. (2019). Transcriptional profiling of the AFL subfamily of B3-type transcription factors during the in *vitro* induction of somatic embryogenesis in the model legume Medicago truncatula. *Plant Cell. Tissue Organ Cult.* 139, 327–337. doi: 10.1007/S11240-019-01687-3/FIGURES/6

Beyene, G., Chauhan, R. D., Villmer, J., Husic, N., Wang, N., Gebre, E., et al. (2022). CRISPR/cas9-mediated tetra-allelic mutation of the 'Green revolution' SEMIDWARF-1 (SD-1) gene confers lodging resistance in tef (Eragrostis tef). *Plant Biotechnol. J.* 20, 1716–1729. doi: 10.1111/PBI.13842

Bilichak, A., Luu, J., Jiang, F., and Eudes, F. (2018). Identification of BABY BOOM homolog in bread wheat. *Agri Gene* 7, 43–51. doi: 10.1016/J.AGGENE.2017.11.002

Birchler, J. A., and Veitia, R. A. (2007). The gene balance hypothesis: from classical genetics to modern genomics. *Plant Cell* 19, 395. doi: 10.1105/TPC.106.049338

Bouchabké-Coussa, O., Obellianne, M., Linderme, D., Montes, E., Maia-Grondard, A., Vilaine, F., et al. (2013). Wuschel overexpression promotes somatic embryogenesis and induces organogenesis in cotton (Gossypium hirsutum L.) tissues cultured in *vitro*. *Plant Cell Rep.* 32, 675–686. doi: 10.1007/S00299-013-1402-9

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., et al. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14, 1737–1749. doi: 10.1105/TPC.001941

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., et al. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PloS Genet.* 7, e1002014. doi: 10.1371/JOURNAL.PGEN.1002014

Brand, A., Quimbaya, M., Tohme, J., and Chavarriaga-Aguirre, P. (2019). Arabidopsis LEC1 and LEC2 orthologous genes are key regulators of somatic embryogenesis in cassava. *Front. Plant Sci.* 10. doi: 10.3389/FPLS.2019.00673/BIBTEX

Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., and Laux, T. (2008). Differential expression of WOX genes mediates apical-basal axis formation in the arabidopsis embryo. *Dev. Cell* 14, 867–876. doi: 10.1016/J.DEVCEL.2008.03.008

Capron, A., Chatfield, S., Provart, N., and Berleth, T. (2009). Embryogenesis: pattern formation from a single cell. *Arabidopsis Book* 7, e0126. doi: 10.1199/TAB.0126

Carbonero, P., Iglesias-Fernández, R., and Vicente-Carbajosa, J. (2017). The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. *J. Exp. Bot.* 68, 871–880. doi: 10.1093/jxb/erw458

Chahal, L. S., Conner, J. A., and Ozias-Akins, P. (2022). Phylogenetically distant BABY BOOM genes from setaria italica induce parthenogenesis in rice. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.863908

Chakraborty, S., Valdes-Lopez, O., Stonoha-Arther, C., and Ane, J. M. (2022). Transcription factors controlling the rhizobium–legume symbiosis: integrating infection, organogenesis and the abiotic environment. *Plant Cell Physiol.* 63, 1326–1343. doi: 10.1093/PCP/PCAC063

Chen, Z., Debernardi, J. M., Dubcovsky, J., and Gallavotti, A. (2022b). The combination of morphogenic regulators BABY BOOM and GRF-GIF improves maize transformation efficiency. *bioRxiv*. doi: 10.1101/2022.09.02.506370

Chen, M., Lin, J.-Y., Wu, X., Apuya, N. R., Henry, K. F., Le, B. H., et al. (2021). Comparative analysis of embryo proper and suspensor transcriptomes in plant embryos with different morphologies. *Proc. Natl. Acad. Sci.* 118, e2024704118. doi: 10.1073/pnas.2024704118

Chen, B., Maas, L., Figueiredo, D., Zhong, Y., Reis, R., Li, M., et al. (2022a). BABY BOOM regulates early embryo and endosperm development. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2201761119. doi: 10.1073/PNAS.2201761119/-/DCSUPPLEMENTAL

Chen, D., Molitor, A., Liu, C., and Shen, W. H. (2010). The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. *Cell Res.* 20, 1332–1344. doi: 10.1038/cr.2010.151

Chen, J., Strieder, N., Krohn, N. G., Cyprys, P., Sprunck, S., Engelmann, J. C., et al. (2017). Zygotic genome activation occurs shortly after fertilization in maize. *Plant Cell* 29, 2106–2125. doi: 10.1105/TPC.17.00099

Chen, N., Veerappan, V., Abdelmageed, H., Kang, M., and Allen, R. D. (2018). HSI2/VAL1 silences AGL15 to regulate the developmental transition from seed maturation to vegetative growth in arabidopsis. *Plant Cell* 30, 600–619. doi: 10.1105/TPC.17.00655

Chopy, M., Cavallini-Speisser, Q., Chambrier, P., Morel, P., Just, J., Hugouvieux, V., et al. (2023). Cell layer-specific expression of the homeotic MADS-box transcription factor PhDEF contributes to modular petal morphogenesis in petunia. *Plant Cell* koad258. doi: 10.1093/PLCELL/KOAD258

Conner, J. A., Mookkan, M., Huo, H., Chae, K., and Ozias-Akins, P. (2015). A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11205–11210. doi: 10.1073/PNAS.1505856112

Conner, J. A., Podio, M., and Ozias-Akins, P. (2017). Haploid embryo production in rice and maize induced by PsASGR-BBML transgenes. *Plant Reprod.* 30, 41–52. doi: 10.1007/S00497-017-0298-X

Curaba, J., Moritz, T., Blervaque, R., Parcy, F., Raz, V., Herzog, M., et al. (2004). AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated

during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in arabidopsis. *Plant Physiol.* 136, 3660–3669. doi: 10.1104/PP.104.047266

Debernardi, J. M., Tricoli, D. M., Ercoli, M. F., Hayta, S., Ronald, P., Palatnik, J. F., et al. (2020). A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat. Biotechnol.* 38, 1274–1279. doi: 10.1038/s41587-020-0703-0

Deng, W., Luo, K., Li, Z., and Yang, Y. (2009). A novel method for induction of plant regeneration *via* somatic embryogenesis. *Plant Sci.* 177, 43–48. doi: 10.1016/J.PLANTSCI.2009.03.009

Devic, M., and Roscoe, T. (2016). Seed maturation: Simplification of control networks in plants. *Plant Sci.* 252, 335–346. doi: 10.1016/J.PLANTSCI.2016.08.012

de Vries, S. C., and Weijers, D. (2017). Plant embryogenesis. *Curr. Biol.* 27, R870–R873. doi: 10.1016/j.cub.2017.05.026

Duarte-Aké, F., Nic-Can, G., and De-la-Peña, C. (2019). "Somatic embryogenesis: Polycomb complexes control cell-to-embryo transition," R. Alvarez-Venegas, et al. (eds) in *Epigenetics in plants of agronomic importance: fundamentals and applications: transcriptional regulation and chromatin remodelling in plants*, *2nd ed.* (Springer Nature Switzerland AG), 339–354. doi: 10.1007/978-3-030-14760-0_13/COVER

Evans, C. E. B., Arunkumar, R., and Borrill, P. (2022). Transcription factor retention through multiple polyploidization steps in wheat. *G3 Genes|Genomes|Genetics* 12, jkac147. doi: 10.1093/G3JOURNAL/JKAC147

Florez, S. L., Erwin, R. L., Maximova, S. N., Guiltinan, M. J., and Curtis, W. R. (2015). Enhanced somatic embryogenesis in Theobroma cacao using the homologous BABY BOOM transcription factor. *BMC Plant Biol.* 15, 1–13. doi: 10.1186/S12870-015-0479-4

Fornari, M., Calvenzani, V., Masiero, S., Tonelli, C., and Petroni, K. (2013). The arabidopsis NF-YA3 and NF-YA8 genes are functionally redundant and are required in early embryogenesis. *PloS One* 8, e82043. doi: 10.1371/journal.pone.0082043

Freitas, N. C., Barreto, H. G., Torres, L. F., Freire, L. L., Rodrigues, L. A. Z., Diniz, L. E. C., et al. (2019). In silico and in *vivo* analysis of ABI3 and VAL2 genes during somatic embryogenesis of Coffea arabica: competence acquisition and developmental marker genes. *Plant Cell. Tissue Organ Cult.* 137, 599–611. doi: 10.1007/S11240-019-01594-7/FIGURES/6

Gazzarrini, S., Tsuchiya, Y., Lumba, S., Okamoto, M., and McCourt, P. (2004). The transcription factor FUSCA3 controls developmental timing in arabidopsis through the hormones gibberellin and abscisic acid. *Dev. Cell* 7, 373–385. doi: 10.1016/J.DEVCEL.2004.06.017

Gnesutta, N., Saad, D., Chaves-Sanjuan, A., Mantovani, R., and Nardini, M. (2017). Crystal structure of the arabidopsis thaliana L1L/NF-YC3 histone-fold dimer reveals specificities of the LEC1 family of NF-Y subunits in plants. *Mol. Plant* 10, 645–648. doi: 10.1016/j.molp.2016.11.006

Gramzow, L., and Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biol.* 11, 1–11. doi: 10.1186/GB-2010-11-6-214/FIGURES/5

Grimault, A., Gendrot, G., Chaignon, S., Gilard, F., Tcherkez, G., Thévenin, J., et al. (2015). Role of B3 domain transcription factors of the AFL family in maize kernel filling. *Plant Sci.* 236, 116–125. doi: 10.1016/J.PLANTSCI.2015.03.021

Gu, C., Guo, Z. H., Hao, P. P., Wang, G. M., Jin, Z. M., and Zhang, S. L. (2017). Multiple regulatory roles of AP2/ERF transcription factor in angiosperm. *Bot. Stud.* 58, 1–8. doi: 10.1186/S40529-016-0159-1/FIGURES/1

Gulzar, B., Mujib, A., Malik, M. Q., Sayeed, R., Mamgain, J., and Ejaz, B. (2020). Genes, proteins and other networks regulating somatic embryogenesis in plants. *J. Genet. Eng. Biotechnol.* 18, 31. doi: 10.1186/s43141-020-00047-5

Gundu, S., Tabassum, N., and Blilou, I. (2020). Moving with purpose and direction: transcription factor movement and cell fate determination revisited. *Curr. Opin. Plant Biol.* 57, 124–132. doi: 10.1016/J.PBI.2020.08.003

Guo, F., Liu, C., Xia, H., Bi, Y., Zhao, C., Zhao, S., et al. (2013). Induced expression of atLEC1 and atLEC2 differentially promotes somatic embryogenesis in transgenic tobacco plants. *PloS One* 8, e71714. doi: 10.1371/JOURNAL.PONE.0071714

Han, X., Yu, H., Yuan, R., Yang, Y., An, F., and Qin, G. (2019). Arabidopsis transcription factor TCP5 controls plant thermomorphogenesis by positively regulating PIF4 activity. *iScience* 15, 611–622. doi: 10.1016/J.ISCI.2019.04.005

Heidmann, I., de Lange, B., Lambalk, J., Angenent, G. C., and Boutilier, K. (2011). Efficient sweet pepper transformation mediated by the BABY BOOM transcription factor. *Plant Cell Rep.* 30, 1107–1115. doi: 10.1007/S00299-011-1018-X

Hoang, X. L. T., Nhi, D. N. H., Thu, N. B. A., Thao, N. P., and Tran, L.-S. P. (2017). Transcription factors and their roles in signal transduction in plants under abiotic stresses. *Curr. Genomics* 18, 483. doi: 10.2174/1389202918666170227150057

Hoerster, G., Wang, N., Ryan, L., Wu, E., Anand, A., McBride, K., et al. (2020). Use of non-integrating Zm-Wus2 vectors to enhance maize transformation: Non-integrating WUS2 enhances transformation. *Vitr. Cell. Dev. Biol. - Plant* 56, 265–279. doi: 10.1007/S11627-019-10042-2/FIGURES/9

Hoffmann, T., Shi, X., Hsu, C. Y., Brown, A., Knight, Q., Courtney, L. S., et al. (2022). The identification of type I MADS box genes as the upstream activators of an endosperm-specific invertase inhibitor in Arabidopsis. *BMC Plant Biol.* 22, 1–13. doi: 10.1186/S12870-021-03399-3/FIGURES/5

Hong, J. C. (2016). ""General aspects of plant transcription factor families,"," in *Plant transcription factors: evolutionary, structural and functional aspects*. Ed. D. Gonzelez (Academic Press), 35–56. doi: 10.1016/B978-0-12-800854-6.00003-8

Horstman, A., Bemer, M., and Boutilier, K. (2017a). A transcriptional view on somatic embryogenesis. *Regeneration* 4, 201–216. doi: 10.1002/reg2.91

Horstman, A., Li, M., Heidmann, I., Weemen, M., Chen, B., Muino, J. M., et al. (2017b). The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis. *Plant Physiol.* 175, 848–857. doi: 10.1104/PP.17.00232

Ikeda, M., Umehara, M., and Kamada, H. (2006). Embryogenesis-related genes; Its expression and roles during somatic and zygotic embryogenesis in carrot and Arabidopsis. *Plant Biotechnol.* 23, 153–161. doi: 10.5511/PLANTBIOTECHNOLOGY.23.153

Ikeuchi, M., Iwase, A., Rymen, B., Harashima, H., Shibata, M., Ohnuma, M., et al. (2015). PRC2 represses dedifferentiation of mature somatic cells in Arabidopsis. *Nat. Plants* 17 (1), 1–7. doi: 10.1038/nplants.2015.89

Itoh, J. I., Sato, Y., Sato, Y., Hibara, K. I., Shimizu-Sato, S., Kobayashi, H., et al. (2016). Genome-wide analysis of spatiotemporal gene expression patterns during early embryogenesis in rice. *Dev.* 143, 1217–1227. doi: 10.1242/DEV.123661/256993/AM/GENOME-WIDE-ANALYSIS-OF-SPATIO-TEMPORAL-GENE

Jha, P., Ochatt, S., and Kumar, V. (2020). WUSCHEL: a master regulator in plant growth signaling. *Plant Cell Rep.* 39, 431–444. doi: 10.1007/s00299-020-02511-5

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

Johnson, K., Cao Chu, U., Anthony, G., Wu, E., Che, P., and Jones, T. J. (2023). Rapid and highly efficient morphogenic gene-mediated hexaploid wheat transformation. *Front. Plant Sci.* 14. doi: 10.3389/FPLS.2023.1151762/BIBTEX

Joshi, S., Awan, H., Paul, P., Tian, R., and Perry, S. (2022a). Revisiting AGAMOUS-LIKE15, a key somatic embryogenesis regulator, using next generation sequencing analysis in arabidopsis. *Int. J. Mol. Sci.* 23, 15082. doi: 10.3390/ijms232315082

Joshi, S., Paul, P., Hartman, J. M., and Perry, S. E. (2022b). AGL15 promotion of somatic embryogenesis: role and molecular mechanism. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.861556

Junker, A., Mönke, G., Rutten, T., Keilwagen, J., Seifert, M., Thi, T. M. N., et al. (2012). Elongation-related functions of LEAFY COTYLEDON1 during the development of Arabidopsis thaliana. *Plant J.* 71, 427–442. doi: 10.1111/J.1365-313X.2012.04999.X

Kadri, A., De March, G. G., Guerineau, F., Cosson, V., and Ratet, P. (2021). WUSCHEL overexpression promotes callogenesis and somatic embryogenesis in medicago truncatula gaertn. *Plants* 10, 715. doi: 10.3390/PLANTS10040715

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., and Hattori, T. (2005). LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46, 399–406. doi: 10.1093/PCP/PCI048

Kao, P., Schon, M. A., Mosiolek, M., Enugutti, B., and Nodine, M. D. (2021). Gene expression variation in Arabidopsis embryos at single-nucleus resolution. *Dev.* 148, dev199589. doi: 10.1242/DEV.199589/268394/AM/GENE-EXPRESSION-VARIATION-IN-ARABIDOPSIS-EMBRYOS

Kerstens, M., Galinha, C., Hofhuis, H., Nodine, M., Scheres, B., and Willemsen, V. (2022). Redundant PLETHORA activity promotes development of early embryonic cell lineages in Arabidopsis. *bioRxiv* 2022, 3.02.482431. doi: 10.1101/2022.03.02.482431

Khanday, I., Santos-Medell In, C., and Sundaresan, V. (2023). Somatic embryo initiation by rice BABY BOOM1 involves activation of zygote-expressed auxin biosynthesis genes. *New Phytol.* 238, 673–687. doi: 10.1111/NPH.18774

Khanday, I., Skinner, D., Yang, B., Mercier, R., and Sundaresan, V. (2018). A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nat* 565, 91–95. doi: 10.1038/s41586-018-0785-8

Kong, J., Martin-Ortigosa, S., Finer, J., Orchard, N., Gunadi, A., Batts, L. A., et al. (2020). Overexpression of the transcription factor GROWTH-REGULATING FACTOR5 improves transformation of dicot and monocot species. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.572319/BIBTEX

Köszegi, D., Johnston, A. J., Rutten, T., Czihal, A., Altschmied, L., Kumlehn, J., et al. (2011). Members of the RKD transcription factor family induce an egg cell-like gene expression program. *Plant J.* 67, 280–291. doi: 10.1111/j.1365-313X.2011.04592.x

Kruglova, N. N., Titova, G. E., and Zinatullina, A. E. (2022). Critical stages of cereal embryogenesis: theoretical and practical significance. *Russ. J. Dev. Biol.* 53, 437–453. doi: 10.1134/S1062360422060042

Kumar, V., Jha, P., and Van Staden, J. (2020). LEAFY COTYLEDONs (LECs): master regulators in plant embryo development. *Plant Cell. Tissue Organ Cult.* 140, 475–487. doi: 10.1007/S11240-019-01752-X/FIGURES/2

Kurdyukov, S., Song, Y., Sheahan, M. B., and Rose, R. J. (2014). Transcriptional regulation of early embryo development in the model legume Medicago truncatula. *Plant Cell Rep.* 33, 349–362. doi: 10.1007/s00299-013-1535-x

Kwong, R. W., Bui, A. Q., Lee, H., Kwong, L. W., Fischer, R. L., Goldberg, R. B., et al. (2003). LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. *Plant Cell* 15, 5–18. doi: 10.1105/TPC.006973

Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., et al. (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *PNAS* 107, 8063–8070. doi: 10.1073/pnas.1003530107

Lee, D. S., Chen, L. J., Li, C. Y., Liu, Y., Tan, X. L., Lu, B. R., et al. (2013). The bsister MADS gene FST determines ovule patterning and development of the zygotic embryo and endosperm. *PloS One* 8, e58748. doi: 10.1371/JOURNAL.PONE.0058748

Lee, S. J., Lee, B. H., Jung, J. H., Park, S. K., Song, J. T., and Kim, J. H. (2018). GROWTH-REGULATING FACTOR and GRF-INTERACTING FACTOR specify meristematic cells of gynoecia and anthers. *Plant Physiol.* 176, 717. doi: 10.1104/PP.17.00960

Li, Q., Deng, C., Xia, Y., Kong, L., Zhang, H., Zhang, S., et al. (2017). Identification of novel miRNAs and miRNA expression profiling in embryogenic tissues of Picea balfouriana treated by 6-benzylaminopurine. *PloS One* 12, e0176112. doi: 10.1371/JOURNAL.PONE.0176112

Li, K., Wang, J., Liu, C., Li, C., Qiu, J., Zhao, C., et al. (2019). Expression of AtLEC2 and AtIPTs promotes embryogenic callus formation and shoot regeneration in tobacco. *BMC Plant Biol.* 19, 1–15. doi: 10.1186/S12870-019-1907-7/FIGURES/5

Li, M., Wrobel-Marek, J., Heidmann, I., Horstman, A., Chen, B., Reis, R., et al. (2022). Auxin biosynthesis maintains embryo identity and growth during BABY BOOM-induced somatic embryogenesis. *Plant Physiol.* 188, 195–1110. doi: 10.1093/plphys/kiab558

Liu, Z., Ge, X. X., Qiu, W. M., Long, J. M., Jia, H. H., Yang, W., et al. (2018). Overexpression of the CsFUS3 gene encoding a B3 transcription factor promotes somatic embryogenesis in Citrus. *Plant Sci.* 277, 121–131. doi: 10.1016/J.PLANTSCI.2018.10.015

Liu, W., Zhang, Y., Fang, X., Tran, S., Zhai, N., Yang, Z., et al. (2022). Transcriptional landscapes of *de novo* root regeneration from detached Arabidopsis leaves revealed by time-lapse and single-cell RNA sequencing analyses. *Plant Commun.* 3, 100306. doi: 10.1016/j.xplc.2022.100306

Lotan, T., Ohto, MA., and Matsudaira Yee, K. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cells* 93, 1195–1205. doi: 10.1016/S0092-8674(00)81463-4

Lowe, K., La Rota, M., Hoerster, G., Hastings, C., Wang, N., Chamberlin, M., et al. (2018). Rapid genotype "independent" Zea mays L. (maize) transformation via direct somatic embryogenesis. *Vitr. Cell. Dev. Biol. - Plant* 54, 240–252. doi: 10.1007/S11627-018-9905-2

Lutova, L. A., Dodueva, I. E., Lebedeva, M. A., and Tvorogova, V. E. (2015). Transcription factors in developmental genetics and the evolution of higher plants. *Russ. J. Genet.* 51, 449–466. doi: 10.1134/S1022795415030084/METRICS

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., et al. (2006). Different polycomb group complexes regulate common target genes in Arabidopsis. *EMBO Rep.* 7, 947–952. doi: 10.1038/sj.embor.7400760

Mayer, K. F. X., Schoof, H., Haecker, A., Lenhard, M., Jürgens, G., and Laux, T. (1998). Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. *Cell* 95, 805–815. doi: 10.1016/S0092-8674(00)81703-1

Mcfarland, F. L., Collier, R., Walter, N., Martinell, B., Kaeppler, S. M., and Kaeppler, H. F. (2023). A key to totipotency: Wuschel-like homeobox 2a unlocks embryogenic culture response in maize (Zea mays L.). *Plant Biotechnol. J.* 21, 1860–1872. doi: 10.1111/pbi.14098

Méndez-Hernández, H. A., Ledezma-Rodríguez, M., Avilez-Montalvo, R. N., Juárez-Gómez, Y. L., Skeete, A., Avilez-Montalvo, J., et al. (2019). Signaling overview of plant somatic embryogenesis. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00077

Mimura, M., Kudo, T., Wu, S., McCarty, D. R., and Suzuki, M. (2018). Autonomous and non-autonomous functions of the maize Shohai1 gene, encoding a RWP-RK putative transcription factor, in regulation of embryo and endosperm development. *Plant J.* 95, 892–908. doi: 10.1111/TPJ.13996

Mitsuda, N., and Ohme-Takagi, M. (2009). Functional analysis of transcription factors in arabidopsis. *Plant Cell Physiol.* 50, 1232–1248. doi: 10.1093/pcp/pcp075

Mookkan, M., Nelson-Vasilchik, K., Hague, J., Zhang, Z. J., and Kausch, A. P. (2017). Selectable marker independent transformation of recalcitrant maize inbred B73 and sorghum P898012 mediated by morphogenic regulators BABY BOOM and WUSCHEL2. *Plant Cell Rep.* 36, 1477–1491. doi: 10.1007/S00299-017-2169-1

Mozgova, I., Köhler, C., and Hennig, L. (2015). Keeping the gate closed: Functions of the polycomb repressive complex PRC2 in development. *Plant J.* 83, 121–132. doi: 10.1111/tpj.12828

Mu, J., Tan, H., Hong, S., Liang, Y., and Zuo, J. (2013). Arabidopsis transcription factor genes NF-YA1, 5, 6, and 9 play redundant roles in male gametogenesis, embryogenesis, and seed development. *Mol. Plant* 6, 188–201. doi: 10.1093/MP/SSS061

Mu, J., Tan, H., Zheng, Q., Fu, Y., Liang, Y., Zhang, J., et al. (2008). LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in arabidopsis. *Plant Physiol.* 148, 1042–1054. doi: 10.1104/PP.108.126342

Mukherjee, K., Brocchieri, L., and Bürglin, T. R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. *Mol. Biol. Evol.* 26, 2775. doi: 10.1093/MOLBEV/MSP201

Nardmann, J., Zimmermann, R., Durantini, D., Kranz, E., and Werr, W. (2007). WOX gene phylogeny in poaceae: A comparative approach addressing leaf and embryo development. *Mol. Biol. Evol.* 24, 2474–2484. doi: 10.1093/MOLBEV/MSM182

Nelson-Vasilchik, K., Hague, J. P., Tilelli, M., and Kausch, A. P. (2022). Rapid transformation and plant regeneration of sorghum (Sorghum bicolor L.) mediated by altruistic Baby boom and Wuschel2. *Vitr. Cell. Dev. Biol. - Plant* 58, 331–342. doi: 10.1007/S11627-021-10243-8

Niu, B., Zhang, Z., Zhang, J., Zhou, Y., and Chen, C. (2021). The rice LEC1-like transcription factor OsNF-YB9 interacts with SPK, an endosperm-specific sucrose synthase protein kinase, and functions in seed development. *Plant J.* 106, 1233–1246. doi: 10.1111/TPJ.15230

Nowak, K., Morończyk, J., Grzyb, M., Szczygieł-Sommer, A., and Gaj, M. D. (2022). miR172 Regulates WUS during Somatic Embryogenesis in Arabidopsis *via* AP2. *Cells* 11, 718. doi: 10.3390/CELLS11040718/S1

Nurbaiti, S., Indrianto, A., Purwestri, Y. A., Daryono, B. S., and Semiarti, E. (2021). Characterization of putative RKD homologous gene with microspore embryogenesis in the local pigmented rice cultivar "SEGRENG". *SABRAO J. Breed. Genet.* 53, 201–212.

Omidbakhshfard, M. A., Proost, S., Fujikura, U., and Mueller-Roeber, B. (2015). Growth-regulating factors (GRFs): A small transcription factor family with important functions in plant biology. *Mol. Plant* 8, 998–1010. doi: 10.1016/j.molp.2015.01.013

Orłowska, A., Igielski, R., Łagowska, K., and Kępczyńska, E. (2017). Identification of LEC1, L1L and Polycomb Repressive Complex 2 genes and their expression during the induction phase of Medicago truncatula Gaertn. somatic embryogenesis. *Plant Cell Tissue Organ Cult.* 129, 119–132. doi: 10.1007/s11240-016-1161-8

Paul, P., Joshi, S., Tian, R., Diogo Junior, R., Chakrabarti, M., and Perry, S. E. (2022). The MADS-domain factor AGAMOUS-Like18 promotes somatic embryogenesis. *Plant Physiol.* 188, 1617–1631. doi: 10.1093/plphys/kiab553

Pelletier, J. M., Kwong, R. W., Park, S., Le, B. H., Baden, R., Cagliari, A., et al. (2017). LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc. Natl. Acad. Sci. U. S. A.* 114, E6710–E6719. doi: 10.1073/PNAS.1707957114/SUPPL_FILE/PNAS.1707957114.SD06.XLSX

Perry, S. E., Lehti, M. D., and Fernandez, D. E. (1999). The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. *Plant Physiol.* 120, 121. doi: 10.1104/PP.120.1.121

Peterson, D., Barone, P., Lenderts, B., Schwartz, C., Feigenbutz, L., St. Clair, G., et al. (2021). Advances in Agrobacterium transformation and vector design result in high-frequency targeted gene insertion in maize. *Plant Biotechnol. J.* 19, 2000–2010. doi: 10.1111/PBI.13613

Petroni, K., Kumimoto, R. W., Gnesutta, N., Calvenzani, V., Fornari, M., Tonelli, C., et al. (2012). The promiscuous life of plant NUCLEAR FACTOR Y transcription factors. *Plant Cell* 24, 4777. doi: 10.1105/TPC.112.105734

Pfreundt, U., James, D. P., Tweedie, S., Wilson, D., Teichmann, S. A., and Adryan, B. (2010). FlyTF: improved annotation and enhanced functionality of the Drosophila transcription factor database. *Nucleic Acids Res.* 38, D443. doi: 10.1093/NAR/GKP910

Priya, P., and Jain, M. (2013). RiceSRTFDB: A database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. *Database J. Biol. Database Curation* 2013, bat027. doi: 10.1093/DATABASE/BAT027

Purwestri, Y. A., Lee, Y. S., Meehan, C., Mose, W., Susanto, F. A., Wijayanti, P., et al. (2023). RWP-RK Domain 3 (OsRKD3) induces somatic embryogenesis in black rice. *BMC Plant Biol.* 23, 1–15. doi: 10.1186/S12870-023-04220-Z/FIGURES/6

Qi, X., Gao, H., Lv, R., Mao, W., Zhu, J., Liu, C., et al. (2022). CRISPR/dCas-mediated gene activation toolkit development and its application for parthenogenesis induction in maize. *Plant Commun.* 4, 100449. doi: 10.1016/j.xplc.2022.100449

Radoeva, T., Albrecht, C., Piepers, M., De Vries, S., and Weijers, D. (2020). Suspensor-derived somatic embryogenesis in Arabidopsis. *Development* 147, dev 188912. doi: 10.1242/dev.188912

Rupps, A., Raschke, J., Rümmler, M., Linke, B., and Zoglauer, K. (2016). Identification of putative homologs of Larix decidua to BABYBOOM (BBM), LEAFY COTYLEDON1 (LEC1), WUSCHEL-related HOMEOBOX2 (WOX2) and SOMATIC EMBRYOGENESIS RECEPTOR-like KINASE (SERK) during somatic embryogenesis. *Planta* 243, 473–488. doi: 10.1007/S00425-015-2409-Y

Sasaki, K. (2018). Utilization of transcription factors for controlling floral morphogenesis in horticultural plants. *Breed. Sci.* 68, 88. doi: 10.1270/JSBBS.17114

Seo, E., and Choi, D. (2015). Functional studies of transcription factors involved in plant defenses in the genomics era. *Brief. Funct. Genomics* 14, 260–267. doi: 10.1093/bfgp/elv011

Shahzad, R., Jamil, S., Ahmad, S., Nisar, A., Amina, Z., Saleem, S., et al. (2021). Harnessing the potential of plant transcription factors in developing climate resilient crops to improve global food security: Current and future perspectives. *Saudi J. Biol. Sci.* 28, 2323–2341. doi: 10.1016/J.SJBS.2021.01.028

Shi, H., Lyu, M., Luo, Y., Liu, S., Li, Y., He, H., et al. (2018). Genome-wide regulation of light-controlled seedling morphogenesis by three families of transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* 115, 6482–6487. doi: 10.1073/PNAS.1803861115/SUPPL_FILE/PNAS.1803861115.SD02.XLS

Shimada, T., Endo, T., Fujii, H., Nakano, M., Sugiyama, A., Daido, G., et al. (2018). MITE insertion-dependent expression of CitRKD1 with a RWP-RK domain regulates somatic embryogenesis in citrus nucellar tissues. *BMC Plant Biol.* 18, 1–19. doi: 10.1186/S12870-018-1369-3/FIGURES/10

Shires, M., Florez, S., Lai, T., and Curtis, W. (2017). Inducible somatic embryogenesis in Theobroma cacao achieved using the DEX-activatable transcription factor-glucocorticoid receptor fusion. *Biotechnol. Lett.* 39, 1747–1755. doi: 10.1007/s10529-017-2404-4

Shiu, S.-H., Shih, M.-C., and Li, W.-H. (2005). Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol.* 139, 18–26. doi: 10.1104/pp.105.065110

Shokri, L., Inukai, S., Hafner, A., Weinand, K., Hens, K., Vedenko, A., et al. (2019). A comprehensive drosophila melanogaster transcription factor interactome. *Cell Rep.* 27, 955. doi: 10.1016/J.CELREP.2019.03.071

Siriwardana, C. L., Kumimoto, R. W., Jones, D. S., and Holt, B. F. (2014). Gene family analysis of the arabidopsis NF-YA transcription factors reveals opposing abscisic acid responses during seed germination. *Plant Mol. Biol. Rep.* 32, 971. doi: 10.1007/S11105-014-0704-6

Sluis, A., and Hake, S. (2015). Organogenesis in plants: Initiation and elaboration of leaves. *Trends Genet.* 31, 300–306. doi: 10.1016/j.tig.2015.04.004

Smertenko, A., and Bozhkov, P. V. (2014). Somatic embryogenesis: life and death processes during apical–basal patterning. *J. Exp. Bot.* 65, 1343–1360. doi: 10.1093/JXB/ERU005

Srinivasan, C., Liu, Z., Heidmann, I., Supena, E. D. J., Fukuoka, H., Joosen, R., et al. (2007). Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (Nicotiana tabacum L.). *Planta* 225, 341–351. doi: 10.1007/S00425-006-0358-1

Stone, S. L., Braybrook, S. A., Paula, S. L., Kwong, L. W., Meuser, J., Pelletier, J., et al. (2008). Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3151–3156. doi: 10.1073/PNAS.0712364105

Strader, L., Weijers, D., and Wagner, D. (2022). Plant transcription factors — being in the right place with the right company. *Curr. Opin. Plant Biol.* 65, 102136. doi: 10.1016/J.PBI.2021.102136

Su, Y. H., Zhao, X. Y., Liu, Y. B., Zhang, C. L., O'Neill, S. D., and Zhang, X. S. (2009). Auxin-induced WUS expression is essential for embryonic stem cell renewal during somatic embryogenesis in Arabidopsis. *Plant J.* 59, 448. doi: 10.1111/J.1365-313X.2009.03880.X

Suo, J., Zhou, C., Zeng, Z., Li, X., Bian, H., Wang, J., et al. (2021). Identification of regulatory factors promoting embryogenic callus formation in barley through transcriptome analysis. *BMC Plant Biol.* 21, 1–19. doi: 10.1186/S12870-021-02922-W/FIGURES/9

Swaminathan, K., Peterson, K., and Jack, T. (2008). The plant B3 superfamily. *Trends Plant Sci.* 13, 647–655. doi: 10.1016/J.TPLANTS.2008.09.006

Ten Hove, C. A., Lu, K. J., and Weijers, D. (2015). Building a plant: cell fate specification in the early Arabidopsis embryo. *Development* 142, 420–430. doi: 10.1242/DEV.111500

Thakare, D., Tang, W., Hill, K., and Perry, S. E. (2008). The MADS-domain transcriptional regulator AGAMOUS-LIKE15 promotes somatic embryo development in arabidopsis and soybean. *Plant Physiol.* 146, 1663–1672. doi: 10.1104/PP.108.115832

Theißen, G., and Gramzow, L. (2016). Structure and Evolution of Plant MADS Domain Transcription Factors. In: D. Gonzalez (ed) Plant Transcription Factors. *Academic Press*, pp 127–138. doi: 10.1016/B978-0-12-800854-6.00008-7

Tian, R., Paul, P., Joshi, S., and Perry, S. E. (2020). Genetic activity during early plant embryogenesis. *Biochem. J.* 477, 3743–3767. doi: 10.1042/BCJ20190161

To, A., Valon, C., Savino, G., Guilleminot, J., Devic, M., Giraudat, J., et al. (2006). A network of local and redundant gene regulation governs arabidopsis seed maturation. *Plant Cell* 18, 1642. doi: 10.1105/TPC.105.039925

Tsukagoshi, H., Morikami, A., and Nakamura, K. (2007). Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2543. doi: 10.1073/PNAS.0607940104

Tu, M., Wang, W., Yao, N., Cai, C., Liu, Y., Lin, C., et al. (2021). The transcriptional dynamics during *de novo* shoot organogenesis of Ma bamboo (Dendrocalamus latiflorus Munro): implication of the contributions of the abiotic stress response in this process. *Plant J.* 107, 1513–1532. doi: 10.1111/TPJ.15398

Tvorogova, V. E., Fedorova, Y. A., Potsenkovskaya, E. A., Kudriashov, A. A., Efremova, E. P., Kvitkovskaya, V. A., et al. (2019). The WUSCHEL-related homeobox transcription factor MtWOX9-1 stimulates somatic embryogenesis in Medicago truncatula. *Plant Cell. Tissue Organ Cult.* 138, 517–527. doi: 10.1007/S11240-019-01648-W

Ueda, M., Aichinger, E., Gong, W., Groot, E., Verstraeten, I., Vu, L. D., et al. (2017). Transcriptional integration of paternal and maternal factors in the Arabidopsis zygote. *Genes Dev.* 31, 617–627. doi: 10.1101/GAD.292409.116

Van Daele, I., Gonzalez, N., Vercauteren, I., de Smet, L., Inzé, D., Roldán-Ruiz, I., et al. (2012). A comparative study of seed yield parameters in Arabidopsis thaliana mutants and transgenics. *Plant Biotechnol. J.* 10, 488–500. doi: 10.1111/j.1467-7652.2012.00687.x

Verma, S., Pardha, V., Attuluri, S., and Robert, H. S. (2022). Transcriptional control of Arabidopsis seed development. *Planta* 255, 90. doi: 10.1007/s00425-022-03870-x

Vernoud, V., Hajduch, M., Khaled, A. S., Depège, N., and Rogowsky, P. M. (2005). Maize embryogenesis. *Maydica* 50, 469–483.

Waki, T., Hiki, T., Watanabe, R., Hashimoto, T., and Nakajima, K. (2011). The arabidopsis RWP-RK protein RKD4 triggers gene expression and pattern formation in early embryogenesis. *Curr. Biol.* 21, 1277–1281. doi: 10.1016/J.CUB.2011.07.001

Wang, H., Caruso, L. V., Downie, A. B., and Perry, S. E. (2004). The embryo MADS domain protein AGAMOUS-like 15 directly regulates expression of a gene encoding an enzyme involved in gibberellin metabolism. *Plant Cell* 16, 1206. doi: 10.1105/TPC.021261

Wang, N., Ryan, L., Sardesai, N., Wu, E., Lenderts, B., Lowe, K., et al. (2023). Leaf transformation for efficient random integration and targeted genome modification in maize and sorghum. *Nat. Plants* 9, 255–270. doi: 10.1038/s41477-022-01338-0

Wang, F. X., Shang, G. D., Wu, L. Y., Xu, Z. G., Zhao, X. Y., and Wang, J. W. (2020). Chromatin accessibility dynamics and a hierarchical transcriptional regulatory network structure for plant somatic embryogenesis. *Dev. Cell* 54, 742–757.e8. doi: 10.1016/J.DEVCEL.2020.07.003

Wang, G., Wang, F., Huang, Q., Li, Y., Liu, Y., and Wang, Y. (2015). Understanding transcription factor regulation by integrating gene expression and DNase I hypersensitive sites. *BioMed. Res. Int.* 2015, 757530. doi: 10.1155/2015/757530

Wendrich, J. R., and Weijers, D. (2013). The Arabidopsis embryo as a miniature morphogenesis model. *New Phytol.* 199, 14–25. doi: 10.1111/nph.12267

Wickramasuriya, A. M., and Dunwell, J. M. (2015). Global scale transcriptome analysis of Arabidopsis embryogenesis in *vitro*. *BMC Genomics* 16, 1–23. doi: 10.1186/s12864-015-1504-6

Winkelmann, T. (2016). Somatic versus zygotic embryogenesis : learning from seeds. *In Vitro Embryogenesis Higher Plants Methods Mol. Biol.* 1359, 25–46. doi: 10.1007/978-1-4939-3061-6

Wójcikowska, B., Jaskóła, K., Gąsiorek, P., Meus, M., Nowak, K., and Gaj, M. D. (2013). LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, *via* YUCCA-mediated auxin biosynthesis. *Planta* 238, 425–440. doi: 10.1007/S00425-013-1892-2

Wu, X. M., Kou, S. J., Liu, Y. L., Fang, Y. N., Xu, Q., and Guo, W. W. (2015). Genomewide analysis of small RNAs in nonembryogenic and embryogenic tissues of citrus: microRNA-and siRNA-mediated transcript cleavage involved in somatic embryogenesis. *Plant Biotechnol. J.* 13, 383–394. doi: 10.1111/PBI.12317

Wu, J., Mohamed, D., Dowhanik, S., Petrella, R., Gregis, V., Li, J., et al. (2020). Spatiotemporal restriction of FUSCA3 expression by class I BPCs promotes ovule development and coordinates embryo and endosperm growth. *Plant Cell* 32, 1886–1904. doi: 10.1105/TPC.19.00764

Xiao, X., Zhang, C., Liu, Y., Wang, X., and You, C. (2023). Functional identification of apple Baby Boom in genetic transformation and somatic embryogenesis. *Vitr. Cell. Dev. Biol. - Plant* 59, 1–13. doi: 10.1007/S11627-022-10292-7/FIGURES/9

Xu, L., Cheng, J., and Jiang, H. (2022a). Mutation of histone H3 serine 28 to alanine influences H3K27me3-mediated gene silencing in Arabidopsis thaliana. *Plant Physiol.* 190, 2417–2429. doi: 10.1093/PLPHYS/KIAC409

Xu, N., Kang, M., Zobrist, J. D., Wang, K., and Fei, S. Z. (2022b). Genetic transformation of recalcitrant upland switchgrass using morphogenic genes. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.781565/BIBTEX

Yamamoto, A., Kagaya, Y., Usui, H., Hobo, T., Takeda, S., and Hattori, T. (2010). Diverse roles and mechanisms of gene regulation by the arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. *Plant Cell Physiol.* 51, 2031–2046. doi: 10.1093/PCP/PCQ162

Yang, Z., Li, C., Wang, Y., Zhang, C., Wu, Z., Zhang, X., et al. (2014). GhAGL15s, preferentially expressed during somatic embryogenesis, promote embryogenic callus formation in cotton (Gossypium hirsutum L.). *Mol. Genet. Genomics* 289, 873–883. doi: 10.1007/S00438-014-0856-Y/FIGURES/5

Yavuz, C., Tillaboeva, S., and Bakhsh, A. (2020). Apprehending the potential of BABY BOOM transcription factors to mitigate cotton regeneration and transformation. *J. Cott. Res.* 3, 1–14. doi: 10.1186/S42397-020-00071-3

Yuan, L., Song, X., Zhang, L., Yu, Y., Liang, Z., Lei, Y., et al. (2021). The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in Arabidopsis. *Nucleic Acids Res.* 49, 98–113. doi: 10.1093/nar/gkaa1129

Zhang, Z., Conner, J., Guo, Y., and Ozias-Akins, P. (2020b). Haploidy in tobacco induced by psASGR-BBML transgenes *via* parthenogenesis. *Genes (Basel).* 11, 1072. doi: 10.3390/GENES11091072

Zhang, Y., Jiao, Y., Jiao, H., Zhao, H., and Zhu, Y.-X. (2016). Two-step functional innovation of the stem-cell factors WUS/WOX5 during plant evolution. *Mol. Biol. Evol.* 34, 640–653. doi: 10.1093/molbev/msw263

Zhang, T., Li, C., Li, D., Liu, Y., and Yang, X. (2020a). Roles of YABBY transcription factors in the modulation of morphogenesis, development, and phytohormone and stress responses in plants. *J. Plant Res.* 133, 751–763. doi: 10.1007/S10265-020-01227-7/FIGURES/2

Zhao, P., Begcy, K., Dresselhaus, T., and Sun, M. X. (2017). Does early embryogenesis in eudicots and monocots involve the same mechanism and molecular players? *Plant Physiol.* 173, 130. doi: 10.1104/PP.16.01406

Zhao, L., Yang, Y., Chen, J., Lin, X., Zhang, H., Wang, H., et al. (2023). Dynamic chromatin regulatory programs during embryogenesis of hexaploid wheat. *Genome Biol.* 24, 1–29. doi: 10.1186/S13059-022-02844-2/FIGURES/7

Zhao, S., Zhang, B., Yang, M., Zhu, J., and Li, H. (2018). Systematic profiling of histone readers in arabidopsis thaliana. *Cell Rep.* 22, 1090–1102. doi: 10.1016/J.CELREP.2017.12.099

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/J.MOLP.2016.09.014

Zheng, Q., and Perry, S. E. (2014). Alterations in the transcriptome of soybean in response to enhanced somatic embryogenesis promoted by orthologs of AGAMOUS-like15 and AGAMOUS-like18. *Plant Physiol.* 164, 1365. doi: 10.1104/PP.113.234062

Zheng, Y., Ren, N., Wang, H., Stromberg, A. J., and Perry, S. E. (2009). Global identification of targets of the arabidopsis MADS domain protein AGAMOUS-like15. *Plant Cell* 21, 2563. doi: 10.1105/TPC.109.068890

Zhou, X., Guo, Y., Zhao, P., and Sun, M. X. (2018). Comparative analysis of WUSCHEL-related homeobox genes revealed their parent-of-origin and cell type-specific expression pattern during early embryogenesis in tobacco. *Front. Plant Sci.* 9. doi: 10.3389/FPLS.2018.00311

Zhou, Q., Liu, M., Xia, X., Gong, T., Feng, J., Liu, W., et al. (2017). A mouse tissue transcription factor atlas. *Nat. Commun. 2017* 81, 1–15. doi: 10.1038/ncomms15089

Zou, X., and Sun, H. (2023). DOF transcription factors: Specific regulators of plant biological processes. *Front. Plant Sci.* 14. doi: 10.3389/FPLS.2023.1044918/BIBTEX

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 23 January 2024
DOI 10.3389/fpls.2023.1305081

Check for updates

OPEN ACCESS

EDITED BY
Daoquan Xiang,
National Research Council Canada (NRC),
Canada

REVIEWED BY
Sheng Wang,
University of Saskatchewan, Canada
Rui Wen,
Agriculture and Agri-Food Canada (AAFC),
Canada

*CORRESPONDENCE
Jianhua Zhang
zhangjianhua@caas.cn
Guomin Zhou
zhouguomin@caas.cn

RECEIVED 30 September 2023
ACCEPTED 26 December 2023
PUBLISHED 23 January 2024

CITATION
Yao Q, Zheng X, Zhou G and Zhang J (2024) SGR-YOLO: a method for detecting seed germination rate in wild rice.
Front. Plant Sci. 14:1305081.
doi: 10.3389/fpls.2023.1305081

COPYRIGHT
© 2024 Yao, Zheng, Zhou 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.

SGR-YOLO: a method for detecting seed germination rate in wild rice

Qiong Yao[1,2], Xiaoming Zheng[2], Guomin Zhou[2,4]* and Jianhua Zhang[2,4]*

[1]College of Agriculture, Henan University, Zhengzhou, China, [2]National Academy of Southern Breeding, Chinese Academy of Agricultural Sciences, Sanya, China, [3]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [4]Agricultural Information Institute of Chinese Academy of Agricultural Sciences/National Agricultural Science Data Center, Beijing, China

Seed germination rate is one of the important indicators in measuring seed quality and seed germination ability, and it is also an important basis for evaluating the growth potential and planting effect of seeds. In order to detect seed germination rates more efficiently and achieve automated detection, this study focuses on wild rice as the research subject. A novel method for detecting wild rice germination rates is introduced, leveraging the SGR-YOLO model through deep learning techniques. The SGR-YOLO model incorporates the convolutional block attention module (efficient channel attention (ECA)) in the Backbone, adopts the structure of bi-directional feature pyramid network (BiFPN) in the Neck part, adopts the generalized intersection over union (GIOU) function as the loss function in the Prediction part, and adopts the GIOU function as the loss function by setting the weighting coefficient to accelerate the detection of the seed germination rate. In the Prediction part, the GIOU function is used as the loss function to accelerate the learning of high-confidence targets by setting the weight coefficients to further improve the detection accuracy of seed germination rate. The results showed that the accuracy of the SGR-YOLO model for wild rice seed germination discrimination was 94% for the hydroponic box and 98.2% for the Petri dish. The errors of germination potential, germination index, and average germination days detected by SGR-YOLO using the manual statistics were 0.4%, 2.2, and 0.9 days, respectively, in the hydroponic box and 0.5%, 0.5, and 0.24 days, respectively, in the Petri dish. The above results showed that the SGR-YOLO model can realize the rapid detection of germination rate, germination potential, germination index, and average germination days of wild rice seeds, which can provide a reference for the rapid detection of crop seed germination rate.

KEYWORDS

wild rice, germination detection, deep learning, SGR-YOLO, BiFPN

1 Introduction

Wild rice is a natural gene pool of rice germplasm resources, and under a long-term natural environment, it has accumulated many excellent genes that cultivated rice does not possess, and has numerous specific traits that can be utilized for breeding and biotechnology of cultivated rice, which is conducive to the genetic improvement of cultivated rice (Zhong et al., 2000). Wild rice is important for breeding because it possesses many valuable genetic traits and stress tolerance, such as drought tolerance, salt tolerance, and resistance to pests and diseases (Quan et al., 2018; Yang et al., 2022). These traits can help to improve cultivated rice varieties for better yield and adaptability. The study of genome and genetic variation in wild rice can help to improve the yield and quality of modern crops (Tanksley and McCouch, 1997; Tian et al., 2006).

The germination rate of wild rice seeds is very low and untidy, which causes great difficulties in the conservation, identification, and utilization studies of wild rice (Zhao et al., 1998). The germination rate of wild rice seeds is one of the most important indicators of seed quality. Seeds with high germination rate germinate quickly in the field and have a high resistance to adversity; seeds with low germination rate germinate slowly in the field, have irregular seedling emergence, and are susceptible to the effects of the growing environment, which may cause a reduction in the yield of agricultural products (Li and Qin, 1998). Wild rice seed germination is usually assessed manually by counting the number of embryonic sheaths visible in the Petri dish. Seeds are generally considered to have germinated when the length of the embryonic sheath exceeds 2 mm (Tobe et al., 2000). Seed germination labeling is performed by an experienced person, who labels seed categories by visually discerning seed radicle length and germ length. Traditional germination detection is through human eye observation. The germination of seeds germinated for 7 days is measured through statistical judgment (Cheng et al., 2002). The inspector needs to have a wealth of experience, as the repetitive germination rate detection is very cumbersome, time-consuming, laborious, and easy to introduce subjective errors, resulting in inconsistent statistical results between different people and poor reproducibility (Li, 2016). Therefore, there is a need for an objective, reproducible, rapid, and economically reliable method of determination.

In recent years, deep learning has developed rapidly and has been widely used in the field of agriculture, and many researchers have applied deep learning to germination detection of seeds (Yuan et al., 2016; Dang et al., 2020; Zhang et al., 2021). Joosen et al. (2010) chose a semiautomatic approach and designed a germinator to make a judgment of whether a seed has germinated or not by high-throughput scoring, which can handle many samples that may germinate under different environmental conditions. However, a good contrast between the radicle and seed coat is required, which may limit its use in several crops. Zhang et al. (2023) utilized the techniques of image segmentation, transformer encoder, small target detection layer, and CDIOU loss to improve the accuracy of detection and developed a convolutional neural network (YOLO-r) that can detect the germination status of rice seeds and automatically evaluate the total number of germinations. The average accuracy was 0.9539, and the average absolute error in predicting the germination rate mainly existed within 0.1. Bai et al. (2023) constructed DB-YOLOv5, a model for seed germination discrimination, by combining machine vision technology with deep learning methods for rapid detection of germination rate, germination potential, germination index, and average days to germination of wheat seeds and carried out testing experiments. The accuracy of the DB-YOLOv5 model for germination discrimination of wheat seeds was 98.5%. Although the above studies achieved good accuracy, they only considered one culture method and a specific period of time and did not take into account the situation of different culture methods and different days.

In this study, a germination discrimination model for wild rice seeds was developed based on a deep learning network model for wild rice seed germination detection. Two different culture methods, namely hydroponic box and Petri dish, were used to detect the seeds in conjunction with a 7-day continuous germination test of wild rice seeds. The SGR-YOLO model was constructed, and based on the YOLOv7 model, the parts of the model were improved by adding the efficient channel attention (ECA) attention mechanism module, the structure of the bi-directional feature pyramid network (BiFPN), and the generalized intersection over union (GIOU) loss function. This method is able to better recognize and evaluate the germination of wild rice seeds, thus providing a fast and accurate solution for the assessment of seed quality. It realizes the detection of the germination rate of wild rice seeds in different culture methods and the rapid detection of germination rate, germination potential, germination index, and average germination days of wild rice seeds. It provides a feasible method for intelligent detection of seed germination rate.

2 Materials and methods

2.1 Construction of wild rice seed germination dataset

2.1.1 Wild rice seed material

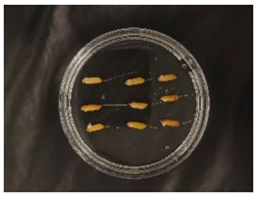
The experimental materials for this study were wild rice numbers 843–1129, with 27 grains of each number, totaling 7,749 seeds, all of which were obtained from the Institute of Crop Science, Chinese Academy of Agricultural Sciences. There are phenotypic differences between different strains of wild rice seeds. Table 1 lists some cereal seeds with different phenotypic characteristics, which vary greatly in length, shape, and color, including medium and short grain lengths; elongated, medium, and short thick shapes; light brown, brown, dark brown, and red colors; and awned versus awnless awns.

2.1.2 Image acquisition

Normally, seed germination is conducted using two methods—hydroponic box and Petri dish (Wang et al., 1996; Wang et al., 2002)—which can save test time and reduce contamination of seeds in the germination process. Therefore, the present study was conducted to test these two methods separately.

10.3389/fpls.2023.1305081

TABLE 1 Seeds with different phenotypic characteristics of seed grain.

Image	Grain length	Grain shape	Grain color	Awn
	Medium	Slender	Brown	×
	Medium	Medium	Brown	×
	Medium	Medium	Light brown	√
	Medium	Medium	Red	√
	Short	Bold	Dark brown	×
	Short	Bold	Red	×
	Medium	Medium	Dark brown	√
	Short	Medium	Light brown	√

(Continued)

10.3389/fpls.2023.1305081

TABLE 1 Continued

Image	Grain length	Grain shape	Grain color	Awn
	Medium	Slender	Red	×

In this experiment, seeds with full grains were baked in an oven at 40°C–45°C for 2–3 days, sterilized with 2%~3% sodium hypochlorite for 0.5 h, and soaked in an incubator at a constant temperature of 37°C, which was used to break seed dormancy (Lei et al., 2004). Next, the treated seeds were placed in Petri dishes and hydroponic boxes. For the Petri dish germination method, nine wild rice seeds of the same variety were placed in a single Petri dish (Figure 1C), and an appropriate amount of distilled water was added to each Petri dish in order to prevent water evaporation. The Petri dishes were then incubated at a constant temperature of 28°C. For the hydroponic box germination method, seeds were selected and transplanted into 96-well black hydroponic boxes containing fresh water (Figure 1A), and the light incubator conditions were set at 16,000 lx of light, 28°C, 12 h; darkness, 28°C, 12 h; and 75% humidity. To ensure that each seed could absorb nutrients evenly, seeds were placed according to varieties, with one variety in each row and another seed in the first row on the right side (as shown in Figure 1B) for a 7-day germination test.

The seeds were placed in Petri dishes and hydroponic boxes, with 1,123 seeds in Petri dishes and 6,626 seeds in hydroponic boxes, totaling 7,749 seeds.

In order to detect the whole process of seed germination, images were acquired from day 1 to day 7 by image acquisition performed on Petri dishes and hydroponic cassettes separately. The first image was taken immediately after the seeds were put into the incubator as the initial image (recorded as day 0), and the images were taken every 24 h for seven consecutive days, as shown in Figure 2. The image acquisition method was vertical shooting, the shooting time was 16:00–18:00, the original image size was 4,096 × 3,072, the acquisition device was a realme cell phone with a 50-megapixel camera and a 2-megapixel rear camera, the shooting process was set to two times digital zoom, and the image format was JPG. The photos taken were characterized by high quality and full color.

2.2 Constructing the dataset

For the two germination methods of hydroponic box and Petri dish, the data of the whole process of wild rice seed germination were collected, and the dataset of seed germination rate detection was formed. As shown in Table 2, in terms of days, on day 0, there were 310 frames; on day 1, there were 293 frames; on day 2, there were 292 frames; on day 3, there were 283 frames; on day 4, there were 281 frames; on day 5, there were 287 frames; on day 6, there were 284 frames; and on day 7, there were 280 frames. In terms of the culture method, there were 1,108 frames in hydroponic boxes and 1,202 frames in Petri dishes, totaling 2,310 frames.

The wild rice seeds in the captured images were labeled using the labelImg software, as shown in Figure 3, where the images were labeled using a rectangular box, with the seeds located in the middle of the rectangular box, and the position of the rectangular box was determined by the coordinates of its two diagonal corners. Different rectangular boxes were labeled in each image according to the location of the wild rice seeds.

The seeds in each image were labeled using labelImg software. The dataset was divided according to stratified sampling, and a total of 1,108 images were obtained from the hydroponic box, with the division ratio of training set:validation set:testing set = 7:2:1, which

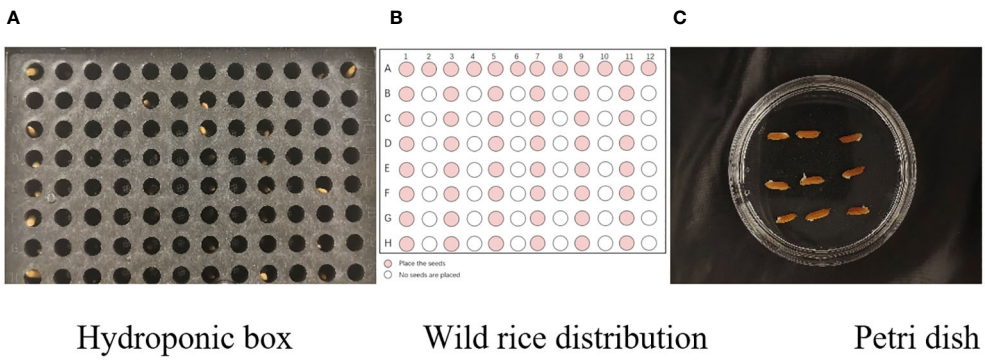


FIGURE 1
Wild rice placement. **(A)** Hydroponic box. **(B)** Wild rice distribution. **(C)** Petri dish.

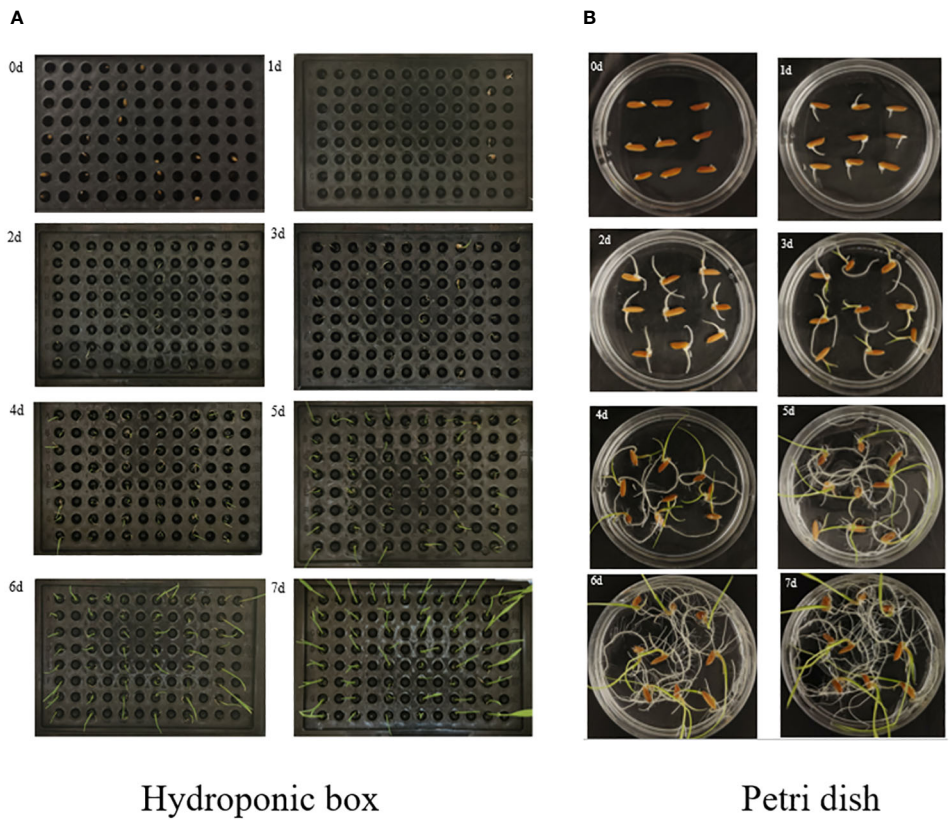


FIGURE 2
Capturing images. **(A)** Hydroponic box. **(B)** Petri dish.

corresponded to the number of images collected as 776, 222, and 110, respectively. A total of 1,202 images were obtained from Petri dishes, and the division ratio was training set:validation set:test set = 7:2:1, which corresponded to the number of acquired images, as 841, 241, and 120, respectively.

TABLE 2 Sample size.

Days to germination	Hydroponic box image data volume/ frame	Petri dish image data volume/ frame	Aggregate/ frame
0	154	156	310
1	137	156	293
2	135	157	292
3	135	148	283
4	136	145	281
5	141	146	287
6	135	149	284
7	135	145	280
Aggregate/frame	1,108	1,202	2,310

2.3 Wild rice seed germination detection algorithm

2.3.1 The SGR-YOLO network model

The small size of the individual wild rice seeds makes it much more difficult to discern germination problems. It shows different morphologies on different days, with dewy whites starting to appear on the first day and the first leaves starting to grow on the third day. During the growth and development of the seeds, the leaves overlapped and crossed over as the number of days changed, and the leaves were not uniform in size due to different growth rates. In a Petri dish, the buds and roots of the seeds grew at the same time, creating a complex background. In hydroponic boxes, each hole appeared as a reflection, thereby reducing the accuracy of the detection model. Therefore, there was a need to improve the thermal visualization, accuracy, and inference speed of the detection algorithm, and an improved feature extraction module was used. By optimizing the feature extraction capability and functionality in order to better capture the key features of the image, a more accurate and lightweight model structure can be obtained, and the network structure is shown in Figure 4.

The SGR-YOLO network structure adds the bi-directional feature pyramid network structure, the ECA attention mechanism module, and the GIOU loss function on the basis of YOLOv7. The small target detection layer was added to detect small targets. The network structure was based on the C3 module and SPPF as the backbone

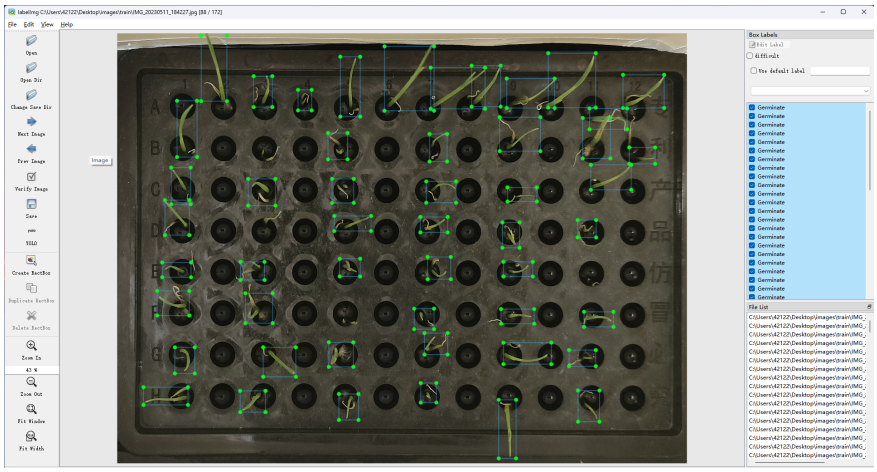


FIGURE 3
Labeling of germinated seeds.

feature extraction network, and the feature pyramid was modified. The model mainly contains four parts: the input (Input), the backbone feature extraction network (Backbone), the neck network (Neck), and the head (Head). The image was sent to the Input after pre-processing operations, such as data enhancement of some of the data, and then sent to the network consisting of the C3 module and CBS module for feature extraction. Subsequently, the extracted features were fused in the Neck using Feature Pyramid Networks (FPNs) to obtain features of three sizes—large, medium, and small—combining the tensor splicing operation of Feature Fusion Networks with the weighted bi-directional feature pyramid BiFPN to target small targets to improve the feature fusion capability. Finally, the Head part generated the predicted category probability and location information of the target. In order to accelerate the convergence of the loss function of the predicted bounding box, the loss function GIOU was used, which effectively improved the performance of the model's bounding box regression and improved the overall performance of the detection model in order to solve the problem of the detection needs of different environments.

2.3.2 YOLOv7

YOLOv7 was proposed by Wang et al. (2022) in August 2022. YOLOv7 utilizes a single-stage detection method, which views the entire detection process as a regression problem. Compared to the traditional two-stage approach, YOLOv7 is faster and can maintain high accuracy in real-time application scenarios. The architecture of YOLOv7 consists of three parts: the Feature Extraction Network (Backbone), the Feature Fusion Layer (Neck), and the Prediction Layer (Head). The Backbone feature extraction network uses a powerful convolutional neural network to extract rich feature representations from the input image. The Neck feature fusion layer fuses feature maps at different scales for detection on targets of different sizes. The Head prediction layer outputs information such as the location, category, and confidence level of the target through

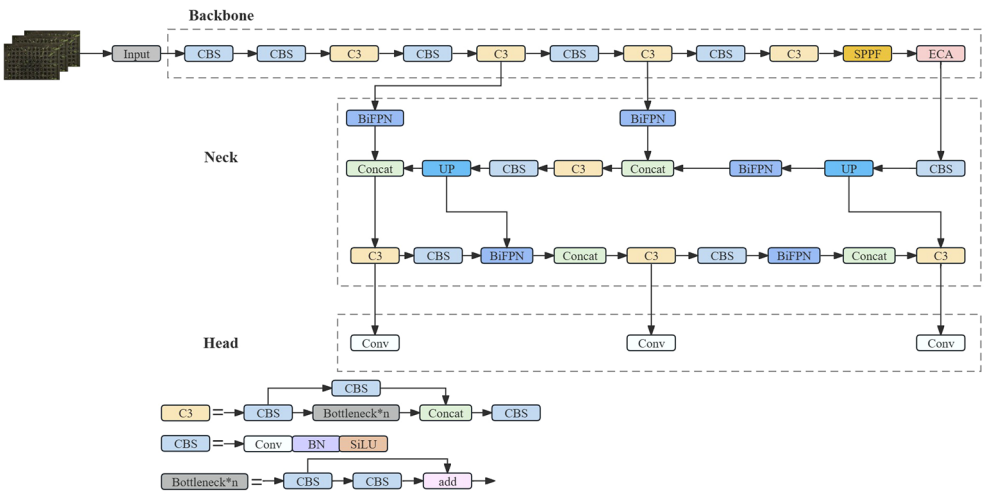


FIGURE 4
SGR-YOLO network structure diagram.

multi-layer convolution and fully connected layers. Therefore, YOLOv7 has fast inference speed and excellent detection performance. It is suitable for real-time applications that require efficient object detection. In conclusion, YOLOv7 is a high-performance object detection model that can accurately and quickly detect a variety of targets.

2.3.3 BiFPN

BiFPN (Tan et al., 2020) is known as the bi-directional feature pyramid network, proposed by Google. BiFPN is a bi-directional feature pyramid network utilizing the idea of bi-directional fusion. In target detection, a feature pyramid can improve the accuracy of target detection because it can combine features of different scales for inspection. BiFPN can realize multi-scale target detection with faster processing speed.

Neck uses an FPN + Path Aggregation Network (PAN) structure. This structure solves the problem of unidirectional information flow limitation of the traditional top-down FPN by adding a bottom-up path aggregation network through PAN. However, due to the dense target density of the wild rice germination detection task, the large number of targets in a single image and the large amount of computation lead to poor real-time performance when the model is applied to the wild rice germination detection task in different environments. In order to ensure the real-time performance of the wild rice germination detection task, BiFPN replaces FPN in Neck, as shown in Figure 5, with P7~P3 representing the layers where the nodes are located.

BiFPN introduces top-down and bottom-up bi-directional connections to realize multi-level information transfer and feedback. This can fully utilize the feature information between different layers, enabling the model to capture the details and contextual information of the object and improve the accuracy and robustness of target detection. BiFPN employs an effective feature fusion strategy, which fuses features from different layers by means of the attention mechanism and layer-by-layer fusion. This strategy enables the model to better balance the information from different layers, enhances the expression of features, and improves the detection performance of the model. The gradient information during backpropagation is passed to the lower-level feature maps through jump connections to accelerate the gradient propagation and improve the convergence speed and training efficiency of the model.

BiFPN mainly realizes cross-layer information transfer and feature fusion. It has two main directions: one is upward convergence from lower feature layers, and the other is downward convergence from higher feature layers. In this way, targeted feature extraction can be performed on feature maps of different layers, and the size variation of the target at different scales can also be handled.

2.3.4 ECA attention mechanism

ECA is a lightweight attention mechanism proposed by Wang et al. (2020). It is an attention model for computer vision tasks. It is mainly used to extract important information from image features for subsequent tasks such as image classification, target detection, and image segmentation. The core idea is to weigh and amplify the important feature channels and suppress the unimportant ones by performing attention operations on the channel dimension of the feature map.

ECA is a dimensionless module for implementing cross-channel interaction as shown in Figure 6. ECA implements a dimensionless cross-channel interaction. In this module, the input

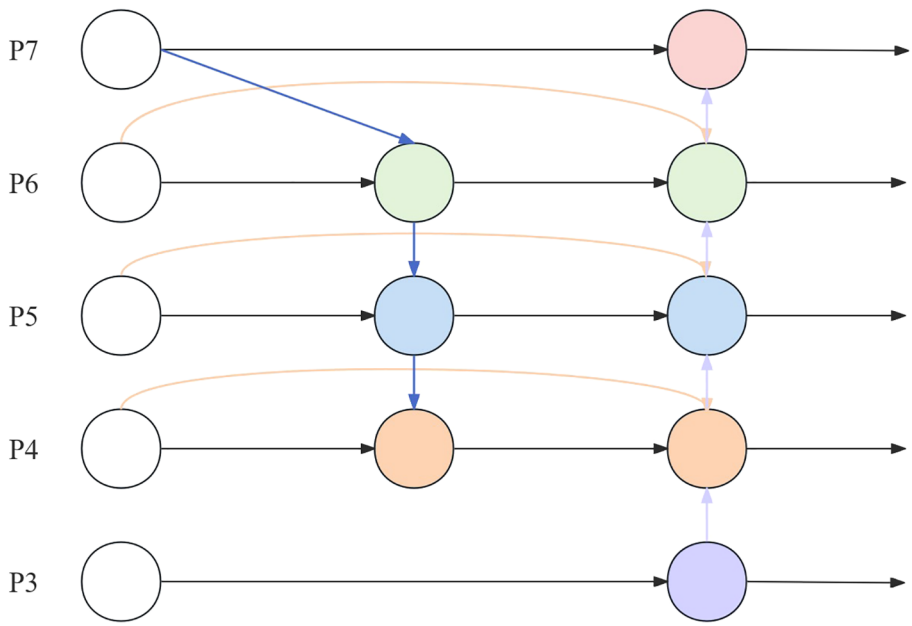


FIGURE 5
BiFPN structure. BiFPN, bi-directional feature pyramid network.

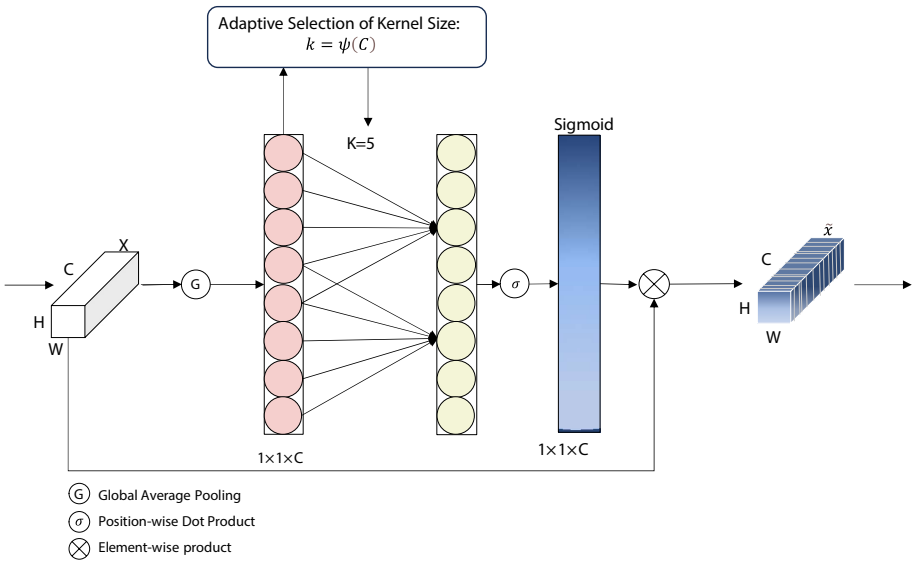


FIGURE 6
Diagram of ECA network structure. ECA, efficient channel attention.

features are first compressed through a (GAP) layer, which in turn is sent into a one-dimensional (1D) convolutional layer for local channel interaction. This is then sent to the Sigmoid function, and this output is then multiplied by the input channel on an element-by-element, basis. The result of the product is the output of the ECA module. In this module, the size of the convolution kernel has a great impact on the sensory field while performing the convolution operation. If the size of the input feature map is large but a small convolution kernel is used, there is a possibility of losing some of the information, and vice versa. Therefore, the concept of dynamic convolution kernel is introduced into the ECA module, and the size of the convolution kernel can be dynamically selected according to the number of channels. ECA captures the local cross-channel interactions by considering each channel and its k nearest neighbors. ECA can be implemented by a fast one-dimensional convolution of size k, where k is the size of the kernel, which indicates the coverage of local cross-channel interactions, i.e., how many immediate neighbors are involved in the attentional prediction of a channel. k is determined using an adaptive determination method, where the kernel size k is proportional to the number of channel dimensions.

The ECA module introduces the feature of dynamic convolution kernels, which allows the selection of different sizes of convolution kernels depending on the number of channels. There exists a mapping relationship between the number of channels and the size of the convolution kernel as shown in Equation 1, and the exact mapping rules will vary depending on the specific implementation. This mechanism of dynamically selecting the size of the convolution kernel allows for the adaptive selection of the appropriate convolution kernel based on the characteristics of the input data, thus improving the performance and adaptability of the model.

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd}, \quad (1)$$

where k denotes the convolutional kernel size, C denotes the channel dimension, and ψ denotes the mapping relationship between k and C. The mapping relationship between k and C is shown in the following table. |t|odd denotes the closest odd number to t. γ and b are set to 2 and 1, respectively.

2.3.5 GIOU loss function

GIOU loss function, proposed by Rezatofighi et al. (2019), is an excellent loss function for target detection tasks. It has the following advantages over other loss functions (e.g., the IOU loss function): first, GIOU is more accurate in terms of the precision of the target frame position and size, which is a better measure of the target detection accuracy. Second, GIOU can effectively prevent overlap and duplication between target frames during the training process, which promotes the stability and robustness of the target detection model. Finally, GIOU is relatively simple to compute, does not require the use of additional parameters or weights, is easy to implement and adjust, and can improve the performance and effectiveness of the target detection model.

The goal of GIOU is equivalent to adding a penalty to the loss function for the closure composed of ground truth and prediction frames, and its penalty term is the area of the closure minus the concatenation of the two frames in the closure that is as small as possible, which is able to more accurately assess the difference between the two bounding boxes. This makes it easier for the network to learn accurate bounding box predictions, leading to improved accuracy and robustness in target detection tasks.

The formula of GIOU is shown in Equation 2.

$$GIOU_{loss} = 1 - GIOU = 1 - \left(IOU - \frac{C - A \cup B}{C}\right) \quad (2)$$

3 Results

3.1 Experimental evaluation indicators

The configuration of the operating environment for this experiment included an operating system environment of Windows 11, a processor of 12th Gen Intel® Core™ i5-12500 3.00 GHz, 32G of machine-banded operating memory, a 1TB SSD, and a graphics card of NVIDIA GeForce RTX 3080 with 10GB of video memory using GPU-accelerated computing. The software environment was as follows: Python 3.9, PyTorch 1.7.0, Torchvision 0.8.2, CUDA 11.0. The number of trial iterations was set to 400, the batch size was set to 8, and Adam was used as the optimizer. The initial learning rate of the model was set to 1e−3, the maximum learning rate was 1e−5, the momentum was 0.937, the weight decay was 0, and the input image resolution was set to 640 × 640. The same training parameters and dataset were used for all the models during the training process.

In order to evaluate the effectiveness of the network model in detecting the germination of wild rice seeds, the trained model was evaluated using the precision rate P (precision), the recall rate R (recall), the F1-score, and the mAP@0.5 (Abouelnaga et al., 2018; Kong et al., 2019; Huang et al., 2021; Mariusz and Marek, 2022) (mean average precision) as assessment metrics so as to validate and compare the performance of the model. Among them, precision rate P denotes the accuracy of the model in predicting the target, recall rate R denotes the success of the model in searching the target, and F1-score is the reconciled average of precision and recall, which is considered to be equally important, with the maximum of 1 and the minimum of 0. mAP@0.5 measures how good the model is in detecting all the categories. The definitions of precision rate P, recall rate R, F1-score, and mAP@0.5 are shown in Equations 3–6.

$$P = \frac{T_P}{T_P + F_P}, \quad (3)$$

$$R = \frac{T_P}{T_P + F_N}, \quad (4)$$

$$V_{F1-score} = 2 \times \frac{P \times R}{P + R}, \quad (5)$$

$$V_{mAP} = \frac{1}{C}\sum_{k=i}^{N} P(k)\Delta R(k) \quad (6)$$

In the formula, TP, FP, and FN denote the number of true-positive, false-positive, and false-negative samples; VF1-score and VmAP represent the values of F1-score and mAP@0.5, respectively; C is the number of categories; N is the number of set thresholds; k is the set threshold; and P(k) and R(k) are the precision and recall, respectively, corresponding to the k value.

In identifying germinated seeds, germinated seeds are the identification target, TP denotes the number of seeds correctly identified as germinated seeds, FP denotes the number of seeds incorrectly identified as germinated seeds, and FN denotes the number of seeds incorrectly identified as ungerminated seeds, precision rate is the ratio of actual germinated seeds among all identified as germinated seeds, and recall rate is the ratio of germinated seeds identified among all actual germinated seeds. In identifying ungerminated seeds, ungerminated seeds are the identification target, TP indicates the number of seeds correctly identified as ungerminated seeds, FP indicates the number of seeds incorrectly identified as ungerminated seeds, FN indicates the number of seeds incorrectly identified as germinated seeds, the precision rate is the rate of actual ungerminated seeds among all identified ungerminated seeds, and the recall rate is the rate of actual ungerminated seeds among all ungerminated seeds that are recognized as ungerminated seeds. The recall rate is the percentage of ungerminated seeds identified out of all ungerminated seeds actually identified.

Germination rate is an important index for detecting seed germination, and in order to validate the germination detection results of this study, the germination rate of manual detection was taken and compared with the germination rate after SGR-YOLO discrimination. In addition, the germination potential, germination index (GI), and mean germination days (MGT) of the manual test were compared with the results of SGR-YOLO. Germination potential refers to the initial count of germination rate on day 3, and germination index and mean days to germination were calculated as shown in Equations 7, 8.

$$V_{GI} = \sum \frac{G_t}{D_t}, \quad (7)$$

$$V_{MGT} = \frac{\sum (G_t + D_t)}{G} \quad (8)$$

In Equations 7, 8, V_{GI} is the value of the germination index, V_{MGT} is the value of average days to germination, D_t is the number of days to germination, G_t is the number of newly germinated seeds per day corresponding to D_t, and G is the germination rate (Zhang and Wang, 2021).

3.2 Analysis of test results of different attention mechanisms

In order to test the effect of different attention mechanism modules on the germination of wild rice seeds, the optimal attention mechanism was screened. The YOLOv7 model was tested on different attention mechanisms for comparison. Replacing different attention mechanisms, such as convolutional block attention module (CBAM) (Woo et al., 2018) attention mechanism, can help deep learning models to better focus and understand important features when processing images. Combining

channel attention and spatial attention enables the model to adaptively select the most meaningful features for the current task, which improves the performance and robustness of the model. The global attention mechanism (GAM) helps the model to better understand the relevance of the input data and better capture the key information, which improves the performance and generalization ability of the model. Squeeze-and-excitation (SE) (Hu et al., 2018) attention mechanism is used to increase the model's importance weights for different features to better learn and process data. It adaptively adjusts the weight of each feature, allowing the model to better focus on task-relevant features and improve model performance. Also included are ECA, Sim AM (A Simple, Parameter-Free Attention Module for Convolutional Neural Networks), and Selective Kernel (SK). The target detection model is trained without changing the parameters other than the backbone. The obtained parameters are analyzed and compared to verify the feasibility of the improved target detection model. The P, R, F1, and mAP@0.5 under different attention mechanisms are compared, and the results are shown in Table 3.

As can be seen from Table 3, the models, after the introduction of the attention mechanism, all have significant improvements in parameters. Among them, the average accuracy of the YOLOv7 network in the hydroponic box was 90.9%, and the average accuracy of the YOLOv7+ECA network was 92.2%, which was improved by 1.3 percentage points. The average accuracy of the YOLOv7 network under the Petri dish was 95.4%, and the average accuracy of the YOLOv7+ECA network was 95.5%, which was improved by 0.1 percentage points. The experimental results show that the introduction of the coordinate attention mechanism is effective for the extraction of wild rice germination in all cases, proving the effectiveness of adding the labeled attention mechanism in the improved version of YOLOv7.

Heat map visualization of the improved YOLOv7 network detection process with the addition of six types of attention mechanisms is shown in Figure 7; after the addition of the attention mechanisms, the attention of each network becomes broad and deep, and the comparison of the heat maps under different days shows that ECA has the best effect. Under the thermodynamic effect of the hydroponic box (Figure 7A), the GAM attention mechanism is slightly worse than ECA, the hydroponic box is denser relative to the Petri dish, and ECA is more effective compared with SimAM, CBAM, SK, and SE attention mechanisms. Under the thermodynamic effect of the Petri dish (Figure 7B), SK and SE are slightly less effective compared with ECA, SimAM, CBAM, and GAM, and ECA is more effective. The experimental results and visualization show that adding the ECA attention mechanism to the network model can effectively improve the detection accuracy of the overall network.

3.3 Experimental results with different loss functions

In order to test the effect of different loss functions on the germination of wild rice seeds, the optimal loss functions were screened and compared with the loss functions tested on the YOLOv7 model. The replacement of different loss functions, such as CIOU, is based on DIOU to increase the loss of the detection box scale and increase the length and width of the loss so that the prediction box will be more in line with the real box. DIOU based on the IOU characteristics, takes into account the shortcomings of the shores of mooching GIOU, a direct regression on the Euclidean distance of the center point of the two frames, to accelerate the convergence. EIOU, on the basis of the CIOU, calculates the

TABLE 3 Comparative analysis of the results of different attention mechanism network modules.

Training method	Network	Precision/%	Recall/%	F1/%	mAP/%
Hydroponic box	YOLOv7	87	91.7	86	90.9
	YOLOv7+CBAM	89.2	83.8	86	91.5
	YOLOv7+SK	89.8	89.9	86	90.6
	YOLOv7+GAM	89.4	88.7	89	88.6
	YOLOv7+SE	93.8	84.6	84	89.5
	YOLOv7+SimAM	89.6	89.1	85	89.4
	YOLOv7+ECA	**86.9**	**87.3**	**89**	**92.2**
Petri dish	YOLOv7	95	96.4	81	95.4
	YOLOv7+CBAM	87.8	96.4	60	94.2
	YOLOv7+SK	90.4	97.2	79	95.3
	YOLOv7+GAM	88.5	95.6	64	94.4
	YOLOv7+SE	91.9	96.7	74	94.7
	YOLOv7+SimAM	91.4	97.3	76	95.4
	YOLOv7+ECA	**86.4**	**97.5**	**74**	**95.5**

The bold values indicates that it has the highest accuracy rate.

10.3389/fpls.2023.1305081

TABLE 4 Comparison of different loss functions.

Training method	Network	Precision/%	Recall/%	F_1/%	mAP/%
Hydroponic box	YOLOv7	87	88.9	86	90.9
	YOLOv7+CIOU	88.6	84.4	86	91.1
	YOLOv7+DIOU	88.2	84.2	86	90.8
	YOLOv7+EIOU	86.6	83.1	85	89.6
	YOLOv7+SIOU	87.9	84.3	86	90.8
	YOLOv7+Focal	86.3	84.8	86	90.1
	YOLOv7+**GIOU**	**88.6**	**84.8**	**87**	**91.5**
Petri dish	YOLOv7	95	96.4	81	95.4
	YOLOv7+CIOU	94.5	96.4	82	94.8
	YOLOv7+DIOU	92.6	96.9	76	95.2
	YOLOv7+EIOU	81.9	92.4	60	89.5
	YOLOv7+SIOU	91.1	96.9	76	94.7
	YOLOv7+Focal	93.5	96.9	76	95
	YOLOv7+**GIOU**	**94.3**	**96.5**	**79**	**95.6**

The bold values indicates that it has the highest accuracy rate.

difference value of width and height to replace the aspect ratio, and at the same time, Focal Loss is introduced to solve the problem of imbalance between difficult and easy samples. GIOU introduces the minimum outer frame label on the basis of IOU characteristics to solve the problem of loss equal to 0 when the detection frame and the real frame do not overlap. SIOU, as well as Focal, trains the target detection model without changing the parameters other than the main stem. The obtained parameters are analyzed and compared to verify the feasibility of the improved target detection model. P, R, F1, and mAP@0.5 are compared under different loss functions.

The effects of different loss functions on the overall performance of the model were analyzed and compared. By replacing different types of bounding box regression loss functions, the prediction box direction drift is improved, and the convergence speed and detection performance of the model are improved. The results are shown in Table 4, which shows that SGR-YOLO, after using GIOU, is significantly improved in all parameters. Among them, the average accuracy of the YOLOv7 network under the hydroponic box was 90.9%, and the average accuracy of the YOLOv7+ECA network was 91.5%, which was improved by 0.6 percentage points. The average accuracy of the YOLOv7 network under the Petri dish was 95.4%, and the average accuracy of the YOLOv7+ECA network was 95.6%, which was improved by 0.2 percentage points. GIOU takes the size of the box into account when calculating the overlap degree, which can more accurately measure the overlap of the box and improve the overall performance of the model.

3.4 Ablation experiments

Ablation experiments were performed on the constructed dataset, and the results are shown in Table 5. YOLOv7 was used as the benchmark model in this experiment. BiFPN structure was adopted in YOLOv7, ECA attention was added in YOLOv7, and the GIOU function was used as a loss function as a way to verify the significance of each module. In the hydroponic box experiment, we found that after adding BiFPN, compared to the initial YOLOv7 model, mAP@0.5 increased by 0.7%. After incorporating the ECA attention mechanism, mAP@0.5 increased by 1.3%. When replacing the loss function with GIOU, mAP@0.5 increased by 0.6%. After using BiFPN together with the ECA module, mAP@0.5 increased by 2.4%. After using BiFPN together with GIOU, mAP@0.5 increased by 2.5%. After using ECA together with GIOU, mAP@0.5 increased by 2.2%. After using BiFPN, ECA, and GIOU together, mAP@0.5 increased by 3.1%. We found that after adding BiFPN, compared to the initial YOLOv7 model, mAP@0.5 increased by 2.3%. After incorporating the ECA attention mechanism, mAP@0.5 increased by 0.1%. After replacing the loss function with GIOU, mAP@0.5 increased by 0.2%. After using BiFPN together with the ECA module, mAP@0.5 increased by 1.7%. After using BiFPN together with GIOU, mAP@0.5 increased by 2.1%. After using ECA together with GIOU, mAP@0.5 increased by 2.5%. After using BiFPN, ECA, and GIOU together, mAP@0.5 increased by 2.8%. All analyses show that the improved model outperforms other models. It has good real-time performance and greatly improves the detection of small targets.

3.5 Comparative tests

In order to evaluate the SGR-YOLO model, under the same experimental conditions, the accuracy P, recall R, harmonic mean F1, and average accuracy were measured. mAP@0.5 of YOLOv5 was compared with YOLOv7 in four aspects. The results are shown in Table 6. The detection performance of the three networks varies.

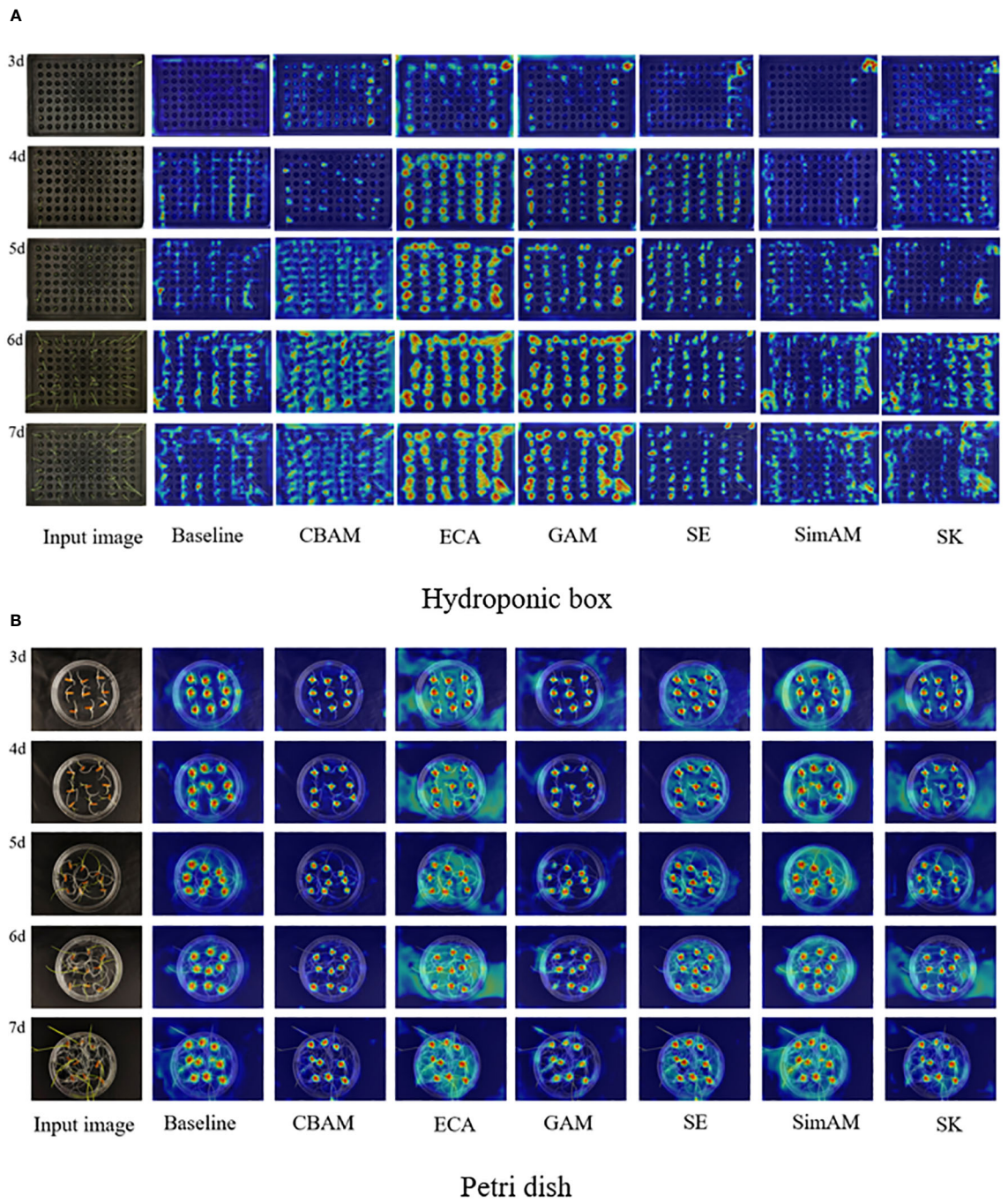


FIGURE 7
Heat map visualization results for different attention mechanism modules.

The SGR-YOLO model has certain advantages in terms of accuracy and recall. The improved model adds the BiFPN structure, which makes the overall model more accurate for the small target detection layer.

The SGR-YOLO network structure proposed in this article was applied to the water culture box, with mAP@0.5 of 94%, and the culture dish, with mAP@0.5 of 98.2%. Compared to the original YOLOv5 and YOLOv7 models, the water culture box increased by 29.8% and 3.1%, respectively, and the plate increased by 19.6% and 2.8%, respectively.

The other values of the SGR-YOLO network structure were higher than the corresponding indexes of the other two models, which proves the effectiveness of the SGR-YOLO network structure in the detection of the germination rate of wild rice seeds.

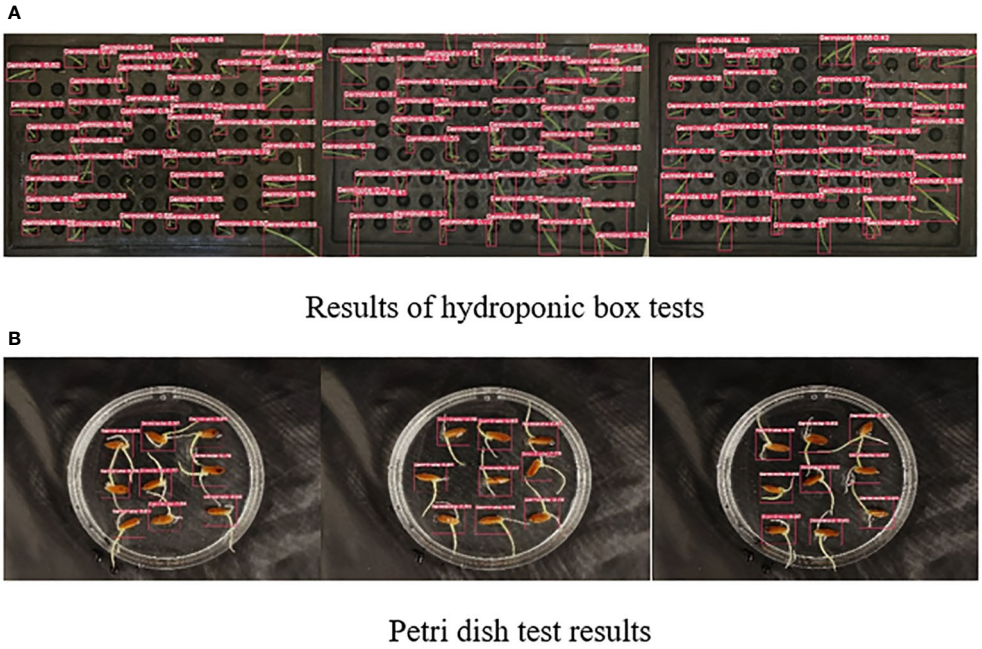


FIGURE 8
Results of SGR-YOLO seed germination test.

3.6 Results of wild rice seed germination rate test

In order to evaluate the performance of the model more accurately, we randomly selected some images for wild rice seed germination detection. Figure 8A demonstrates the detection results of the hydroponic box. During the germination process of wild rice, the leaves overlapped and crossed over as time passed, and the size of the leaves varied due to different growth rates. The SGR-YOLO assay performed very well in the face of these complex conditions.

TABLE 5 Results of ablation experiments for the model.

Training method	Network	Precision/%	Recall/%	F_1/%	mAP/%
Hydroponic box	YOLOv7	87	88.9	86	90.9
	YOLOv7+BiFPN	89	86	88	91.6
	YOLOv7+ECA	86.9	87.3	89	92.2
	YOLOv7+GIOU	88.6	84.8	87	91.5
	YOLOv7+BiFPN+ECA	88.2	89.8	89	93.3
	YOLOv7+BiFPN+GIOU	90.4	90.9	91	93.4
	YOLOv7+ECA+GIOU	91.1	87	89	93.1
	YOLOv7+BiFPN+ECA+GIOU	91.1	88.9	90	94
Petri dish	YOLOv7	95	96.4	81	95.4
	YOLOv7+BiFPN	90.4	98.1	81	97.7
	YOLOv7+ECA	86.4	97.5	74	95.5
	YOLOv7+GIOU	94.3	96.5	79	95.6
	YOLOv7+BiFPN+ECA	88.9	98.1	73	97.1
	YOLOv7+BiFPN+GIOU	88.9	98.8	80	97.5
	YOLOv7+ECA+GIOU	91.5	98.1	85	97.9
	YOLOv7+BiFPN+ECA+GIOU	95.1	97.2	87	98.2

TABLE 6 Comparison of detection performance of different networks.

Training method	Model	Precision/%	Recall/%	F_1/%	mAP/%
Hydroponic box	YOLOv5	73.7	74.4	74	64.8
	YOLOv7	87	88.9	86	90.9
	SGR-YOLO	91.1	91.7	90	94
Petri dish	YOLOv5	85.9	73	78	78.6
	YOLOv7	95	96.4	81	95.4
	SGR-YOLO	95.1	97.2	87	98.2

However, Figure 8B shows the assay results of Petri dishes. In the Petri dishes, the buds and roots of the seeds grew at the same time, resulting in a complex background. The SGR-YOLO assay still performed well under this complex background.

The germination rates of the seeds in the test set were compared between manual detection and the SGR-YOLO model. The seeds did not germinate on days 0 and 1, started germinating on day 2, reached the peak germination period on day 3, with most seeds having germinated, and the germination rate stabilized by day 5. In the hydroponic box, the germination rate of the seeds gradually increased from day 1 to day 7. On day 1, the seeds had not germinated. On day 2, the manual detection reported a germination rate of 31.5%, while SGR-YOLO detected 28.5%. On day 3, the manual detection reported a germination rate of 86.3%, while SGR-YOLO detected 85%. By day 5, the manual detection reported a germination rate of 91%, while SGR-YOLO detected 90.5%. In the petri dish, on day 1, the seeds had not germinated. On day 2, the manual detection reported a germination rate of 36.2%, while SGR-YOLO detected 35.7%. On day 3, the manual detection reported a germination rate of 87.3%, while SGR-YOLO detected 84%. The germination rate stabilized by day 5, with the manual detection reporting a rate of 98.4% and SGR-YOLO detecting 97.4%. Due to simultaneous growth of roots and seedlings in the petri dish, SGR-YOLO occasionally recognized them as separate entities, resulting in a slightly lower germination rate during detection.

Combined with the detection model, the 7-day germination data of the test set were counted, and the seed germination rate, germination potential, germination index, average days to germination, and calculation time were measured. The results are shown in Table 7. The discrepancies between SGR-YOLO and manual detection in terms of germination vigor, germination index, and average germination days in the hydroponic box were 0.4%, 2.2, and 0.9 days. For the petri dish detection, the disparities in germination vigor, germination index, and average germination days between SGR-YOLO and manual detection were 0.5%, 0.5, and 0.24 days. In the hydroponic boxes, the time for manual detection of the test set images was 3,850 s, while SGR-YOLO needed only 32.46 s. In Petri dishes, the time for manual detection of the test set images was 1,920 s, while SGR-YOLO needed only 32.98 s. The SGR-YOLO significantly reduces the detection time by eliminating the need for manual individual processing, whereas manual detection requires processing each item one by one, resulting in lower efficiency. SGR-YOLO can be used to discriminate wild rice seed germination and calculate seed germination rate.

4 Discussion

In order to solve the problem of seed germination rate detection, Jin et al. (2022) used the full spectrum and feature wavelengths selected by principal component analysis (PCA) to

TABLE 7 Comparison of seed germination test indicators.

Training method	Means	Germination rate/%	Germination potential/%	Germination index	Average days to germination/days	Computation time/s
Hydroponic box	Manual detection	91.6	83.4	68.83	2.56	3850
	SGR-YOLO	91.2	83	67.93	2.67	32.46
Petri dish	Manual detection	98.4	86.5	73.87	2.12	1920
	SGR-YOLO	97.4	86	72.76	2.36	32.98

construct a convolutional neural network (CNN) and traditional machine learning methods (support vector machine (SVM) and logistic regression (LR)) for predicting the vigor of different varieties of rice seeds under natural aging conditions. The accuracy of most models was above 85%. Peng et al. (2022) designed the DDST-Center Net algorithm proposed for the automatic monitoring system of seed germination test. The algorithm is computationally efficient, but it is only applicable to seed germination of oilseed rape. Wang et al. (2021). proposed a non-destructive monitoring method for the growth process of cucumber seedlings based on a Kinect camera, which carried out non-destructive monitoring of the germination rate, plant height, leaf area, and other parameters of cucumber seedlings, and the germination rate error was not more than 1.567%. These methods all use machine vision technology to extract germination features and realize seed germination discrimination through morphological detection, but the seed germination characteristics of different crops are different, which leads to the limitations of the application of these methods.

This paper proposed a model method based on a deep learning model to detect the seed germination rate of wild rice. First, dynamic image collection was carried out on two different germination methods, Petri dish and hydroponic box; then, the comparative test of different attention mechanisms and loss functions was carried out, among which the ECA attention mechanism and GIOU function had the best effect, and then the ablation test was carried out. Finally, the improved model SGR-YOLO was used to analyze the germination rate of wild rice seeds, and the following conclusions were drawn.

1. The SGR-YOLO model adds the ECA attention mechanism to the backbone to focus on the feature differences, reduces the data dimension, and improves the accuracy and speed; after the BiFPN structure is introduced in the Neck part, the computational efficiency of the model is improved, but the accuracy is not significantly improved; the detection accuracy is improved by improving the loss function in the Prediction part.
2. By comparing the results of different attention mechanisms, it was found that the ECA attention mechanism has the best effect, the heat map under different days shows that the ECA has the best effect, and the attention of each network becomes broad and deep. Comparing the results of different loss functions, the GIOU function has the best effect. First, when the BiFPN structure was introduced into the hydroponic box, the accuracy rate reached 91.6%, and that of the Petri dish reached 97.7%; then, the ECA attention mechanism and GIOU function were added, and the mAP@0.5 of the model in the hydroponic box was 94%, which was increased by 3.1% compared with the YOLOv7 model, and the mAP@0.5 in the Petri dish was 98.2%, which was 2.8% higher than that of the YOLOv7 model. Compared with the YOLOv5 and YOLOv7 models, the SGR-YOLO model is better than the comparison model in terms of accuracy, recall, F1 value, parameter quantity, and computational cost.
3. In order to further optimize SGR-YOLO, a comparative test was carried out on SGR-YOLO, and the error between the germination rate of SGR-YOLO and the manual detection in the hydroponic box was 0.4% considering the seed germination rate, germination potential, germination index, average germination days, and calculation time. The discrepancy between SGR-YOLO and manual detection in terms of germination rate in the petri dish was 1%. In the hydroponic box, the image time of the manual test set was 3,850 s, and the image time of SGR-YOLO only needed 32.46 s; in the Petri dish, the image time of the manual test set was 1,920 s, and the image time of SGR-YOLO only needed 32.98 s.
4. The SGR-YOLO model can calculate the seed germination rate more intuitively and quickly through multi-day image combination detection, which provides a feasible method for automatic seed germination detection. This research method has certain accuracy and scientificity, which lays a research foundation for the subsequent intelligent identification of the germination rate of wild rice seeds and also provides a certain theoretical basis and reference value for target detection and recognition using images obtained by mobile phones. Seed germination models can be used to assess the quality and vigor of seeds. By predicting the germination rate and germination time of seeds, it is possible to judge the health and viability of seeds. It can help select high-quality seeds and improve the yield and quality of crops.

Data availability statement

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

Author contributions

QY: Writing – original draft. XZ: Writing – review & editing. GZ: Writing – review & editing. JZ: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (No.31971792,No.32160421) and supported by the Project of Sanya Yazhou Bay Science and Technology City (No.SCKJ-JYRC-2023-45), National Key Research and Development

Program of China (No.2022YFF0711805), the Innovation Project of the Chinese Academy of Agricultural Sciences (No.CAAS-ASTIP-2023-AII, No.ZDXM23011), the Special Fund of Chinese Central Government for Basic Scientific Research Operations in Commonweal Research Institutes (No.JBYW-AII-2023-06, No.Y2022XK24, No.Y2022QC17, No.JBYW-AII-2022-14), and Nanfan special project, CAAS Grant Nos. YBXM2312, YDLH01, YDLH07, YBXM10.

Acknowledgments

We thank all the participants in this study.

Conflict of interest

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

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

Abouelnaga, Y., Eraqi, H. M., and Moustafa, M. N. (2018). Real-time Distracted Driver Posture Classification. Monntain View, United States. doi: 10.48550/arXiv.1706.09498

Bai, W. W., Zhao, X. N., Luo, B., Zhao, W., Huang, S., and Zhang, H. (2023). Research on Wheat Seed Germination Detection Method Based on YOLOv5. *Acta Agricult Zhejiangensis* 35, 445–454.

Cheng, X. H., Zhu, Y. J., and Fan, F. Q. (2002). Factors affecting rice seed germination and preventive measure. *Modern Seed Industry* 4, 21.

Dang, M. Y., Meng, Q. K., Gu, F., Gu, B., and Hu, Y. (2020). Rapid recognition of potato late blight based on machine vision. *Trans. Chin. Soc Agric. Eng.* 36, 193–200.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-Excitation Networks. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE 2018, 7132–7141. doi: 10.1109/CVPR.2018.00745

Huang, T., Fu, R., and Chen, Y. (2021). Deep driver behavior detection model based on human brain consolidated learning for shared autonomy systems. *Measurement* 179, 109463. doi: 10.1016/j.measurement.2021.109463

Jin, B. C., Qi, H. N., Jia, L. Q., Tang, Q. Z., Gao, L., Li, Z. N., et al. (2022). Determination of viability and vigor of naturally-aged rice seeds using hyperspectral imaging with machine learning. *Infrared Phys. Technol.* 122, 104097. doi: 10.1016/j.infrared.2022.104097

Joosen, R. V., Kodde, J., Willems, L. A., Ligterink, W., van der Plas, L. H., and Hilhorst, H. W. (2010). GERMINATOR: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. *Plant J.* 62 (1), 148–159. doi: 10.1111/j.1365-313X.2009.04116.x

Kong, D., Hu, S., Wang, J., Liu, Z., Chen, T., Yu, Q., et al. (2019). Study of recall time of associative memory in a memristive Hopfield neural network. *IEEE Access* 7, 58876–58882. doi: 10.1109/ACCESS.2019.2915271

Lei, D. Y., He, Q., Deng, H. B., and Chen, L. Y. (2004). Preliminary exploration of methods for breaking the dormancy of rice seeds. *Hunan Agric. Sci.*, 16–17. doi: 10.16498/j.cnki.hnnykx.2004.05.007

Li, C. H. (2016). Common problems and solutions for detecting rice seed germination rate. *Seed World*, 54–55.

Li, R. B., and Qin, X. Y. (1998). Preliminary report on germination experiment of wild rice seeds. *Seed*, 12–15. doi: 10.16590/j.cnki.1001-4705.1988.01.039

Mariusz, P., and Marek, B. (2022). Neural network training with limited precision and asymmetric exponent. *J. Big Data* 9 (1).

Peng, Q., Tu, L., Wu, Y., Yu, Z., and Tang, G. (2022). Automatic monitoring system for seed germination test based on deep learning. *J. Electrical Comput. Eng* 2022, 15. doi: 10.1155/2022/4678316

Quan, R., Wang, J., Hui, J., Bai, H., Lyu, X., Zhu, Y., et al. (2018). Improvement of salt tolerance using wild rice genes. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02269

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). *Generalized intersection over union: A Metric and a Loss for Bounding Box Regression [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE). doi: 10.1109/CVPR.2019.00075

Tan, M., Pang, R., and Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. *IEEE Computer Society*. 10781–10790.

Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *science* 277 (5329), 1063–1066.

Tian, F., Li, D. L., Fu, Q., Zhu, Z. F., Fu, Y. C., Wang, X. K., et al. (2006). Construction of introgression lines carrying wild rice (Oryza rufipogon Griff.) segments in cultivate drice (Oryza sativa L.) background and characterization of introgressed segments associated with yield-related traits. *TAG Theor. Appl. Genet. Theoretische Und Angewandte Genetik* 112 (3), 570–580.

Tobe, K., Li, X., and Omasa, K. (2000). Seed germination and radicle growth of a halophyte, Kalidium caspicum (Chenopodiaceae). *Ann. Bot.* 85 (3), 391–396. doi: 10.1006/anbo.1999.1077

Wang, A. G., Cao, H., and Chen, Z. H. (2002). A new method for seed germination experiment. *Pratacult Sci.*, 31–32.

Wang, J. Z., Gur, R., Sun, L., and Zhang, Y. (2021). Non-destructive monito ring of plug seedling growth process based on kinect camera. *Trans. Chin. Soc. Agric. Machinery* 52 (2), 227–235.

Wang, Y. R., Li, W. J., and Yu, J. J. (1996). Screening of rice genotypes with low potassium tolerance using nutrient hydroponic method. *Acta Agricult Universitatis Jiangxiensis*, 193–199.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. IEEE Computer Society. 11534–11542.

Wang, C., Bochkovskiy, A., and Liao, H. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.arxiv.

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). Cbam: Convolutional block attention module. *In Proc. Eur. Conf. Comput. Vision (ECCV)*, 3–19. doi: 10.1007/978-3-030-01234-2_1

Yang, Z. Y., Xu, Z. J., Yang, Q. W., and Qiao, W. H.. (2022). Conservation and utilization of genetic resources of wild rice in China. *Rice Sci.* 29 (3), 216–224.

Yuan, J., Zhu, D., Sun, B., Sun, L., Wu, L., Song, Y., et al. (2016). Machine vision based segmentation algorithm for rice seedling. *Acta Agricult Zhejiangensis* 28 (6), 1069–1075.

Zhang, H. S., and Wang, Z. F. (2021). *Seed Science. 3rd* (Beijing: Science Press).

Zhang, H., Wang, C., Dong, H. T., et al. (2021). Study on the seed selection method of cabbage based on machine vision. *J. Agric. Mechanization Res.* 43 (12), 31–36.

Zhao, S. Y., Huang, L. Q., and Mao, X. F. (1998). Research on improving the germination ability of wild rice seeds. *Seed World*, 19–21.

Zhang, H., Wang, C., Dong, H. T., Wu, X. D., Yan, N., and Luo, B. (2023). Deep-learning-based automatic evaluation of rice seed germination rate. *J. Sci. Food Agric.* 103 (4), 1912–1924. doi: 10.1002/jsfa.12318

Zhong, D. B., Luo, L. J., and Ying, C. S. (2000). Research progress on favorable gene transfer in wild rice. *Chin. J. Rice Sci.*, 40–43. doi: 10.16819/j.1001-7216.2000.02.008

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 12 April 2024
DOI 10.3389/fpls.2024.1355902

Check for updates

OPEN ACCESS

EDITED BY
Zhaorong Hu,
China Agricultural University, China

REVIEWED BY
Le Wang,
Jilin University, China
Mei Zheng,
Beijing Academy of Agricultural and Forestry Sciences, China
Yu Long,
Henan University, China

*CORRESPONDENCE
Jitao Zou
jitao.zou@nrc-cnrc.gc.ca
Jingpu Song
jingpu.song@nrc-cnrc.gc.ca

†These authors have contributed equally to this work

RECEIVED 14 December 2023
ACCEPTED 26 March 2024
PUBLISHED 12 April 2024

CITATION
Song J, Mavraganis I, Shen W, Yang H and Zou J (2024) Applying a non-GMO breeding approach with an identified natural variation to reduce food allergen Len c3 in *Lens culinaris* seeds. *Front. Plant Sci.* 15:1355902. doi: 10.3389/fpls.2024.1355902

COPYRIGHT
© 2024 His Majesty the King in Right of Canada. 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.

Applying a non-GMO breeding approach with an identified natural variation to reduce food allergen Len c3 in *Lens culinaris* seeds

Jingpu Song*†, Ioannis Mavraganis†, Wenyun Shen, Hui Yang and Jitao Zou*

Aquatic and Crop Resource Development Research Centre, National Research Council of Canada, Saskatoon, SK, Canada

Lentils (*Lens culinaris*) are produced in diverse agroecological regions and are consumed as one of the most important food legumes worldwide. Lentils possess a nutritional profile from a human health perspective that is not only nutrient dense but also offers a better balance between protein and carbohydrates. However, lentil causes food allergy, which has been a significant concern due to increased consumption in parts of the world. Len c3, a non-specific lipid transfer protein (LTP), was identified as one of the allergens in lentil seeds. In this study, we identified an LTP gene *Lcu.2RBY.4g013600* that encodes the lentil allergen Len c3. We then focused on gene screening from a collection of natural accessions to search for natural mutations of the Len c3 allergen-encoding gene. A natural lentil line M11 was identified with mutations at *LcLTP3b* and low accumulation of vicilin through genomic-assisted approaches. Furthermore, we generated a pool of lentil germplasms with *LcLTP3b* mutation background through crossing the identified lentil plant M11 with two lentil cultivars, CDC Redmoon and CDC Gold. These generated lentil hybrids can be used as a breeding resource targeting at reducing allergen risk in lentil consumption.

KEYWORDS

LTP, Len c3, lentil allergen, legume, natural variation

Introduction

Lentils (*Lens culinaris*) have nutritional, economical, and environmental advantages as an excellent source of protein-based human diet (Kumar et al., 2016). Lentils have twice the level of proteins than cereals and are rich in minerals, polyphenols, and vitamins (Yu et al., 2023). Lentil proteins have been a source of novel food formulations in milk substitute,

curd-like products, meat products, extruded products, and baked goods (Boye et al., 2010). Notwithstanding its superb nutritional characteristics, consumption of lentil proteins causes food allergy among some individuals, particularly in pediatric population.

Lentils are the major causes of IgE-mediated allergic reactions in children of the Mediterranean population (Pascual et al., 1999). In Spain, allergy to lentils ranks the fifth most common cause of IgE-mediated food allergy in children (Martínez San Ireneo et al., 2008; Valenta et al., 2015). In Turkey, where most of the populations consume legumes as staple food, lentils are the sixth most common food allergen in the pediatric population as well (Akarsu et al., 2021). Similarly, in India, where legumes are essential protein sources of vegetarian diet, lentils and chickpeas are recognized as key contributing factors to legume allergy (Patil et al., 2001). Three major allergen groups from lentils have been identified. Len c1 is derived from vicilin, the most abundant component of seed storage protein (López-Torrejón et al., 2003). Len c2 is a 66-kDa biotinylated protein (Sánchez-Monge et al., 2000), and Len c3 is a non-specific lipid transfer protein (LTP) of approximately 9 kDa (Akkerdaas et al., 2012). Len c3 was extracted from germinated lentils seeds and verified as immunologically potential allergens using immunoblot analysis (Akkerdaas et al., 2012). LTPs can easily bind to multiple types of lipid molecules such as fatty acids and phospholipids (Shenkarev et al., 2017). Although Len c3 is less abundant compared with Len c1 and Len c2, LTPs are highly cross-reactive and are considered as one of the main plant allergens (Shenkarev et al., 2017; Halima et al., 2022).

While crop breeding and agronomic advances have greatly increased crop yield, food security on a global scale urgently demands improvements in nutritional quality. Natural variations harbor numerous mutations and abundant historical recombination and are cost effective for population development phenotyping and repeated phenotyping (Liang et al., 2021). A major challenge is to identify and utilize the advantageous traits in a breeding program (Gur and Zamir, 2004). Molecular tools for lentil breeding such as genome sequencing and transcriptome profiling are being developed rapidly (Fedoruk et al., 2013; Haile et al., 2020; Song et al., 2022; Yu et al., 2023), which have rendered this relatively facile for identifying the target genes. In this study, we identified the LTP gene that encodes Len c3 in lentil and performed genetic screening via gene sequencing for LTP mutation from a collection of natural lentil accessions. We further conducted transcriptome analysis of developing seeds to investigate gene expression of vicilin in the identified natural variation. Lastly, the identified natural *LTP* mutation was introduced into two cultivated lines, CDC Redmoon and CDC Gold, to generate an allergen-less germplasm pool.

Materials and methods

Plant materials and growth conditions

Plants were grown in a growth chamber under 16 h light, 23°C and 8 h dark, 18°C, with far red light for flowering. Lentil seeds, collected individually from 400 natural accessions (Supplementary File 1) originated from the Mediterranean regions, were received from Plant Gene Resources of Canada (PGRC). Two lentil cultivated lines, CDC Redmoon and CDC Gold, were also included in this study.

Phylogenetic analysis

Lentil LTP peptide sequences were obtained from lentil genome assembly v2.0 (Ramsay et al., 2019). The information on Len c3 was found on Allergome (www.allergome.org), and the peptide sequence of Len c3, shown as LTP2, was downloaded from NCBI (accession no. A0AT29.1). Nucleotide and peptide sequences were aligned by using MEGA 11 software (Tamura et al., 2021), and phylogenetic tree was generated by using online tool iTOL (Letunic and Bork, 2021).

LcLTP3b gene cloning and sequence alignment

Genomic DNA of each lentil line was isolated from lentil leaves. *LcLTP3b* gene amplification was conducted by using Phusion DNA polymerase (New England Biolabs, MA, USA). The amplified PCR products were purified by using a PCR Purification Kit (Qiagen, Canada) before Sanger sequencing. The sequencing results were analyzed by using MegAlign Pro software.

Identification of *lcltp3b* allele

To whether the hybrid plants harbor the *lcltp3b* allele, the first exon of *LcLTP3b* gene was cloned and subjected to Mse1 treatment for 30 min before agarose gel electrophoresis.

Storage protein analysis

Seed storage protein isolation and separation was performed as described previously (Song et al., 2021). Briefly, three seeds of each lentil line were ground, and 0.1 g was used for protein isolation. Protein samples were separated on a 15% SDS-PAGE gels. After separation by electrophoresis using a Biochrom Novaspec Plus Visible Spectrophotometer (Bio-RAD), the protein gels were stained with Coomassie Brilliant Blue R250 for 30 min, followed by de-staining for 1 h with de-staining solution before imaging with ChemiDoc Imaging System (Bio-RAD).

RNA extraction and data analysis

Total RNA was extracted from dissected embryos and seed coats containing endosperms using a RNeasy plant mini kit (Qiagen, Germany) according to the manufacturer's instruction. For transcriptome sequencing, cDNA libraries were constructed from the isolated RNA samples by using a TruSeq RNA Sample Preparation kit v2 (Illumina). The cDNA libraries were used for

RNAseq. RNAseq was conducted on an Illumina NOVAseq 6000 pair-end sequencing. RNAseq data analysis was conducted as previously described (Song et al., 2022).

Seed imaging and analysis

Seed images were taken using a Canon EOS70D with a MACRO 100 mm lens. Seed diameters were determined by Image J.

Data availability

RNAseq data that support the findings of this study have been deposited in the Gene Expression Omnibus under accession code GSE255951.

Results

LcLTP3b encodes the allergen Len c3 in lentil

To identify the LTP that encodes Len c3, we first conducted a phylogenetic analysis of the known 26 lentil LTP peptide sequences and Len c3. The phylogenetic tree showed that Lcu.2RBY.4g013600 and Len c3 were grouped together (Figure 1A), and further sequence alignment result indicated that they were 100% identical (Supplementary Figure S1). We concluded that lentil LTP gene *Lcu.2RBY.4g013600* (*LcLTP3b*) encodes the allergen protein Len c3.

Identification of natural lentil germplasm with *LcLTP3b* mutation

To perform a genomic screening of a collection of natural lentil germplasms, we first retrieved the genomic DNA sequence of *LcLTP3b* from the lentil genome. Next, we designed a pair of primers: the forward primer (F1: CTAACACCCGTTAAGACATTGC) was at ~100 bp upstream of the translation start site (ATG) and the reverse primer (R1: ATAGCCTTGGAACCGCAACA) was at ~20 bp downstream of the stop codon (TAA) (Figure 1B). Furthermore, we cloned the *LcLTP3b* gene from the 400 lentil lines, respectively, followed by Sanger sequencing. Sequencing results were aligned to the *LcLTP3b* sequence. Among the collected lentil accessions, we identified one line "CN45073" (Supplementary File 1) with several mutations at *LcLTP3b*, including two insertions, one deletion, and one stop-gain point mutation (Figures 1C, D). This lentil line was referred as M11 in this study, and the mutated gene was marked as *lcltp3b*.

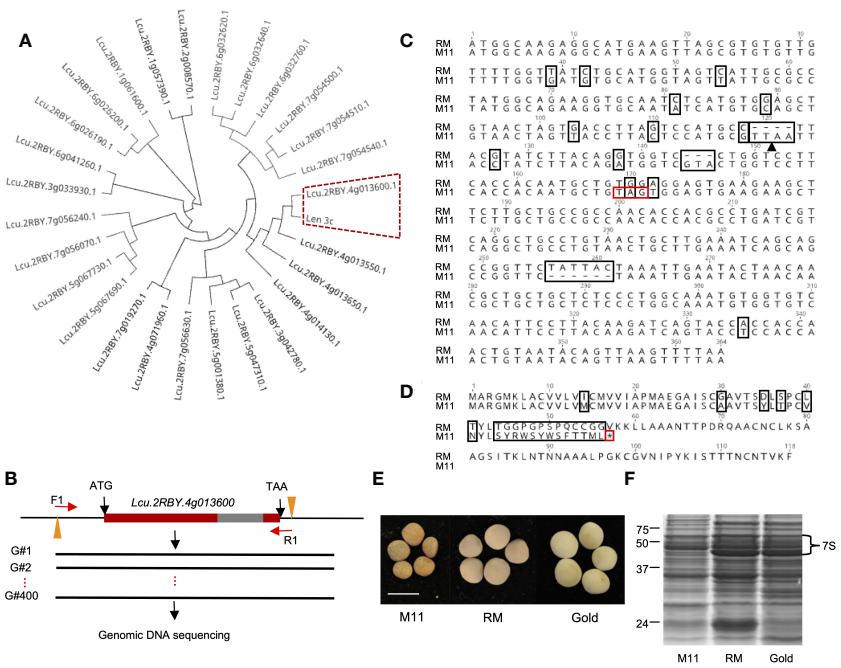


FIGURE 1
Identification of natural lentil line harboring mutated allele of *LTP* gene that encodes Len c3 allergen. **(A)** Phylogenetic analysis of lentil LTP peptides and Len c3. Len c3 and its closest lentil LTP are highlighted with red dash lines. **(B)** Schematic workflow showing the genetic screening on 400 wild lentil lines. Red bars indicate exon, and gray bar represents intron. F1, forward primer; R1, reverse primer. **(C)** Nucleotide sequence alignment of *LcLTP3b* sequenced from CDC Redmoon (RM) and the natural mutation line 11 (M11). All alignment disagreements are highlighted in black boxes; black arrow indicates Mse1 restriction site; the induced stop codon is highlighted with a red box. **(D)** Peptide sequence alignment of LcLTP3b translated from RM and M11. All alignment disagreements are highlighted in black boxes. The induced translation termination site is highlighted with a red box. **(E)** Image of mature seeds collected from M11, RM, and CDC Gold (Gold). Scale bars: 5 mm. **(F)** SDS-PAGE gel image showing the 7S vicilin in seeds from M11, RM, and Gold.

Low vicilin accumulation in the M11 seeds

Lentil seed storage protein quantity and yield are important agronomic traits in lentil breeding (Subedi et al., 2021). Next, we investigated the seed storage protein levels of the identified M11 seeds. We isolated total proteins from M11 mature dry seeds and two CDC (Crop Development Centre) cultivars, CDC Redmoon (Redmoon, hereafter), and CDC Gold (Gold, hereafter) (Figure 1E) and conducted seed storage protein analysis via SDS-PAGE electrophoresis. The results indicated that M11 seeds had much lower 7S protein level compared to Redmoon and Gold (Figure 1F). 7S proteins are enriched with vicilin where another lentil allergen *Len c1* is derived (López-Torrejón et al., 2003). Thus, we performed transcriptome analysis of the mature green seeds to investigate the expression levels of seed storage protein genes encoding 7S vicilin. In seed embryos, the majority of 7S vicilin encoding genes in M11 exhibited lower expression levels compared to Redmoon and Gold, while in the seed coat attaching endosperm tissues, M11 had lower expression levels compared to Redmoon, but higher than that of Gold (Figure 2, Supplementary File 2). Seed storage proteins are mainly accumulated in seed embryos. Overall, these results indicated that M11 had lower accumulation levels of 7S vicilin compared to the other two cultivars.

Creation of a pool of lentil hybrids with allergen-less traits

To incorporate the mutation allele *lcltp3b* into commercial cultivars, we pollinated M11 with pollen grains collected from Redmoon and Gold, respectively. Hybrids of Redmoon × M11 (R×11) and Gold × M11 (G×11) were produced and identified by PCR-based genotyping. We first amplified the full-length genomic DNA of *LcLTP3b* gene with the first set of primers (F1, CTAACACCCGTTAAGACATTGC; R1, ATAGCCTTGG AACCGCAACA) from the plant genomic DNA. Of note, the first insertion site located at the first exon of *lcltp3b* created a Mse1 restriction site. Thus, we took advantage of this mutation site and designed a second set of primers (F2, ATGGCAAGAGGCATGAA; R2, TTAGAAAAAGACATACGTATTAC) to clone only the first exon of *LcLTP3b (LcLTP3b-exon)* by using the genomic DNA of *LcLTP3b* gene as template. After Mse1 treatment on the amplified PCR products, *lcltp3b-exon* originated from M11 was cut into two fragments, while *LcLTP3b-exon* obtained from either Redmoon or Gold remained intact (Figure 3A). We further propagated the second generation to screen for the *lcltp3b* homozygous lines. The *lcltp3b* homozygous lines (five R×11 lines and four G×11) were identified by using the same PCR-based enzyme restriction assays, and their mature seeds were collected, separately (Figure 3B). In addition, agronomic traits of seeds from these lines are documented in Table 1. Taken together, we have successfully incorporated the natural mutation *lcltp3b* into other lentil cultivars and generated a pool of lentil germplasms.

Discussion

Allergens in lentil have been a significant concern due to increased consumption in parts of the world (Sackesen et al., 2020). Numerous efforts have been made to identify and qualify lentil allergens, but lentil allergy management through modifications in food systems requires further development

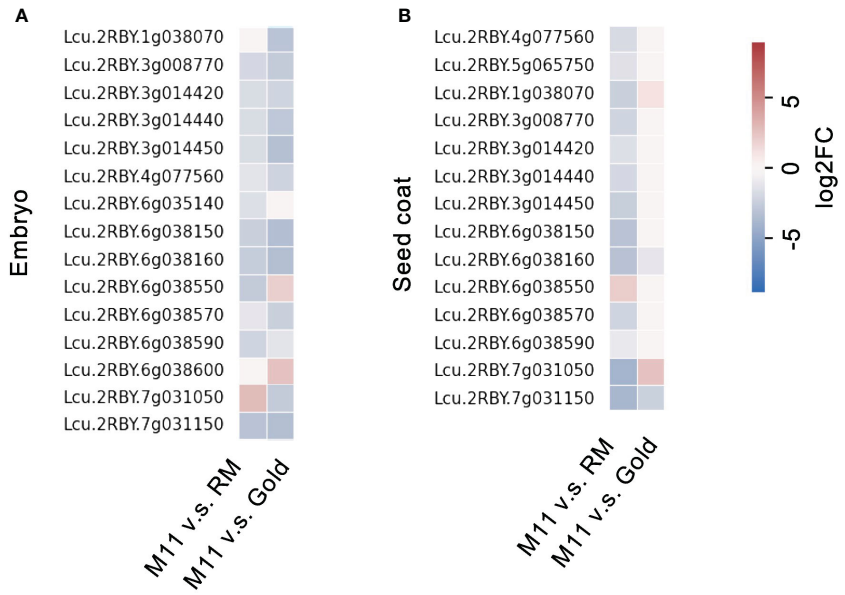


FIGURE 2
Fold change of transcription levels of seed storage protein genes encoding 7S vicilin in the dissected mature green seeds (**A**, embryo; **B**, seed coat with endosperm) of M11 compared to RM and Gold.

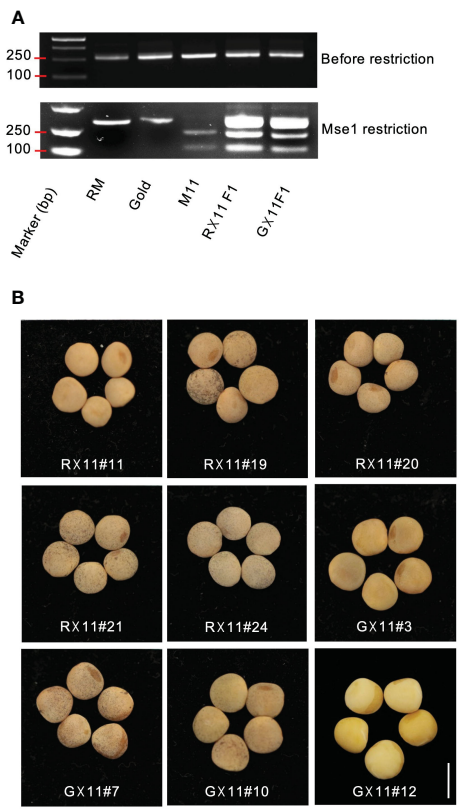


FIGURE 3
Identification of the hybrid lentil lines generated from the crosses of Redmoon×M11 (Rx11) and Gold×M11 (Gx11). **(A)** PCR and Mse1 restriction-based genotyping for lentil hybrid identification. **(B)** Lentil seeds collected from the F2 plants of Rx11 and Gx11 harboring homozygous *lcltp3b*. Scale bars: 5 mm.

(Halima et al., 2022). One goal in crop breeding for better seed nutrition quality is to reduce levels of anti-nutritional factors, such as allergens (Song et al., 2022). In this study, we took a non-GMO approach, focusing on exploring natural variations to search for lentil accessions that harbor natural mutated allergen-encoding genes.

Up to date, three major allergens are known in lentil seeds. Although Len c3 is a rather minor component of lentils' seed protein portfolio, its interacting IgE is detected in 9 out of 10 patients' sera, indicating its prevalence in invoking allergic responses (Akkerdaas et al., 2012). LTPs are a group of plant proteins initially defined by their ability to bind polar lipids in a non-specific manner *in vitro* (Salminen et al., 2016). Voluminous literature indicates that sensitization to LTPs can lead to cross-reaction to homologous food allergens (Rial and Sastre, 2018). From a plant productivity point of view, LTPs represent a minuscule portion of total seed proteins in lentils. Hence, eliminating LTPs will not affect seed protein yield or nutritional profiles. In this study, Len c3-encoding gene *LcLTP3b* was first identified and used as a target for genetic screening to identify natural mutations in a large collection of natural accessions (Figures 1A, B). The screening process identified one natural variation, named M11, harboring mutation at *LcLTP3b*, naming *lcltp3b* (Figure 1C). A previous study has shown that mutations at Thr41, Arg45, and/or Tyr80 significantly affect the ligand-binding capacity and the allergenic potential of Len c3 (Melnikova et al., 2021). The *Lcu.2RBY.4g013600* gene encodes a functional 118-aa (amino acid) protein, while the mutated gene only encodes a 56-aa abnormal protein with substitutions at sites 40/45 (Figure 1D), suggesting that the mutation of *Lcu.2RBY.4g013600* reduces allergens in M11. Moreover, the identified lentil variation M11

TABLE 1 Agronomic traits of Redmoon, Gold, 11 and hybrid seeds.

Genotype	Diameter (cm)	Weight (mg)	Seed coat	
			Color	Dark spots
CDC Redmoon	0.49 ± 0.02	49.0 ± 1.3	red	no
CDC Gold	0.53 ± 0.04	53.6 ± 3.8	golden	no
Mutant(11)	0.41 ± 0.04	37.8 ± 1.2	red	yes
Rx11#11	0.46 ± 0.04	46.0 ± 2.6	red	no
Rx11#19	0.53 ± 0.02	52.7 ± 4.0	red	yes
Rx11#20	0.48 ± 0.02	46.7 ± 1.3	red	yes
Rx11#21	0.50 ± 0.02	51.9 ± 2.8	red	yes
Rx11#24	0.50 ± 0.02	46.0 ± 0.8	red	yes
Gx11#3	0.51 ± 0.02	48.5 ± 3.7	golden	yes
Gx11#7	0.54 ± 0.01	59.0 ± 4.3	red	no
Gx11#10	0.55 ± 0.01	53.1 ± 1.2	red	yes
Gx11#12	0.56 ± 0.01	64.1 ± 2.6	golden	no

has a much lower vicinlin (7S seed storage protein) level, which might be due to a lower transcription activity (Figures 1F, 2) compared to two commercial cultivars, CDC Redmoon and CDC Gold. Taken together, the genetic evidence suggests that M11 appear to produce seeds containing less allergens.

For breeding purpose, we introduced the mutated allele *lcltp3b* from M11 into Redmoon and Gold, which are in different genetic backgrounds through genetic crosses. We have obtained *lcltp3b* homozygous lines from the two crosses, five lines from the Redmoon × M11 crosses and four lines form the Gold × M11 crosses (Figure 3B). The F2 generation lines produced seeds varying in seed size, weight, and seed coat color (Table 1). However, these agronomic traits need continuous breeding process to be fixed. It would be useful to develop a haploid inducer line to shorten the stabilization process of these *lcltp3b* homozygous lines (Gilles et al., 2017; Chen et al., 2023). The *lcltp3b* homozygous lines are the results of different genetic combinations of M11 and other cultivars; thus, they can be directly used as breeding materials in searching for favorable agronomic traits in in the allergen-less background. The findings of this study reveal the potential of lentil natural variation M11 as a breeding material for reducing allergy risk.

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

JS: Investigation, Methodology, Writing – original draft. IM: Formal Analysis, Methodology, Software, Writing – review & editing. WS: Methodology, Writing – review & editing. HY: Methodology, Writing – review & editing. JZ: Conceptualization, Project administration, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the SSP project (A1-022294 to JS) as its contribution to the Aquatic and Crop Resource and Development Centre of the National Research Council of Canada.

Acknowledgments

We thank the Plant Gene Resources of Canada (PGRC) for providing us with the lentil seeds.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1355902/full#supplementary-material

References

Akarsu, A., Ocak, M., Köken, G., Şahiner, Ü.M., Soyer, Ö., and Şekerel, B. E. (2021). Ige mediated food allergy in Turkey: different spectrum, similar outcome. *Turk. J. Pediatr.* 63, 554. doi: 10.24953/turkjped.2021.04.002

Akkerdaas, J., Finkina, E. I., Balandin, S. V., Santos Magadán, S., Knulst, A., Fernandez-Rivas, M., et al. (2012). Lentil (Lens culinaris) Lipid Transfer Protein Len c 3: A Novel Legume Allergen. *Int. Arch. Allergy Immunol.* 157, 51–57. doi: 10.1159/000324946

Boye, J., Zare, F., and Pletch, A. (2010). Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Res. Int.* 43, 414–431. doi: 10.1016/j.foodres.2009.09.003

Chen, X., Li, Y., Ai, G., Chen, J., Guo, D., Zhu, Z., et al. (2023). Creation of a watermelon haploid inducer line via ClDMP3-mediated single fertilization of the central cell. *Hortic. Res.* 10. doi: 10.1093/hr/uhad081

Fedoruk, M. J., Vandenberg, A., and Bett, K. E. (2013). Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers. *Plant Genome* 6. doi: 10.3835/plantgenome2013.05.0012

Gilles, L. M., Martinant, J.-P., Rogowsky, P. M., and Widiez, T. (2017). Haploid induction in plants. 27, R1095–R1097. doi: 10.1111/pbi.1280

Gur, A., and Zamir, D. (2004). Unused natural variation can lift yield barriers in plant breeding. *PloS Biol.* 2, 1610–1615. doi: 10.1371/journal.pbio.0020245

Haile, T. A., Heidecker, T., Wright, D., Neupane, S., Ramsay, L., Vandenberg, A., et al. (2020). Genomic selection for lentil breeding: Empirical evidence. *Plant Genome* 13. doi: 10.1002/tpg2.20002

Halima, O., Najar, F. Z., Wahab, A., Gamagedara, S., Chowdhury, A. I., Foster, S. B., et al. (2022). Lentil allergens identification and quantification: An update from omics perspective. *Food Chemist.: Mol. Sci.* 4. doi: 10.1016/j.fochms.2022.100109

Kumar, J., Gupta, D. S., Kumar, S., Gupta, S., and Singh, N. P. (2016). Current knowledge on genetic biofortification in lentil. *J. Agric. Food Chem.* 64, 6383–6396. doi: 10.1021/acs.jafc.6b02171

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Liang, Y., Liu, H.-J., Yan, J., and Tian, F. (2021). Annual review of plant biology natural variation in crops: realized understanding, continuing promise. *Ann. Rev. Plant Biol.* 72, 357–385 doi: 10.1146/annurev-arplant-080720

López-Torrejón, G., Salcedo, G., Martín-Esteban, M., Díaz-Perales, A., Pascual, C. Y., and Sánchez-Monge, R. (2003). Len c 1, a major allergen and vicilin from lentil seeds: Protein isolation and cDNA cloning. *J. Allergy Clin. Immunol.* 112, 1208–1215. doi: 10.1016/j.jaci.2003.08.035

Martínez San Ireneo, M., Ibáñez, M. D., Fernández-Caldas, E., and Carnés, J. (2008). *In vitro* and in *vivo* Cross-Reactivity Studies of Legume Allergy in a Mediterranean Population. *Int. Arch. Allergy Immunol.* 147, 222–230. doi: 10.1159/000142045

Melnikova, D. N., Finkina, E. I., Bogdanov, I. V., Ignatova, A. A., Matveevskaya, N. S., Tagaev, A. A., et al. (2021). Effect of point mutations on structural and allergenic properties of the lentil allergen len c 3. *Membr. (Basel)* 11, 939. doi: 10.3390/membranes11120939

Pascual, C. Y., Fernandez-Crespo, J., Sanchez-Pastor, S., Padial, M. A., Diaz-Pena, J. M., Martin-Muñoz, F., et al. (1999). Allergy to lentils in Mediterranean pediatric patients. *J. Allergy Clin. Immunol.* 103, 154–158. doi: 10.1016/S0091-6749(99)70539-7

Patil, S. P., Niphadkar, P. V., and Bapat, M. M. (2001). Chickpea: a major food allergen in the Indian subcontinent and its clinical and immunochemical correlation. *Ann. Aller. Asthma Immunol.* 87, 140–145. doi: 10.1016/S1081-1206(10)62209-0

Ramsay, L., Koh, C., Konkin, D., Cook, D., Penmetsa, V., Dongying, G., et al. (2019). *Lens culinaris CDC Redberry genome assembly v2.0*. Available at: https://knowpulse.usask.ca/genome-assembly/Lcu.2RB.

Rial, M. J., and Sastre, J. (2018). Food allergies caused by allergenic lipid transfer proteins: what is behind the geographic restriction? *Curr. Allergy Asthma Rep.* 18, 56. doi: 10.1007/s11882-018-0810-x

Sackesen, C., Erman, B., Gimenez, G., Grishina, G., Yilmaz, O., Yavuz, S. T., et al. (2020). IgE and IgG4 binding to lentil epitopes in children with red and green lentil allergy. *Pediatr. Allergy Immunol.* 31, 158–166. doi: 10.1111/pai.13136

Salminen, T. A., Blomqvist, K., and Edqvist, J. (2016). Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta* 244, 971–997. doi: 10.1007/s00425-016-2585-4

Sánchez-Monge, R., Pascual, C. Y., Díaz-Perales, A., Fernández-Crespo, J., Martín-Esteban, M., and Salcedo, G. (2000). Isolation and characterization of relevant allergens from boiled lentils. *J. Allergy Clin. Immunol.* 106, 955–961. doi: 10.1067/mai.2000.109912

Shenkarev, Z. O., Melnikova, D. N., Finkina, E. I., Sukhanov, S. V., Boldyrev, I. A., Gizatullina, A. K., et al. (2017). Ligand binding properties of the lentil lipid transfer protein: molecular insight into the possible mechanism of lipid uptake. *Biochemistry* 56, 1785–1796. doi: 10.1021/acs.biochem.6b01079

Song, J., Xie, X., Chen, C., Shu, J., Thapa, R. K., Nguyen, V., et al. (2021). LEAFY COTYLEDON1 expression in the endosperm enables embryo maturation in Arabidopsis. *Nat. Commun.* 12, 3963. doi: 10.1038/s41467-021-24234-1

Song, J., Mavraganis, I., Shen, W., Yang, H., Cram, D., Xiang, D., et al. (2022). Transcriptome dissection of candidate genes associated with lentil seed quality traits. *Plant Biol.* 24, 815–826. doi: 10.1111/plb.13426

Subedi, M., Khazaei, H., Arganosa, G., Etukudo, E., and Vandenberg, A. (2021). Genetic stability and genotype × environment interaction analysis for seed protein content and protein yield of lentil. *Crop Sci.* 61, 342–356. doi: 10.1002/csc2.20282

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Valenta, R., Hochwallner, H., Linhart, B., and Pahr, S. (2015). Food allergies: the basics. *Gastroenterology* 148, 1120–1131.e4. doi: 10.1053/j.gastro.2015.02.006

Yu, B., Gao, P., Song, J., Yang, H., Qin, L., Yu, X., et al. (2023). Spatiotemporal transcriptomics and metabolic profiling provide insights into gene regulatory networks during lentil seed development. *Plant J.* 115, 253–274. doi: 10.1111/tpj.16205

frontiers | Frontiers in Plant Science

TYPE Mini Review
PUBLISHED 03 May 2024
DOI 10.3389/fpls.2024.1393438

Check for updates

OPEN ACCESS

EDITED BY
Zhaorong Hu,
China Agricultural University, China

REVIEWED BY
Xin Wei,
Shanghai Normal University, China

*CORRESPONDENCE
Yong Lei
leiyong@caas.cn
Boshou Liao
lboshou@hotmail.com

RECEIVED 29 February 2024
ACCEPTED 18 April 2024
PUBLISHED 03 May 2024

CITATION
Wang Z, Lei Y and Liao B (2024) Omics-driven advances in the understanding of regulatory landscape of peanut seed development.
Front. Plant Sci. 15:1393438.
doi: 10.3389/fpls.2024.1393438

COPYRIGHT
© 2024 Wang, Lei and Liao. 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.

Omics-driven advances in the understanding of regulatory landscape of peanut seed development

Zhihui Wang[1,2], Yong Lei[1*] and Boshou Liao[1*]

[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences (CAAS), Wuhan, China, [2]National Key Laboratory of Crop Genetic Improvement, National Center of Crop Molecular Breeding Technology, National Center of Oil Crop Improvement (Wuhan), Huazhong Agricultural University, Wuhan, China

Peanuts (*Arachis hypogaea*) are an essential oilseed crop known for their unique developmental process, characterized by aerial flowering followed by subterranean fruit development. This crop is polyploid, consisting of A and B subgenomes, which complicates its genetic analysis. The advent and progression of omics technologies—encompassing genomics, transcriptomics, proteomics, epigenomics, and metabolomics—have significantly advanced our understanding of peanut biology, particularly in the context of seed development and the regulation of seed-associated traits. Following the completion of the peanut reference genome, research has utilized omics data to elucidate the quantitative trait loci (QTL) associated with seed weight, oil content, protein content, fatty acid composition, sucrose content, and seed coat color as well as the regulatory mechanisms governing seed development. This review aims to summarize the advancements in peanut seed development regulation and trait analysis based on reference genome-guided omics studies. It provides an overview of the significant progress made in understanding the molecular basis of peanut seed development, offering insights into the complex genetic and epigenetic mechanisms that influence key agronomic traits. These studies highlight the significance of omics data in profoundly elucidating the regulatory mechanisms of peanut seed development. Furthermore, they lay a foundational basis for future research on trait-related functional genes, highlighting the pivotal role of comprehensive genomic analysis in advancing our understanding of plant biology.

KEYWORDS

omics, seed development, peanut, yield, quality

Introduction

As a source of edible vegetable oil and protein, peanut (*Arachis hypogaea* L.) is an oil and economic crop of worldwide importance. The peanut are now cultivated in more than 100 countries, mainly distributed in developing countries in Asia, Africa and South America. The global peanut production has been about 54 million tons annually, with a consistent cultivation area of approximately 31 million hectares (ha) in recent years (http://faostat.fao.org). Peanuts have a high nutritional value, as they are rich in fats (35%~60%) and proteins (22%~35%), and also provide dietary fiber, minerals, vitamins and bioactive macromolecules (Zhao et al., 2020b; Chen et al., 2021; Zhou et al., 2021). The traits of peanuts cover various characteristics such as seed weight, oil content, protein content, fatty acid composition, sucrose content, seed coat color, etc. The improvement of these traits is currently a key area of focus in in peanut genetic breeding. These traits are intricately linked to the expression and regulation of genes during seed development. Consequently, the elucidation of the peanut seed development process based on omics data has become a research hotspot in recent years, shedding light on the regulatory mechanisms governing the formation and variation of essential traits in peanut seeds.

The peanut seed development process spans from flowering to subterranean fruiting, illustrating the unique geocarpic growth habit of peanuts. In this process, the flower pollinates above ground, and then the peg, carrying the fertilized ovule, elongates and burrows into the soil to form the seed. The development process of peanut seeds is highly intricate, governed by numerous genes that regulate various seed traits such as size, weight, oil content, seed coat color, fatty acid composition, and the concentration of functional substances. Therefore, researching the regulatory genes and related molecular mechanisms involved in the peanut development process is of significant importance for the genetic improvement of peanut traits.

The omic-technology with illumina or long sequencing reads was utilized to construct the reference genome of peanut. The genome of the diploid progenitors of cultivated peanut, *A. duranensis* and *A. ipaensis*, was sequenced first (Bertioli et al., 2016; Chen et al., 2016b; Hu et al., 2018), followed by the genome of the cultivated allotetraploid peanut *A. hypogaea* (Bertioli et al., 2019; Zhuang et al., 2019). Since the release of the peanut reference genome, significant progress have been achieved in the investigation of quantitative trait loci (QTL) mapping, expression regulation, epigenetics, and other facets pertaining to seed-related traits (Table 1). These advancements have been facilitated through the comprehensive analysis of diverse omics datasets, such as resequencing data, transcriptome, proteome, metabolome, and epigenome, etc (Figure 1).

QTL mapping and GWAS analysis of peanut seed-related traits driven by re-sequencing data

The foundation for mapping QTLs associated with seed traits in peanuts has been laid through the use of a genetic map, where the quality and precision of QTL mapping, as well as the accurate localization of QTL regions, are significantly influenced by the number and density of markers. Recent research efforts employing specific locus amplified fragment sequencing (SLAF-seq) (Wang et al., 2018b; Zhang et al., 2019; Zhao et al., 2022), genotyping-by-sequencing (GBS) (Zhang et al., 2017), and double-digest restriction-site-associated DNA sequencing (ddRAD-seq) (Wang et al., 2019; Luo et al., 2021; Zhang et al., 2021) have contributed to the generation of over 2,000 SNP markers on genetic maps, underscoring the role of omics data in enhancing genetic analysis. With the declining cost of sequencing, whole-genome re-sequencing has emerged as a powerful approach for generating large-scale SNP markers and constructing high-density genetic maps. Notably, this approach has led to the development of four high-density genetic linkage maps, each containing over 8,000 SNPs (Agarwal et al., 2018, 2019; Liu et al., 2020b; Jiang et al., 2021), making a significant advance in the ability to identify markers on a large scale, especially in the context of low genetic diversity in peanut germplasm (Jiang et al., 2011; Wang et al., 2011; Mukri et al., 2012).

In recent genetic studies, various quantitative trait loci (QTL) associated with seed-related traits have been identified, offering insights into the complex genetic foundations of these traits. Noteworthy discoveries include the identification of one stable QTL linked to seed weight on the terminal regions of chromosome B07 (Wang et al., 2018b). Further research identified additional stable QTLs influencing seed weight located on chromosomes A02 and B06 (Zhang et al., 2019). In terms of nutritional traits, a major QTL, *qA05.1*, was found to have a significant impact on oil, protein, and six fatty acids across diverse environments, highlighting the intricate genetic interactions shaping the nutritional composition of peanuts (Sun et al., 2022). A detailed examination conducted by Hu et al. (2018) identified QTLs related to oleic acid (C18:1), linoleic acid (C18:2), and the oleic-to-linoleic acid ratio (O/L) on chromosomes A03, A04, A09, B09, and B10, illuminating the genetic regulation of fatty acid composition. Liu et al. (2020c) discovered a stable QTL, *qOCA08.1*, on chromosome A08, which explained a substantial proportion of phenotypic variation in oil content. Fine-mapping of this QTL revealed a ~0.8-Mb genomic region harboring two annotated genes influencing oil synthesis, providing vital insights into the genetic determinants of oil-related traits in peanut. Employing BSA-seq technology, Guo et al. (2023) uncovered four QTLs for sucrose content on chromosomes A03 and A06, while Wang et al. (2024) further identified two homologous QTLs on chromosomes A06 and B06, providing valuable information on the genetic factors impacting this essential trait. Furthermore, studies on color traits identified key genes controlling red testa color. QTL analysis and fine-mapping identified the *AhRt2* gene on chromosome 12, associated with a SNP in the third exon, as crucial for red testa color (Zhang et al., 2022b). Additionally, the *AhTc1* gene, encoding an R2R3-MYB transcription factor, was found to regulate purple testa color (Zhao et al., 2020b), while *AhRt1* was mapped to a region on chromosome A03, associated with a bHLH transcription factor gene, further elucidating the genetics underlying testa color in peanut (Chen et al., 2021).

Advances in genome-wide association study (GWAS) analyses have significantly contributed to the understanding of peanut seed traits. Zhao et al. (2022) identified SNP markers associated with

TABLE 1 Reference list for omics-driven research on peanut seed development.

Re-sequencing data were employed for SNP genotyping to construct high-resolution genetic maps, identify quantitative trait loci (QTL), or conduct genome-wide association studies (GWAS) focusing on seed-related traits					
Omics technology	Traits/application	Reference	Omics Data	Traits	Reference
SLAF-seq	Yield-related Traits/QTL	Wang et al., 2018b	BSA-seq	Sucrose content/QTL	Guo et al., 2023
SLAF-seq	Seed weight/QTL	Zhang et al., 2019	BSA-seq	Red testa/QTL	Zhang et al., 2022b
SLAF-seq	Seed weight/QTL	Zhao et al., 2022	ddRAD-seq	Yield-related traits/GWAS	Wang et al., 2019;
SLAF-seq	Oleic and Linoleic Acid/QTL	Hu et al., 2018	ddRAD-seq	Fatty acid components/GWAS	Zhang et al., 2021
WGS-seq	Quality traits/QTL	Sun et al., 2022	Axiom_Arachis2 SNP array	Seed weight/GWAS	Zhao et al., 2022
WGS-seq	Sucrose content/QTL	Wang et al., 2024	WGS-seq	Yield-related traits/GWAS	Zhou et al., 2021
WGS-seq	Purple testa/QTL	Zhao et al., 2020b	WGS-seq	Fatty acid components/GWAS	Zhou et al., 2022
ddRAD-seq	Trans-resveratrol content/QTL	Luo et al., 2021	Axiom_Arachis2 SNP array	Fatty acid components/GWAS	Otyama et al., 2022
ddRAD-seq	Oil content/QTL	Liu et al., 2020c	SLAF-seq	Seed weight/GWAS	Zhang et al., 2017
Transcriptome, Protome, Metabolome data or multi-omics data joint analysis were employed for seed development or seed-related traits					
RNA-seq	SD	Clevenger et al., 2016	RNA-seq	Sucrose content	Li et al., 2021a
RNA-seq	SD	Gupta et al., 2016	RNA-seq	Seed coat color	Wan et al., 2016
RNA-seq	SD	Yin et al., 2013	RNA-seq	Seed coat color	Huang et al., 2020
RNA-seq	SD	Zhang et al., 2012	Proteomic	SD and lipid metabolism	Wang et al., 2016;
RNA-seq	SD	Chen et al., 2013	Proteomic	SD and allergen proteins	Li et al., 2020
RNA-seq	SD	Zhu et al., 2014	RNA-seq and DNA Methylation	Oil content	Liu et al., 2022;
RNA-seq	SD	Zhang et al., 2016	Methylation	SD and seed size	Li et al., 2023
RNA-seq	SD	Chen et al., 2016a	CircRNAs	SD and seed size	Feng et al., 2019
RNA-seq	SD	Zhao et al., 2020a	miRNA	SD	Chen et al., 2019a
RNA-seq	SD	Chen et al., 2019b	miRNA	SD	Ma et al., 2018
RNA-seq	SD	Yu et al., 2015	Metabolomics	Seed coat color	Zhang et al., 2022a
RNA-seq	SD	Liu et al., 2020a	Metabolomics	SD	Kefale et al., 2023
RNA-seq	SD	Li et al., 2017	Metabolomics	SD	Li et al., 2022
RNA-seq	SD	Yang et al., 2020	QTL-seq and RNA-seq	Pod length	Lv et al., 2024
RNA-seq	Seed size	Wu et al., 2022	QTL-seq and RNA-seq	Seed weight	Wang Z. et al., 2022
RNA-seq	Seed size	Li et al., 2021b	Metabolomics-Transcriptomics joint analysis	Seed coat color	Xue et al., 2021
RNA-seq	Seed size and Oil content	Guo et al., 2022	Metabolomics-Transcriptomics joint analysis	Seed coat color	Hu et al., 2021

(Continued)

TABLE 1 Continued

Re-sequencing data were employed for SNP genotyping to construct high-resolution genetic maps, identify quantitative trait loci (QTL), or conduct genome-wide association studies (GWAS) focusing on seed-related traits					
Omics technology	Traits/application	Reference	Omics Data	Traits	Reference
Transcriptome, Protome, Metabolome data or multi-omics data joint analysis were employed for seed development or seed-related traits					
RNA-seq	Oil Content	Wang et al., 2018a	Metabolomics-Transcriptomics joint analysis	Seed coat color	Wang X. et al., 2022
RNA-seq	Seed size and oil content	Yang et al., 2023	Metabolomics-Transcriptomics joint analysis	SD	Li et al., 2022
RNA-seq	Oleic acid content	Liu et al., 2018	Metabolomics-Transcriptomics joint analysis	Pod size	Lv et al., 2022
			Lipidomics and proteomicsjoint analysis	Oleic acid content	Liu et al., 2020

WGS-seq, Whole genome resequencing; BSA-seq, Bulked segregant analysis based on deep sequencing; SLAF-seq, Specific locus amplified fragment sequencing; ddRAD-seq, Double digest restriction-site associated sequencing; SD, Seed developmental.

hundred seed weight, branch number, and pod shape. In the Chinese peanut core collection, Zhou et al. (2021) uncovered two major loci exhibiting pleiotropic effects on yield-related traits, explaining about 20% of phenotypic variation. Furthermore, Zhou et al. (2022) identified three stable major associated loci, including two on chromosome A09 for oleic acid and linoleic acid and one on B06 for stearic acid. Extending the research to the USDA peanut core collection, Otyama et al. (2022) explored genetic markers tied to variations in fatty acid composition, unveiling 10 markers affecting oleic and linoleic acid contents, with the alleles having inverse impacts on these acid concentrations. Moreover, Zhang et al. (2017) uncovered 18 significant markers related to seed weight distributed across six chromosomes. Wang et al. (2019) further identified seven peak SNPs associated with yield per plant, pod weight, and seed weight. Zhang et al. (2021) identified five stable significant SNPs associated with oil content and three stable significant SNPs associated with C24:0. Collectively, these GWAS findings offer a comprehensive view of the genetic architecture underlying various peanut seed traits, facilitating targeted breeding efforts for improved cultivars.

The regulation of peanut seed development based on transcriptome data

The study of peanut seed development through transcriptome analysis has led to significant insights, with numerous publications highlighting the intricate genetic networks involved. A pivotal study by Clevenger et al. (2016) sequenced a comprehensive transcriptome map covering 22 tissue types throughout the peanut's reproductive development, from flowering to seed maturation. This work, in conjunction with additional RNA-seq data from Yin et al. (2013); Gupta et al. (2016), and Zhang et al. (2012), has provided a detailed gene expression landscape during seed development.

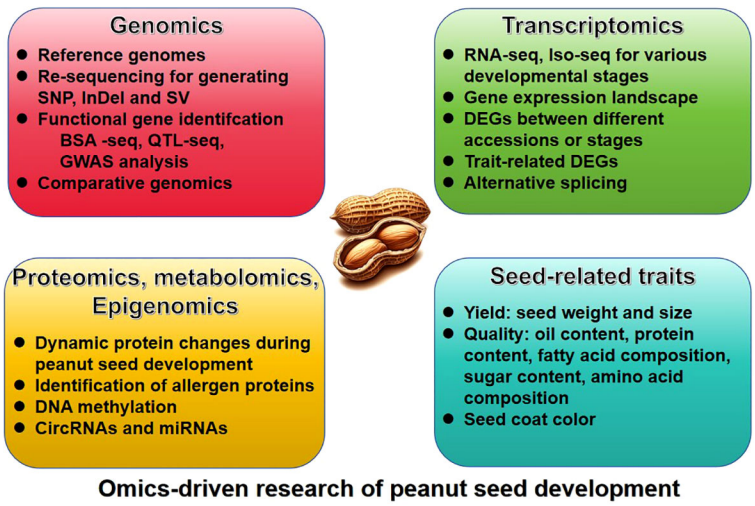


FIGURE 1
The overview of omics-driven research of peanut seed development.

Peanuts exhibit a unique botanical feature: aerial flowering followed by subterranean fruit development. The failure of peg penetration into the soil inhibits the start of pod swelling, resulting in the development of aerial pods and ultimately leading to seed abortion. Comparative transcriptomic analyses between aerial and subterranean pods have identified genes associated with early embryo abortion, including up-regulated photosynthesis-related genes and senescence-associated genes in aerial pods, which may hinder pod swelling (Chen et al., 2013; Zhu et al., 2014). Further transcriptome analyses have identified crucial genes in the embryo and basal regions of the peg, both before and after soil penetration. These genes, including MADS-box transcription factors and cellulose synthase, are vital for embryo development and pod formation (Zhang et al., 2016). Chen et al. (2016a) expanded the research to encompass two whole pod stages and nine stages of isolated pod walls, revealing a developmental gradient of gene expression and highlighting the roles of transcription factors in pod development. Similarly, MADS-box transcription factors play a pivotal role in regulating seed development in both grapevine (Grimplet et al., 2016) and Arabidopsis (Simona et al., 2011), highlighting their fundamental importance across diverse plant species. Zhao et al. (2020a) explored alternative splicing in early swelling pods, finding it mainly related to ovule development, root hair cells enlargement, root apex division, and seed germination. Chen et al. (2019b) and Yu et al. (2015) focused on oil metabolism, identifying over 2,500 genes related to lipid biosynthesis. Their expression patterns during seed development offer insights into peanut lipid biosynthesis.

Studies on the dynamic transcriptomic changes during pod filling have shed light on genotypic variations in lipid metabolism and pod filling efficiency, with the "Hanoch" genotype showing superior pod-filling capabilities (Gupta et al., 2016). Liu et al. (2020a) examined the developmental transcriptome of underground peanut pods and identified 165,689 transcripts, revealing a shift from DNA synthesis and cell division to cell expansion and storage during seed development, with photosynthetic genes active in both aerial and subterranean pods. Moreover, the role of calcium, a crucial signaling molecule, in peanut pod development has been explored. Li et al. (2017) and Yang et al. (2020) investigated the effects of calcium deficiency on gene expression related to calcium signaling and hormone regulation during pod development. Similar to findings in peanuts, studies on wheat and Chinese cabbage reveal the pivotal role of calcium in sustaining plant health, underscoring how calcium deficiency influences gene expressions linked to calcium signaling and hormonal regulation in these diverse agricultural crops (Aslam et al., 2017; Zhang et al., 2022). These studies illustrate the complex interplay between calcium signal transduction, hormone pathways, and the genetic regulation of peanut seed development, providing valuable insights for targeted breeding and genetic improvement initiatives.

The trait-related regulation of peanut seed development based on transcriptome data

Comprehensive transcriptomic analyses have shed light on the genetic mechanisms behind peanut seed development, size, and oil content. For instance, genes such as PNC, YUC, and GASA were found to influence auxin synthesis and seed size, while specific variant sites like GCP4 and RPPL1 within QTL intervals play roles in cell tissue microtubule nucleation (Wu et al., 2022). These findings were consistent with the research in other plants (Cao et al., 2020), which highlighted the importance of auxin and related genes in seed development and grain yield. RNA-seq data from cultivated peanuts and wild Arachis monticola identified genes uniquely expressed during seed development, with certain proteins potentially linked to increased seed size in cultivated varieties (Li et al., 2021b). Differences in gene expression between genotypes with varying seed size and oil content have identified pathways related to plant hormones and fatty acid biosynthesis as critical for seed related traits (Guo et al., 2022).

In terms of oil content, the analysis of 49 cultivars uncovered significant markers on chromosome A03, aiding marker-assisted selection in breeding (Wang et al., 2018a). Comparative studies between peanut varieties with different seed sizes and oil levels identified genes and networks involved in fatty acid synthesis, suggesting strategies for improving seed yield and quality (Yang et al., 2023). Similar to research conducted on peanuts, studies in other crops like safflower, soybean, and rapeseed have also explored key genes and networks involved in regulating traits such as like seed size, oil content, and fatty acid composition (Fan et al., 2023; Guan et al., 2023; Zhao et al., 2023). Research on oleic acid content between cultivars highlighted the role of FAB2 in unsaturated fatty acid biosynthesis and lipid oxidation (Liu et al., 2018), which agreed with the reports in other plants (Dar et al., 2017; He et al., 2020).

For sucrose content, comparative transcriptomics between high- and low-sucrose peanut varieties revealed genes linked to sucrose metabolism, offering targets for molecular breeding (Li et al., 2021a). Studies on seed coat color have utilized transcriptomics to identify genes and markers associated with testa color, providing valuable information for breeding peanuts with desired coat characteristics (Wan et al., 2016; Huang et al., 2020). These findings collectively enhance our understanding of peanut genomics and support targeted breeding efforts for trait improvement.

Deciphering peanut seed development based on proteomic, metabolic, and epigenetic data

The exploration of peanut seed development through integrated approaches combining proteomics, metabolomics, and epigenetics has yielded profound insights into the molecular underpinnings of key seed traits such as size, oil content, allergenicity, and amino acid composition.

Proteomic analyses have significantly advanced our understanding of the dynamic protein changes during peanut seed development. Studies have identified a diverse array of proteins involved in carbohydrate, amino acid, and lipid metabolism, highlighting their crucial roles in seed development (Wang et al., 2016; Li et al., 2020). Some of the proteins identified during the development of peanut seeds were also found in other crops such as rice (Yoon et al., 2023), soybeans

(Wei et al., 2020), wheat (Dong et al., 2015), and barley (Chen et al., 2022). Notably, the identification of allergen proteins and their expression patterns offers valuable insights into allergen accumulation processes, informing breeding strategies aimed at reducing allergenicity (Li et al., 2020). Moreover, the differential expression of proteins related to lipid metabolism during seed development and post-germination stages underscores the complex regulatory mechanisms governing oil accumulation and degradation (Wang et al., 2016).

Epigenetic modifications, specifically DNA methylation, have been elucidated as pivotal regulatory mechanisms influencing peanut seed development. Comparative analyses have demonstrated global methylation changes accompanying seed development, with significant correlations between methylation levels and gene expression, particularly in pathways related to seed size and oil content (Liu et al., 2022; Li et al., 2023). These findings suggest that epigenetic regulation plays a substantial role in modulating seed trait expression.

The regulation of gene expression during seed development has also been a focus, with studies revealing the importance of circRNAs and miRNAs in this process. CircRNAs have been implicated in seed development and size regulation, pointing to their involvement in post-transcriptional regulatory networks (Feng et al., 2019). Similarly, miRNA-mediated regulatory networks have been identified as key contributors to embryo development under calcium deficiency and seed expansion, highlighting the roles of specific miRNAs in modulating gene expression related to growth and development processes (Ma et al., 2018; Chen et al., 2019a).

Metabolomics has recently been embraced in peanut research, offering novel insights into our understanding of its seed develpment. Zhang et al. (2022) employs UPLC-MS/MS to profile metabolites in the testa of four peanut germplasms with varied colors, identifying 85 metabolites and highlighting the significant diversity and differential accumulation of these compounds, including proanthocyanidins, isoflavones, flavonols, and anthocyanidins (Zhang, et al., 2022a). Kefale et al. (2023) identified many differentially accumulated metabolites related to amino acid metabolism, phenylpropanoid biosynthesis, flavonoid biosynthesis, and lipid metabolism between peanut and other oil crops. Li et al. (2022) found that during the early stages of development, most amino acids were present at significantly lower levels. However, this trend shifted in the middle and late stages, where the levels of amino acids were notably higher.

Deciphering peanut seed development and seed-related traits based on joint multi-omics analysis

The joint multi-omics analysis, including genomics-transcriptomics, metabolomics-transcriptomics, proteomics-transcriptomics, serves as a powerful toolkit for elucidating the mechanisms regulating peanut seed development and controlling seed-related traits. In terms of genomics-transcriptomics joint analysis, QTL-seq and RNA-seq have been successfully applied to identify candidate genes for pod length (Lv et al., 2024) and seed weight (Wang Z, et al., 2022). For metabolomics-transcriptomics joint analysis, three studies untilize a integrated approach to investigate the regulatory mechanisms behind testa pigmentation in peanuts. Xue et al. (2021) led this exploration, uncovering the intricate anthocyanin metabolism, highlighting the importance of petunidin 3-O-glucoside and cyanidin O-acetylhexoside in color differentiation. Their analysis identified crucial genes and transcription factors, such as CHS, DFR, MYB, bHLH, and WD40, as pivotal in regulating the distinct pigmentation of peanut testa. Hu et al. (2021) further analyzed the flavonoid biosynthesis pathway, identifying 27 significantly differentially expressed genes (DEGs) associated with testa color development, emphasizing the roles of cyanidin and delphinidin. Wang et al. (2022) broadened this investigation by profiling 133 flavonoids across four peanut cultivars, correlating specific flavonoid components with a variety of testa colors and detailing the roles of cyanidin-based anthocyanins, MYB-like transcription factors, anthocyanidin reductases (ANR), and UDP-glycosyltransferases (UGT) in color modulation (Wang X, et al., 2022). Additionally, transcriptomic and metabolomic analyses have shed light on the genetic and metabolic pathways involved in seed development, with the identification of genes and proteins involved in amino acid metabolism, notably arginine biosynthesis, providing avenues for enhancing the nutritional quality of peanut seeds (Li et al., 2022). Lv et al. (2022) identifies that the accumulation of p-coumaryl alcohol and its associated biosynthesis pathway, particularly the differential expression of gene LOC112771695, plays a critical role in determining peanut pod size. By integrating lipidomics and proteomics, Liu et al. (2020) unravel the complex dynamics of lipid molecular species and their association with the FAD2 mutation in high-oleic acid peanut seeds.

Perspectives

This review demonstrates the pivotal role of omics technologies and related data in achieving a comprehensive peanut reference genome and deepening our insight into the regulatory mechanisms governing peanut seed development. Extensive analysis across various developmental phases, encompassing gene expression, proteomics, metabolomics, and epigenetics, has unveiled the molecular underpinnings and identified key regulatory mechanisms and QTLs linked to seed traits. Such discoveries have substantially contributed greatly to our comprehension of peanut seed development and trait regulation. However, there is still a relative scarcity of research on the cloning and functional study of peanut functional genes. Accelerating the fine mapping of QTLs and employing multi-omics techniques to identify functional genes are essential next steps. Additionally, the application of gene-editing technologies for the improvement of seed traits and the creation of new germplasms represents a crucial research direction for the future. The genome navigation system, including platforms like RiceNavi (Wei et al., 2021), was anticipated to be developed for the purpose of QTN pyramiding and optimizing breeding routes in peanuts. These approaches will not only deepen our understanding

of peanut biology but also facilitate the breeding of varieties with enhanced yield, quality, and environmental resilience.

Author contributions

ZW: Writing – original draft, Writing – review & editing. YL: Writing – original draft, Writing – review & editing. BL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (No. 32201770), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Science (CAAS-ASTIP-2021-OCRI), Natural Science Foundation of Hubei Province (2022CFB332).

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

Agarwal, G., Clevenger, J., Kale, S. M., Wang, H., Pandey, M. K., Choudhary, D., et al. (2019). A recombination bin-map identified a major QTL for resistance to Tomato Spotted Wilt Virus in peanut (*Arachis hypogaea*). *Sci. Rep.* 9, 18246. doi: 10.1038/s41598-019-54747-1

Agarwal, G., Clevenger, J., Pandey, M. K., Wang, H., Shasidhar, Y., Chu, Y., et al. (2018). High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. *Plant Biotechnol. J.* 16, 1954–1967. doi: 10.1111/pbi.12930

Aslam, R., Williams, L. E., Bhatti, M. F., and Virk, N. (2017). Genome-wide analysis of wheat calcium ATPases and potential role of selected ACAs and ECAs in calcium stress. *BMC Plant Biol.* 17, 174. doi: 10.1186/s12870-017-1112-5

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G. D., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nat Genet* 48 (4), 438–46. doi: 10.1038/ng.3517

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D. Y., Seijo, G., et al. (2019). The genome sequence of segmental allotetraploid peanut Arachis hypogaea. *Nat Genet* 51 (5), 877–884. doi: 10.1038/s41588-019-0405-z

Cao, J., Li, G., Qu, D., Li, X., and Wang, Y. (2020). Into the seed: auxin controls seed development and grain yield. *Int. J. Mol. Sci.* 21, 1662. doi: 10.3390/ijms21051662

Chen, H., Chen, X. Y., Xu, R. R., Liu, W. J., Liu, N. A., Huang, L., et al. (2021). Fine-mapping and gene candidate analysis for a major dominant locus responsible for testa color in cultivated peanut. *Theor. And Appl. Genet.* 134, 3721–3730. doi: 10.1007/s00122-021-03924-w

Chen, X. P., Li, H. J., Pandey, M. K., Yang, Q. L., Wang, X. Y., Garg, V., et al. (2016b). Draft genome of the peanut A-genome progenitor provides insights into geocarpy, oil biosynthesis, and allergens. *P Natl. Acad. Sci. U.S.A.* 113, 6785–6790. doi: 10.1073/pnas.1600899113

Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., et al. (2019b). Sequencing of Cultivated Peanut, *Arachis hypogaea*, Yields Insights into Genome Evolution and Oil Improvement. *Mol. Plant* 12, 920–934. doi: 10.1016/j.molp.2019.03.005

Chen, Y., Wang, J., Yao, L., Li, B., Ma, X., Si, E., et al. (2022). Combined proteomic and metabolomic analysis of the molecular mechanism underlying the response to salt stress during seed germination in barley. *Int. J. Mol. Sci.* 23, 10515. doi: 10.3390/ijms231810515

Chen, H., Yang, Q., Chen, K., Zhao, S., Zhang, C., Pan, R., et al. (2019a). Integrated microRNA and transcriptome profiling reveals a miRNA-mediated regulatory network of embryo abortion under calcium deficiency in peanut (*Arachis hypogaea* L.). *BMC Genomics* 20, 392. doi: 10.1186/s12864-019-5770-6

Chen, X., Yang, Q., Li, H., Li, H., Hong, Y., Pan, L., et al. (2016a). Transcriptome-wide sequencing provides insights into geocarpy in peanut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 14, 1215–1224. doi: 10.1111/pbi.12487

Chen, X., Zhu, W., Azam, S., Li, H., Zhu, F., Li, H., et al. (2013). Deep sequencing analysis of the transcriptomes of peanut aerial and subterranean young pods identifies candidate genes related to early embryo abortion. *Plant Biotechnol. J.* 11, 115–127. doi: 10.1111/pbi.12018

Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. (2016). A developmental transcriptome map for allotetraploid *arachis hypogaea*. *Front. Plant Sci.* 7, 1446. doi: 10.3389/fpls.2016.01446

Dar, A. A., Choudhury, A. R., Kancharla, P. K., and Arumugam, N. (2017). The FAD2 Gene in plants: occurrence, regulation, and role. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01789

Dong, K., Zhen, S., Cheng, Z., Cao, H., Ge, P., and Yan, Y. (2015). Proteomic analysis reveals key proteins and phosphoproteins upon seed germination of wheat (*Triticum aestivum* L.). *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01017

Fan, K., Qin, Y., Hu, X., Xu, J., Ye, Q., Zhang, C., et al. (2023). Identification of genes associated with fatty acid biosynthesis based on 214 safflower core germplasm. *BMC Genomics* 24, 763. doi: 10.1186/s12864-023-09874-5

Feng, G., Zhang, M., Ma, S., Zhang, X., Ma, X., Ning, L., et al. (2019). Genome-wide identification of circular RNAs in peanut (*Arachis hypogaea* L.). *BMC Genomics* 20, 653. doi: 10.1186/s12864-019-6020-7

Grimplet, J., Martínez-Zapater, J. M., and Carmona, M. J. (2016). Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genomics* 17. doi: 10.1186/s12864-016-2398-7

Guan, M., Shi, X., Chen, S., Wan, Y., Tang, Y., Zhao, T., et al. (2023). Comparative transcriptome analysis identifies candidate genes related to seed coat color in rapeseed. *Front Plant Sci* 14, 1154208. doi: 10.3389/fpls.2023.1154208

Guo, J. J., Qi, F. Y., Qin, L., Zhang, M. N., Sun, Z. Q., Li, H. Y., et al. (2023). Mapping of a QTL associated with sucrose content in peanut kernels using BSA-seq. *Front. In Genet.* 13. doi: 10.3389/fgene.2022.1089389

Guo, F., Zhu, X., and Zhao, C. (2022). Transcriptome analysis and gene expression profiling of the peanut small seed mutant identified genes involved in seed size control. *Int. J. Mol. Sci.* 23 (17), 9726. doi: 10.3390/ijms23179726

Gupta, K., Kayam, G., Faigenboim-Doron, A., Clevenger, J., Ozias-Akins, P., and Hovav, R. (2016). Gene expression profiling during seed-filling process in peanut with emphasis on oil biosynthesis networks. *Plant Sci.* 248, 116–127. doi: 10.1016/j.plantsci.2016.04.014

He, M., Qin, C. X., Wang, X., and Ding, N. Z. (2020). Plant unsaturated fatty acids: biosynthesis and regulation. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00390

Hu, M., Li, J., Hou, M., Cui, S., Yang, X., Liu, L., et al. (2021). Transcriptomic and metabolomic joint analysis reveals distinct flavonoid biosynthesis regulation for variegated testa color development in peanut (*Arachis hypogaea* L.). *Sci. Rep.* 11, 10721. doi: 10.1038/s41598-021-90141-6

Hu, X. H., Zhang, S. Z., Miao, H. R., Cui, F. G., Shen, Y., Yang, W. Q., et al. (2018). High-density genetic map construction and identification of QTLs controlling oleic and linoleic acid in peanut using SLAF-seq and SSRs. *Sci. Rep.* 8 (1), 5479. doi: 10.1038/S41598-018-23873-7

Huang, L., Liu, X., Pandey, M. K., Ren, X., Chen, H., Xue, X., et al. (2020). Genome-wide expression quantitative trait locus analysis in a recombinant inbred line population for trait dissection in peanut. *Plant Biotechnol. J.* 18, 779–790. doi: 10.1111/pbi.13246

Jiang, Y., Luo, H., Yu, B., Ding, Y., Kang, Y., Huang, L., et al. (2021). High-density genetic linkage map construction using whole-genome resequencing for mapping QTLs of resistance to *aspergillus flavus* infection in peanut. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.745408

Jiang, C., Ramchiary, N., Ma, Y., Jin, M., Feng, J., Li, R., et al. (2011). Structural and functional comparative mapping between the Brassica A genomes in allotetraploid *Brassica napus* and diploid *Brassica rapa*. *Theor. Appl. Genet.* 123, 927–941. doi: 10.1007/s00122-011-1637-1

Kefale, H., Segla, K. D. S., Li, F., Jiang, N., Zhou, R., Wang, L., et al. (2023). Widely targeted metabolic profiling provides insights into variations in bioactive compounds and antioxidant activity of sesame, soybean, peanut, and perilla. *Food Res. Int.* 174, 113586. doi: 10.1016/j.foodres.2023.113586

Li, W., Huang, L., Liu, N., and Pandey, M. K. (2021a). Key regulators of sucrose metabolism identified through comprehensive comparative transcriptome analysis in peanuts. *Int. J. Mol. Sci.* 22 (14), 7266. doi: 10.3390/ijms22147266

Li, H., Liang, X., Zhou, B., Chen, X., Hong, Y., Zhou, R., et al. (2020). A proteomic analysis of peanut seed at different stages of underground development to understand the changes of seed proteins. *PloS One* 15, e0243132. doi: 10.1371/journal.pone.0243132

Li, C., Lai, X., Luo, K., Zheng, Y., Liu, K., and Wan, X. (2022). Integrated metabolomic and transcriptomic analyses of two peanut (*Arachis hypogaea* L.) cultivars differing in amino acid metabolism of the seeds. *Plant. Physiol. Biochem.* 185, 132–143. doi: 10.1016/j.plaphy.2022.05.037

Li, Z., Liu, Q., Zhao, K., Cao, D., Cao, Z., Zhao, K., et al. (2023). Dynamic DNA methylation modification in peanut seed development. *iScience* 26, 107062. doi: 10.1016/j.isci.2023.107062

Li, Y., Meng, J., Yang, S., Guo, F., Zhang, J., Geng, Y., et al. (2017). Transcriptome analysis of calcium- and hormone-related gene expressions during different stages of peanut pod development. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01241

Li, Z., Zhang, X., Zhao, K., Zhao, K., Qu, C., Gao, G., et al. (2021b). Comprehensive transcriptome analyses reveal candidate genes for variation in seed size/weight during peanut (*Arachis hypogaea* L.) domestication. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.666483

Liu, N., Guo, J. B., Zhou, X. J., Wu, B., Huang, L., Luo, H. Y., et al. (2020c). High-resolution mapping of a major and consensus quantitative trait locus for oil content to a ~ 0.8-Mb region on chromosome A08 in peanut. *Theor. And Appl. Genet.* 133, 37–49. doi: 10.1007/s00122-019-03438-6

Liu, H., Hong, Y., Lu, Q., Li, H., Gu, J., Ren, L., et al. (2020). Integrated analysis of comparative lipidomics and proteomics reveals the dynamic changes of lipid molecular species in high-oleic acid peanut seed. *J. Agric. Food Chem.* 68, 426–438. doi: 10.1021/acs.jafc.9b04179

Liu, H., Li, H., Gu, J., Deng, L., Ren, L., Hong, Y., et al. (2018). Identification of the Candidate Proteins Related to Oleic Acid Accumulation during Peanut (*Arachis hypogaea* L.) Seed Development through Comparative Proteome Analysis. *Int. J. Mol. Sci.* 19 (4), 1235. doi: 10.3390/ijms19041235

Liu, H., Liang, X., Lu, Q., Li, H., Liu, H., Li, S., et al. (2020a). Global transcriptome analysis of subterranean pod and seed in peanut (*Arachis hypogaea* L.) unravels the complexity of fruit development under dark condition. *Sci. Rep.* 10, 13050. doi: 10.1038/s41598-020-69943-7

Liu, H., Sun, Z., Zhang, X., Qin, L., Qi, F., Wang, Z., et al. (2020b). QTL mapping of web blotch resistance in peanut by high-throughput genome-wide sequencing. *BMC Plant Biol.* 20, 249. doi: 10.1186/s12870-020-02455-8

Liu, N., Wu, B., Pandey, M. K., Huang, L., Luo, H., Chen, Y., et al. (2022). Gene expression and DNA methylation altering lead to the high oil content in wild allotetraploid peanut (A. monticola). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1065267

Luo, H. Y., Guo, J. B., Yu, B. L., Chen, W. G., Zhang, H., Zhou, X. J., et al. (2021). Construction of ddRADseq-based high-density genetic map and identification of quantitative trait loci for trans-resveratrol content in peanut seeds. *Front. In Plant Sci.* 12. doi: 10.3389/fpls.2021.644402

Lv, Z., Lan, G., Bai, B., Yu, P., Wang, C., Zhang, H., et al. (2024). Identification of candidate genes associated with peanut pod length by combined analysis of QTL-seq and RNA-seq. *Genomics* 116, 110835. doi: 10.1016/j.ygeno.2024.110835

Lv, Z., Zhou, D., Shi, X., Ren, J., Zhang, H., Zhong, C., et al. (2022). Comparative multi-omics analysis reveals lignin accumulation affects peanut pod size. *Int. J. Mol. Sci.* 23, 13533. doi: 10.3390/ijms232113533

Ma, X., Zhang, X., Zhao, K., Li, F., Li, K., Ning, L., et al. (2018). Small RNA and Degradome Deep Sequencing Reveals the Roles of microRNAs in Seed Expansion in Peanut (*Arachis hypogaea* L.). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00349

Mukri, G., Nadaf, H. L., Bhat, R. S., Gowda, M. V. C., Upadhyaya, H. D., and Sujay, V. (2012). Phenotypic and molecular dissection of ICRISAT mini core collection of peanut (Arachis hypogaea L.) for high oleic acid. *Plant Breed.* 131, 418–422. doi: 10.1111/j.1439-0523.2012.01970.x

Otyama, P. I., Chamberlin, K., Ozias-Akins, P., Graham, M. A., Cannon, E. K. S., Cannon, S. B., et al. (2022). Genome-wide approaches delineate the additive, epistatic, and pleiotropic nature of variants controlling fatty acid composition in peanut (*Arachis hypogaea* L.). *G3 (Bethesda)* 12 (1), jkab382. doi: 10.1093/g3journal/jkab382

Simona, M., Lucia, C., Paul, E. G., Arp, S., and Martin, M. K. (2011). The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* 23, 865–872. doi: 10.1105/tpc.110.081737

Sun, Z. Q., Qi, F. Y., Liu, H., Qin, L., Xu, J., Shi, L., et al. (2022). QTL mapping of quality traits in peanut using whole-genome resequencing. *Crop J.* 10, 177–184. doi: 10.1016/j.cj.2021.04.008

Wan, L., Li, B., Pandey, M. K., Wu, Y., Lei, Y., Yan, L., et al. (2016). Transcriptome analysis of a new peanut seed coat mutant for the physiological regulatory mechanism involved in seed coat cracking and pigmentation. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01491

Wang, Z., Huai, D., Zhang, Z., Cheng, K., Kang, Y., Wan, L., et al. (2018b). Development of a high-density genetic map based on specific length amplified fragment sequencing and its application in quantitative trait loci analysis for yield-related traits in cultivated peanut. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00827

Wang, X., Liu, Y., Ouyang, L., Yao, R., He, D., Han, Z., et al. (2022). Metabolomics combined with transcriptomics analyses of mechanism regulating testa pigmentation in peanut. *Front. Plant Sci.* 16. doi: 10.3389/fpls.2022.1065049

Wang, Y., Ma, X., Zhang, X., He, X., Li, H., Cui, D., et al. (2016). ITRAQ-based proteomic analysis of the metabolic mechanisms behind lipid accumulation and degradation during peanut seed development and postgermination. *J. Proteome Res.* 15, 4277–4289. doi: 10.1021/acs.jproteome.6b00345

Wang, M. L., Sukumaran, S., Barkley, N. A., Chen, Z., Chen, C. Y., Guo, B., et al. (2011). Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. *Theor. Appl. Genet.* 123, 1307–1317. doi: 10.1007/s00122-011-1668-7

Wang, X., Xu, P., Yin, L., Ren, Y., Li, S., Shi, Y., et al. (2018a). Genomic and transcriptomic analysis identified gene clusters and candidate genes for oil content in peanut (Arachis hypogaea L.). *Plant Mol. Biol. Rep.* 36, 518–529. doi: 10.1007/s11105-018-1088-9

Wang, Z., Yan, L., Chen, Y., Wang, X., Huai, D., Kang, Y., et al. (2022). Detection of a major QTL and development of KASP markers for seed weight by combining QTL-seq, QTL-mapping and RNA-seq in peanut. *Theor. Appl. Genet.* 135, 1779–1795. doi: 10.1007/s00122-022-04069-0

Wang, J., Yan, C. X., Li, Y., Li, C. J., Zhao, X. B., Yuan, C. L., et al. (2019). GWAS discovery of candidate genes for yield-related traits in peanut and support from earlier QTL mapping studies. *Genes* 10 (10), 803. doi: 10.3390/genes10100803

Wang, Z., Zhang, Y., Huai, D., Chen, Y., Wang, X., Kang, Y., et al. (2024). Detection of two homologous major QTLs and development of diagnostic molecular markers for sucrose content in peanut. *Theor. Appl. Genet.* 137, 61. doi: 10.1007/s00122-024-04549-5

Wei, J., Liu, X., Li, L., Zhao, H., Liu, S., Yu, X., et al. (2020). Quantitative proteomic, physiological and biochemical analysis of cotyledon, embryo, leaf and pod reveals the effects of high temperature and humidity stress on seed vigor formation in soybean. *BMC Plant Biol.* 20, 127. doi: 10.1186/s12870-020-02335-1

Wei, X., Qiu, J., Yong, K., Fan, J., Zhang, Q., Hua, H., et al. (2021). A quantitative genomics map of rice provides genetic insights and guides breeding. *Nat. Genet.* 53, 243–253. doi: 10.1038/s41588-020-00769-9

Wu, Y., Sun, Z., Qi, F., Tian, M., Wang, J., Zhao, R., et al. (2022). Comparative transcriptomics analysis of developing peanut (*Arachis hypogaea* L.) pods reveals candidate genes affecting peanut seed size. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.958808

Xue, Q., Zhang, X., Yang, H., Li, H., Lv, Y., Zhang, K., et al. (2021). Transcriptome and metabolome analysis unveil anthocyanin metabolism in pink and red testa of peanut (*Arachis hypogaea* L.). *Int. J. Genomics* 2021, 5883901. doi: 10.1155/2021/5883901

Yang, S., Wang, J., Tang, Z., Guo, F., Zhang, Y., Zhang, J., et al. (2020). Transcriptome of peanut kernel and shell reveals the mechanism of calcium on peanut pod development. *Sci. Rep.* 10, 15723. doi: 10.1038/s41598-020-72893-9

Yang, L., Yang, L., Ding, Y., Chen, Y., Liu, N., et al. (2023). Global transcriptome and co-expression network analyses revealed hub genes controlling seed size/weight and/or oil content in peanut. *Plants* 12 (17), 3144. doi: 10.3390/plants12173144

Yin, D., Wang, Y., Zhang, X., Li, H., Lu, X., Zhang, J., et al. (2013). *De novo* assembly of the peanut (*Arachis hypogaea* L.) seed transcriptome revealed candidate unigenes for oil accumulation pathways. *PloS One* 8, e73767. doi: 10.1371/journal.pone.0073767

Yoon, J., Min, C. W., Kim, J., Baek, G., Kim, D., Jang, J. W., et al. (2023). Quantitative proteomic analysis deciphers the molecular mechanism for endosperm nuclear division in early rice seed development. *Plants* 12, 3715. doi: 10.3390/plants12213715

Yu, M., Liu, F., Zhu, W., Sun, M., Liu, J., and Li, X. (2015). New features of triacylglycerol biosynthetic pathways of peanut seeds in early developmental stages. *Funct. Integr. Genomics* 15, 707–716. doi: 10.1007/s10142-015-0447-y

Zhang, S., Gao, H., Wang, L., Zhang, Y., Zhou, D., Anwar, A., et al. (2022). Comparative Transcriptome and Co-Expression Network Analyses Reveal the Molecular Mechanism of Calcium-Deficiency-Triggered Tipburn in Chinese Cabbage (*Brassica rapa* L. ssp. Pekinensis). *Plants* 11, 3555. doi: 10.3390/plants11243555

Zhang, S., Hu, X., Miao, H., Chu, Y., Cui, F., Yang, W., et al. (2019). QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (Arachis hypogaea L.). *BMC Plant Biol.* 19, 537. doi: 10.1186/s12870-019-2164-5

Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., et al. (2012). *De novo* assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genomics* 13, 90. doi: 10.1186/1471-2164-13-90

Zhang, K., Ma, J., Gangurde, S. S., Hou, L., Xia, H., Li, N., et al. (2022a). Targeted metabolome analysis reveals accumulation of metabolites in testa of four peanut germplasms. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.992124

Zhang, H., Wang, M. L., Dang, P., Jiang, T., Zhao, S. Z., Lamb, M., et al. (2021). Identification of potential QTLs and genes associated with seed composition traits in peanut using GWAS and RNA-Seq analysis. *Gene* 769, 145215. doi: 10.1016/j.gene.2020.145215

Zhang, Y., Wang, P., Xia, H., Zhao, C., Hou, L., Li, C., et al. (2016). Comparative transcriptome analysis of basal and zygote-located tip regions of peanut ovaries provides insight into the mechanism of light regulation in peanut embryo and pod development. *Genes (Basel)* 17, 606. doi: 10.1186/s12864-016-2857-1

Zhang, K., Yuan, M., Xia, H., He, L. Q., Ma, J., Wang, M. X., et al. (2022b). BSA-seq and genetic mapping reveals as a candidate gene responsible for red testa of peanut. *Theor. And Appl. Genet.* 135, 1529–1540. doi: 10.1007/s00122-022-04051-w

Zhang, X. G., Zhang, J. H., He, X. Y., Wang, Y., Ma, X. L., and Yin, D. M. (2017). Genome-wide association study of major agronomic traits related to domestication in peanut. *Front. In Plant Sci.* 8. doi: 10.3389/fpls.2017.01611

Zhao, X., Li, C., Zhang, H., Yan, C., Sun, Q., Wang, J., et al. (2020a). Alternative splicing profiling provides insights into the molecular mechanisms of peanut peg development. *BMC Plant Biol.* 20, 488. doi: 10.1186/s12870-020-02702-y

Zhao, Y. H., Ma, J. J., Li, M., Deng, L., Li, G. H., Xia, H., et al. (2020b). Whole-genome resequencing-based QTL-seq identified gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. *Plant Biotechnol. J.* 18, 96–105. doi: 10.1111/pbi.13175

Zhao, H. L., Tian, R. Z., Xia, H., Li, C. S., Li, G. H., Li, A. Q., et al. (2022). High-density genetic variation map reveals key candidate loci and genes associated with important agronomic traits in peanut. *Front. In Genet.* 13. doi: 10.3389/fgene.2022.845602

Zhao, X., Wang, J., Xia, N., Liu, Y., Qu, Y., Ming, M., et al. (2023). Combined analysis of the metabolome and transcriptome provides insight into seed oil accumulation in soybean. *Biotechnol. Biofuels* 16 (1), 70. doi: 10.1186/s13068-023-02321-3

Zhou, X. J., Guo, J. B., Pandey, M. K., Varshney, R. K., Huang, L., Luo, H. Y., et al. (2021). Dissection of the genetic basis of yield-related traits in the Chinese peanut mini-core collection through genome-wide association studies. *Front. In Plant Sci.* 12. doi: 10.3389/fpls.2021.637284

Zhou, X. J., Luo, H. Y., Yu, B. L., Huang, L., Liu, N. A., Chen, W. G., et al. (2022). Genetic dissection of fatty acid components in the Chinese peanut minicore collection under multi-environments. *PloS One* 17 (12), e0279650. doi: 10.1371/journal.pone.0279650

Zhu, W., Chen, X., Li, H., Zhu, F., Hong, Y., Varshney, R. K., et al. (2014). Comparative transcriptome analysis of aerial and subterranean pods development provides insights into seed abortion in peanut. *Plant Mol. Biol.* 85, 395–409. doi: 10.1007/s11103-014-0193-x

Zhuang, W. J., Chen, H., Yang, M., Wang, J. P., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865. doi: 10.1038/s41588-019-0402-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 June 2024
DOI 10.3389/fpls.2024.1393621

Check for updates

OPEN ACCESS

EDITED BY
Zhaorong Hu,
China Agricultural University, China

REVIEWED BY
Liang Xiao,
Beijing Forestry University, China
Xinye Liu,
Hebei Normal University, China

*CORRESPONDENCE
Isobel A. P. Parkin
Isobel.Parkin@agr.gc.ca

RECEIVED 29 February 2024
ACCEPTED 29 April 2024
PUBLISHED 05 June 2024

CITATION
Zhang W, Higgins EE, Robinson SJ, Clarke WE, Boyle K, Sharpe AG, Fobert PR and Parkin IAP (2024) A systems genomics and genetics approach to identify the genetic regulatory network for lignin content in *Brassica napus* seeds. *Front. Plant Sci.* 15:1393621. doi: 10.3389/fpls.2024.1393621

COPYRIGHT
© 2024 His Majesty the King in Right of Canada. 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.

A systems genomics and genetics approach to identify the genetic regulatory network for lignin content in *Brassica napus* seeds

Wentao Zhang[1], Erin E. Higgins[2], Stephen J. Robinson[2], Wayne E. Clarke[2], Kerry Boyle[1], Andrew G. Sharpe[3], Pierre R. Fobert[4] and Isobel A. P. Parkin[2]*

[1]Aquatic and Crop Resource Development, National Research Council of Canada, Saskatoon, SK, Canada, [2]Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, [3]Global Institute for Food Security (GIFS), University of Saskatchewan, Saskatoon, SK, Canada, [4]Aquatic and Crop Resource Development, National Research Council of Canada, Ottawa, ON, Canada

Seed quality traits of oilseed rape, *Brassica napus* (*B. napus*), exhibit quantitative inheritance determined by its genetic makeup and the environment via the mediation of a complex genetic architecture of hundreds to thousands of genes. Thus, instead of single gene analysis, network-based systems genomics and genetics approaches that combine genotype, phenotype, and molecular phenotypes offer a promising alternative to uncover this complex genetic architecture. In the current study, systems genetics approaches were used to explore the genetic regulation of lignin traits in *B. napus* seeds. Four QTL (qLignin_A09_1, qLignin_A09_2, qLignin_A09_3, and qLignin_C08) distributed on two chromosomes were identified for lignin content. The qLignin_A09_2 and qLignin_C08 loci were homologous QTL from the A and C subgenomes, respectively. Genome-wide gene regulatory network analysis identified eighty-three subnetworks (or modules); and three modules with 910 genes in total, were associated with lignin content, which was confirmed by network QTL analysis. eQTL (expression quantitative trait loci) analysis revealed four cis-eQTL genes including lignin and flavonoid pathway genes, *cinnamoyl-CoA-reductase* (*CCR1*), and *TRANSPARENT TESTA* genes *TT4*, *TT6*, *TT8*, as causal genes. The findings validated the power of systems genetics to identify causal regulatory networks and genes underlying complex traits. Moreover, this information may enable the research community to explore new breeding strategies, such as network selection or gene engineering, to rewire networks to develop climate resilience crops with better seed quality.

KEYWORDS

systems genomics, systems genetics, lignin, QTL, eQTL, regulatory network, seed quality, *Brassica*

1 Introduction

Understanding the genetic and molecular architecture underlying complex quantitative traits or phenotypic variation is the central goal of current biology (Kliebenstein, 2009). Changing transcript abundance arising from sequence polymorphism is one possible underlying molecular mechanism for complex traits (Rockman and Kruglyak, 2006). A wide range of complex traits within multiple species have been shown to be controlled by their underlying gene expression (Kliebenstein, 2009). A large-scale analysis of the relationship between complex diseases and known non-synonymous single nucleotide polymorphisms (SNPs) found that an increased number of SNPs did not find a significant number of associations, further indicating the potential role of variation in gene expression in controlling phenotypic traits (Burton et al., 2007). Together, these findings support the hypothesis that gene expression acts as the intermediate molecular phenotype, connecting genotypic variation to phenotypic variation (Rockman and Kruglyak, 2006).

The traditional quantitative trait loci (QTLs) mapping approach has been widely used to dissect the molecular mechanism underlying complex traits. However, most complex traits are controlled by multiple genes (Kliebenstein, 2009), and multiple QTLs with marginal effects (Mackay et al., 2009), which can be difficult to dissect with traditional QTL analysis. With the advent of advanced high-throughput technologies, including expression arrays and genome sequencing technologies, a new approach emerged to meet the challenge of complex traits, gene expression quantitative trait locus (eQTL) analyses. This approach is characterized by its combination of population genomic and quantitative genetics (Jansen and Nap, 2001). Briefly, QTL analysis is applied to the transcript abundances measured from a well-genotyped segregating population, allowing genetic loci (eQTL) controlling variation of each transcript to be identified on a genome-wide level (Keurentjes et al., 2007). A locus of sequence polymorphism affecting gene expression that is tightly linked to the physical position of the gene itself is defined as a cis-eQTL; all other eQTLs act in trans (trans-eQTLs) (Brem et al., 2002). The heritability of transcript variation and the genetic effect on this variation can also be inferred from eQTL analysis. Studies from Arabidopsis (West et al., 2007; Keurentjes et al., 2007), yeast (Brem and Kruglyak, 2005), maize (Schadt et al., 2003), and barley (Potokina et al., 2007), revealed an uneven distribution of trans-eQTLs with some genomic regions identified as eQTL hotspots (Schadt et al., 2003), enriched with numerous trans-eQTLs. Identification and dissection of eQTL hotspots may help to identify underlying master regulators or pleiotropic genes, and concomitantly the affected downstream genes or biological pathways. Thus, global eQTL analysis has the potential to yield great insight into the complex genetic architecture of transcript variation, and when linked to phenotypic variation (pQTLs), molecular mechanisms underlying complex traits can be directly inferred.

Several *in silico* large-scale data analyses have demonstrated the power of gene co-expression networks in dissecting a range of biological problems such as secondary wall biosynthesis (Persson et al., 2005), glucosinolate accumulation (Hirai et al., 2007), and seed germination and dormancy (Bassel et al., 2011) at the systems level. Gene regulatory networks were also hypothesized to constitute important information that can be inferred from eQTL data (Jansen and Nap, 2001); yet, this component has often been overlooked. Studies on seed development of Arabidopsis demonstrated that seed filling was controlled by a dynamic and distinct developmental transcriptional program, and genes involved in central carbon transport and multiple other metabolic pathways were implicated (Ruuska et al., 2002). However, most eQTL analyses have been limited to single gene-level analysis.

Fibre in seed meal negatively affects digestion and processibility for animal feed, and reducing the seed fibre content in oilseed rape could result in increased seed oil and protein content (Wittkop et al., 2009; Snowdon et al., 2010; Liu et al., 2012, 2013; Yu et al., 2016; Behnke et al., 2018; Miao et al., 2019; Chao et al., 2022; Yusuf and Möllers, 2024). Therefore, the low fibre trait was identified as a high-value target for canola (*Brassica napus*) breeding programs. Crude fibre is mainly comprised of lignin, including neutral detergent fibre (NDF), acid detergent fibre (ADF), acid detergent lignin (ADL), and polysaccharides including cellulose and hemicellulose (Van Soest et al., 1991; Wittkop et al., 2012). Genetics studies revealed a few major QTLs for lignin traits concentrated on chromosomes A09, C09, and C05 (Badani et al., 2006; Liu et al., 2012, 2013; Stein et al., 2013; Yu et al., 2016; Behnke et al., 2018; Miao et al., 2019; Chao et al., 2022). Lignin biosynthetic pathway(s) are complex. Forward and reverse genetics and biochemical analyses have identified more than 600 involved genes, with a subset of key enzymes that include *phenylalanine ammonia-lyase1* (*PAL1*) and *PAL2*; *cinnamate 4-hydroxylase* (*C4H*); *4-coumarate: CoA ligase1* (*4CL1*) and *4CL2*; *caffeoyl-CoA O-methyltransferase1* (*CCoAOMT1*); *cinnamoyl-CoA reductase1* (*CCR1*); *ferulate 5-hydroxylase* (*F5H1*); *caffeic acid O-methyltransferase* (*COMT*); and *cinnamyl alcohol dehydrogenase6* (*CAD6*) (Vanholme et al., 2010). Previous studies also reported that QTLs for seed colour colocalize with lignin content in *B. napus* seeds (Badani et al., 2006; Liu et al., 2012, 2013; Yu et al., 2016; Chao et al., 2022). In Arabidopsis, genes in the flavonoid pathway, including those from the *TRANSPARENT TESTA* (TT) family (eg. *TT1-TT19*) and transcription factor complex MYB-bHLH-WD40, are key regulators controlling seed colour (Lepiniec et al., 2006; Xu et al., 2014, 2015). The biosynthetic pathways of lignin and pigments share the same precursors (Vanholme et al., 2012), indicating that the two complex biosynthetic pathways may act together to control both lignin and seed colour, and possibly other quality traits in *B. napus* seeds.

As Arabidopsis is a very close relative of *B. napus* within the Brassicaceae family, it is reasonable to infer that seed quality traits of *B. napus*, e.g., lignin, are also controlled by interactions of multiple metabolic and biological pathways. Therefore, a systems genetic approach that combines quantitative genetics, regulatory network analysis, and eQTL analysis would be more appropriate to investigate seed quality traits within *B. napus* (Kliebenstein, 2009 and Marshall-Colón and Kliebenstein, 2019; Kliebenstein, 2020). In the present study, we aimed to fully exploit the power of this

approach to decipher the regulatory network underlying lignin traits in *B. napus* seeds.

2 Materials and methods

2.1 Plant materials

The *B. napus* lines used in this study include the two parental lines DH12075, a Canadian spring-type doubled haploid canola line (generated by Dr. Gerhard Rakow and Dr. Ginette Séguin-Swartz, Agriculture and Agri-Food Canada), and PSA12, a resynthesised *B. napus* line (generated by Dr. Monica Beschorner and Dr. Derek Lydiate, Agriculture and Agri-Food Canada), and 96 doubled haploid segregating (SG) lines generated from the cross of these two parental lines. PSA12 was developed from an interspecific cross between *B. oleracea ssp. alboglabra* line A12DH and *B. rapa* line Parkland Sunshine, and DH12075 was derived from a cross between varieties Cresor and Westar (Mayerhofer et al., 2005).

The developing seeds at 21 days post-anthesis were collected from *B. napus* lines grown under field conditions at the AAFC Saskatoon Research Farm in 2009. This time point was selected because approximately 20 days after flowering (DAF) had previously been identified as a critical stage for cell proliferation and oil deposition (O'Hara et al., 2002; Dong et al., 2004). Total RNA was extracted from the developing seeds using a method modified from that of Oñate-Sánchez and Vicente-Carbajosa (2008) and described in detail by Parkin et al. (2010). Seeds from field grown plants from two trials (2009 and 2010) were harvested for seed quality analysis.

2.2 Seed quality analysis

Seed quality traits including oil content (OilHD), protein, lignin content of NDF, ADF, and ADL, iodine value (IV), hydrogenated density (HD) total glucosinolate, and chlorophyll were estimated by near-infrared reflectance (NIR Systems Model 6500, FOSS, Eden Prairie, MN) following methods described by Yu et al. (2016). Results are presented as percentage of whole seed dry matter with zero moisture. The system was calibrated for oil content using the method of Raney and Serblowski (2007), using samples hexane extracted following the method described by Raney et al. (1987). Fatty acid composition of the extracted oil samples was also analysed using the Agilent 6890 GC-FID (Agilent Technologies, Santa Clara, CA) (OilEx).

2.3 Array processing and data analysis

cRNAs were amplified from total RNA samples (2 μg) and labelled with either cy3 or cy5 using the Quick Amp labelling kit, Two Colour (Agilent, Catalogue: 5190-0444) according to the manufacturer's instructions, followed by purification with the Qiagen RNeasy Mini Kit (Qiagen, Catalogue 74104) and quantification with a NanoDrop ND-1000. Purified cRNAs (2 μg) were fragmented and hybridized to the Agilent 4x44K *Brassica* arrays (Parkin et al., 2010) for 17 h at 65°C with a rotation of 10 rpm according to the manufacturer's protocol. Arrays were washed with the Agilent Gene Expression Washing buffers according to the manufacturer's protocol and scanned with GenePix 4000B as described by Parkin et al. (2010). Four and two biological replications were performed for the parental lines and the SG lines, respectively, with dye-swaps applied to the biological replicates to minimize dye effects from the two colour systems (Lee et al., 2004).

The raw array datasets were extracted with Gene Pix Pro 6.0, and were firstly normalized by the lowess (Yang et al., 2002) method within arrays, followed by the quantile (Bolstad et al., 2003) method between arrays, using the R package Limma (Smyth, 2005). The normalized datasets of the SG lines were averaged across replications. Probes with intensities over 20 and with a frequency less than 30 percent from our datasets were treated as non-significant signal probes and removed, resulting in a final total of 29,752 quality probes. Finally, the normalized gene levels lower than 20 were adjusted to 20 as background intensity. The log2 transformed normalized data were treated as gene expression traits (e-traits) and were used to perform eQTL analysis.

2.4 Differential gene expression analysis of parental lines DH12075 and PSA12

Normalized and adjusted data of parental lines with four biological replications were imported to GeneSpring GX 10.0 (Agilent Technologies) and analysed differential gene expression with Student's t test. The p-value was adjusted with Benjamini-Hochberg's method and controlled at FDR = 0.05. Genes with adjusted p-value < 0.05 and fold changes equal or larger than 2.0 were claimed as significantly differentially expressed genes.

2.5 Genotyping and linkage map construction

Genomic DNA was extracted with a modified CTAB method and genotyped with simple sequence repeat (SSR) markers developed at the Saskatoon Research Centre, Agriculture and Agri-Food Canada (http://aafc-aac.usask.ca/BrassicaMAST/). The SSRs were analysed using fluorescently labelled tail PCR amplifications in 384-well format on an Applied Biosystems 3730xl. Genotype data was imaged using genographer (available at http://www.genographer.com/), and polymorphic loci were scored visually. The SNP genotyping data of the SG population were generated by the Brassica 60 K Infinium array from Clarke et al. (2016).

2.6 Linkage map construction and QTL mapping

Genetic maps were developed with Mapdisto (Lorieux, 2012) using 694 SSR markers and 1266 SNP (binned map) markers with a

cut-off recombination value of 0.3, and a threshold logarithm of odds (LOD) of 6.5. Physical positions of mapped markers were retrieved by BLASTN-searching SNP flanking sequences against the Darmor-bzh reference genome (Chalhoub et al., 2014).

QTL analysis was performed by composite interval mapping (CIM) as implemented in WinQTL Cartographer (Wang et al., 2007) with a walking speed of 1 cM and a window size of 10 cM. The significance of a QTL was claimed at the threshold of 5% significance level by the 3000-permutation test in WinQTL Cartographer (Churchill and Doerge, 1994; Basten et al., 2004; Wang et al., 2007). QTL within an interval of 10 cM were merged into one QTL, represented by the QTL peaks. QTL were visualized by the R package LinkageMapView (Ouellette et al., 2018).

2.7 Gene co-regulatory network construction

Normalized and filtered transcripts were used to construct a weighted gene co-regulatory network (WGCNA) with the step-by-step method implemented in the WGCNA R package (Zhang and Horvath, 2005; Langfelder and Horvath, 2008). The soft-thresholding power function analysis was applied correlation of adjacency of genes with the power value from 1 to 20 (Supplementary Figure S1). The optimized power value was 12, at which an approximal scale-free network was reached (Supplementary Figure S1). With this value, the adjacency matrix was transformed into a topological overlap matrix (TOM). Gene hierarchical clustering tree was generated by the dynamic tree-cutting method, hcluster. Modules with eigenvalue similarity >= 0.75 were merged, and modules with less than 30 genes were removed. Modules and their relationship to external traits were also obtained through the method within this package (Langfelder and Horvath, 2008). The *B. napus* seed network was generated with a hard cut-off value of |r|= 0.75 which made the network follow a roughly power-law distribution (Ravasz et al., 2002; Barabási and Oltvai, 2004). Identified gene interactions were imported into Cytoscape (Shannon et al., 2003) version 2.7.0 for visualization and additional analysis.

2.8 Module-trait association and module QTL (network QTL) analysis

The gene expression profile in each module was characterized by the eigenvalues of each module, and is referred to here as the module eigengene value (ME). The module-trait association was performed by Pearson correlation analysis between MEs and seed quality traits implemented within the WGCNA package. Eigengene value of the network can extract the key information (features) of the larger number of correlated genes within the network, and retain its most important variance of the network dataset, and reduce its dimensionality (Zhang and Horvath, 2005; Langfelder and Horvath, 2008). Thus, the ME that captures the summarized variance of the regulatory network can be treated as a classical phenotypic trait representing the correlated gene expression patterns of the whole network. Subsequently, QTL mapping of MEs (network QTL) was performed as described above (section 2.6).

2.9 Gene annotation, functional and enrichment analysis

ESTs of array probes were used in BLASTP (E-value <1e−5, coverage >70%, identity >70%) against the *Arabidopsis* Information Resource protein database (TAIR 10). The BLASTP output and the *Arabidopsis* functional annotations were parsed with in-house scripts to derive functional annotation and relevant GO terms.

Gene function terms were characterized with Mapman (Usadel et al., 2005) ontology terms and mapped into different biological pathways with tools embedded in the Mapman software (Usadel et al., 2005). Gene function enrichment analysis within detected modules was performed with the Fisher's exact test and significance was claimed at $p < 0.05$.

2.10 eQTL analysis

The log transformed gene expression values were treated as a trait and analysed with the composite interval mapping (CIM) (Zeng, 1994) module "zmapqtl", which was implemented in QTL Cartographer version 1.17 (Basten et al., 2004) under the Linux system, with a walking speed of 1 cM and a window size of 10 cM. Due to the limited capacity of the computation, a global permutation threshold (GPT) described by West et al., 2007, was performed to obtain a threshold criterion for declaring a statistically significant eQTLs. The GPT value ($p = 0.05$) equals a LOD score of 4.75 for this eQTL analysis. eQTLs within 10 cM intervals were merged into one QTL represented by the QTL peak.

The sequence of the array probes was also used with BLASTN (E-value < 1e-10) against the Darmor-bzh reference genome (Chalhoub et al., 2014) and physical positions of the best-hit were retrieved. The eQTL was defined as cis- if the gene and the marker linked to it co-localized within a physical interval of 4 Mb; otherwise, they were defined trans-eQTLs. According to this definition, cis- and trans-eQTL of *B. napus* were identified.

3 Results

3.1 Characterization of QTL for lignin content

The lignin contents on the seeds of this DH population exhibited a normal distribution, assessed using the Shapiro-Wilk test ($p < 0.05$) (Supplementary Figure S2). The parental line DH12075 consistently showed higher lignin contents than the resynthesized PSA12 in different environments (Supplementary Figure S2). We consistently identified major QTLs for lignin content NDF, ADF, and ADL from two-years of field trials, on chromosomes C08 and A09 (Figure 1), while a minor QTL from

C01 was also identified in 2010. The NDF, ADF, and ADL shared the same QTLs on A09 and C08 (Figure 1), and since it has been suggested that investigation of ADF or NDF alone was sufficient to dissect the quantitative traits of fibre and lignin content (Cardinal et al., 2003), we therefore decided to use ADF as a representative trait to explore the genetic architecture underlying seed lignin traits. We defined three QTLs on A09: qLignin_A09_1 (comprised of qNDF_A09_1, qADF_A09_1 and qADL_A09_1); qLignin_A09_2 (from qNDF_A09_2, qADF_A09_2 and qADL_A09_2), and qLignin_A09_3 (from qNDF_A09_3, qADF_A09_3 and qADL_A09_3). The QTL on C08 was defined as qLignin_C08 (consisting of qNDF_C08, qADF_C08 and qADL_C08). qLignin_A09_1 spanned an interval of 130-145 cM and was located at 19-25 Mb on A09. It conferred effects with R^2 of 7% in 2009 and 12% in 2010. The qLignin_A09_2 conferred the largest effects in the two-year test, and localized between 145-160 cM with a physical interval of 25-30 Mb. The qLignin_A09_3 localized between 160-170 cM with an interval of 30-33 Mb on A09. The qLignin_C08 was found in the interval of 82-92 cM with the physical interval of 30-40 Mb on C08. The QTL qLignin_A09_1, qLignin_A09_3, and qLignin_C08 conferred a similar magnitude of effect on lignin content and with ~3-5% less than the qLignin_A09_2. Synteny analysis identified qLignin_A09_2 and qLignin_C08 are homologous QTL from the A and C subgenomes, respectively (Supplementary Figure S3). The alleles for all QTL on A09 that confer the higher level of lignin content are from parental line DH12075, while the alleles for the higher lignin QTL on C08 are from the synthetic parental line, PSA12.

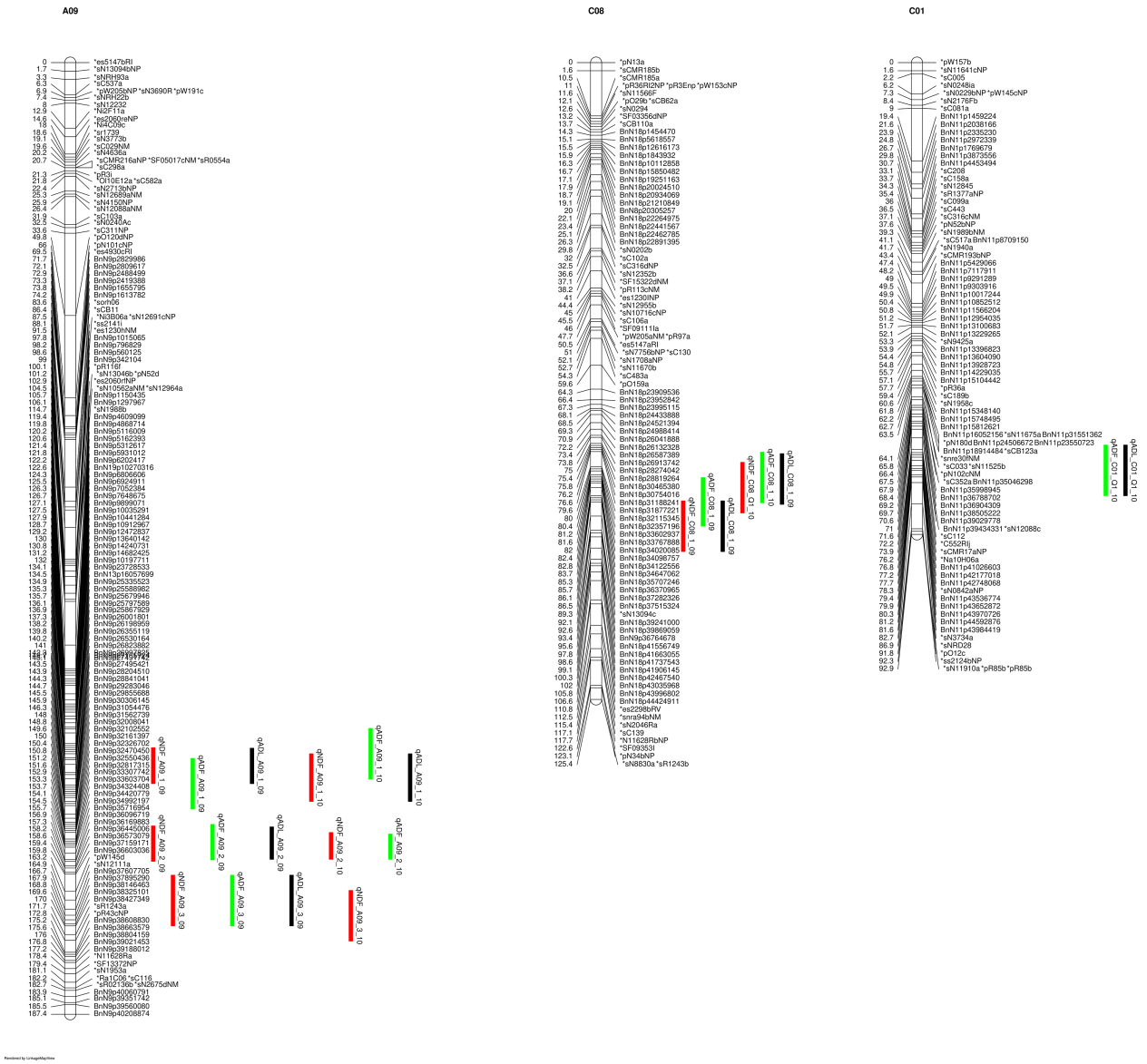


FIGURE 1
QTL for lignin content on identified on chromosomes A09 and C08 of the SG genetic map. NDF (neutral detergent fibre), indicated with red bars; ADF (acid detergent fibre), indicated by green bars, ADL (acid detergent lignin), indicated with black bars. 09 and 10 indicate the years 2009 and 2010.

10.3389/fpls.2024.1393621

3.2 Identification of genetic regulatory networks underlying lignin content

A total of 29,725 transcripts were used to build a co-regulatory network with WGCNA. The final co-expression network consisted of 12,670 nodes connected by 468,974 edges, containing ~60% of the information from the input datasets. The entire network was partitioned into 83 distinct gene modules (or subnetworks). A global view of the co-regulatory network with modules coloured distinctly on the network is shown in Figure 2. The number of genes within these subnetworks ranged from 30 to 7,303 genes.

Module-trait association analysis was performed to identify regulatory networks associated with seed quality traits in *B. napus* seeds; including oil content, protein, NDF, ADF, ADL, IV, glucosinolate, hydrogenated density, glucosinolate, and chlorophyll. Module 33 and module 35 were identified as significantly positively associated with NDF, ADF, and ADL, while module 67 was significantly negatively correlated with these three forms of lignin (Figure 3). As shown in Figure 3, the correlation coefficients between ADF and module 33, module 35, and module 67 were 0.6 (p-value: 3E-12), 0.6 (p-value: 9E-12), and -0.8 (p-value: 2e-22), respectively.

To validate that these modules are causal networks of lignin traits, a network QTL mapping approach was applied, utilizing linkage mapping to identify the genetic locus that controlled these modules. As shown in Figure 4, the QTL for module 33, qModule33_A09_1 overlapped with lignin pQTL qLignin_A09_1 on A09. The two QTLs for module 35, qModule35_A09_1 and qModule35_A09_2, colocalized with qLignin_A09_2 and qLignin_A09_3, respectively (Figure 4). The qModule67_A09_1 colocalized with qLignin_A09_2 and qModule_67_C08_1 collocated with qLignin_C08 (Figure 4). To further explore these regulatory networks, genes were extracted from these three modules and a sub-network built, which we termed the lignin causative network (Figure 5).

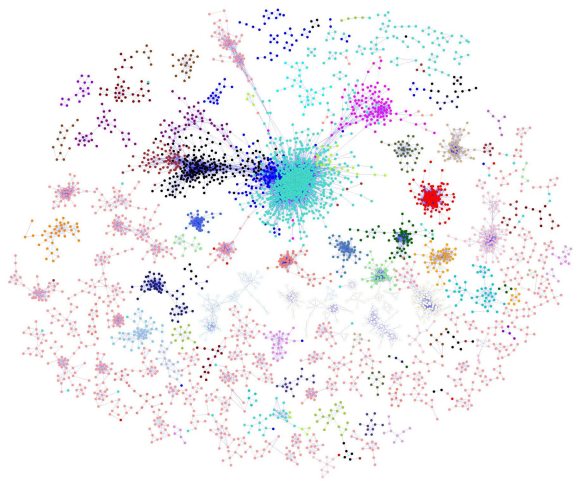


FIGURE 2
B. napus co-regulatory network. Nodes are probe sets from the customized *B. napus* array. The node colours depict the different modules identified from the network. Edges indicate significant co-expression between probe sets above a hard threshold. In order to improve the clarity of the display, only correlation coefficient (|r|>= 0.85) were shown on the network.

3.3 Gene function and enrichment analysis of the genetic regulatory network for lignin content

From the metabolic overview map, the major metabolite pathways that corresponded with the change of lignin content were characterized by genes involved in secondary metabolism, lipid biosynthesis, and cell wall modification (Supplementary Figure S4). Gene function enrichment analysis revealed that the lignin causal network was enriched with genes related to secondary metabolites (p< 1.8-E04) (Supplementary Table S1). These secondary metabolism-related genes were mapped mainly on the flavonoid, phenylpropanoids, lignin, shikimate, and mevalonate (MVA) pathways (Supplementary Figure S5). Major genes involved in lignin and flavonoid pathways, including *BANYULS (BAN)*, *CCR1*, *C4H*, *PAL2*, and the *TT*-related genes, *TT3*, *TT4*, *TT6*, *TT7*, *TT8*, *TT18*, *TT19*, were found positively correlated with lignin as shown in Supplementary Table S2, Supplementary Figure S6.

Stress-related genes were also enriched in the regulatory network (p-value: 0.01) (Supplementary Table S1); therefore, we explored the patterns of transcripts involved in abiotic and biotic stress. An interesting change in gene expression patterns associated with biotic stress in response to the change of seed lignin was observed. Almost all secondary metabolite-related genes involved in biotic stress were positively correlated with the change of lignin, while many of the major genes involved in biotic stress, including WRKY transcription factors (*WRKY25* and *WRKY33*), ethylene-related AP2 transcription factors, ethylene-related genes, and MAP kinases, were negatively correlated with lignin accumulation (Supplementary Table S2, Supplementary Figure S7).

If the genes are causal for the trait, it can be hypothesized the eQTL of the gene will colocalize with the trait QTL (pQTL). Prioritized causal genes will be shown as cis-eQTLs, in which the gene controls the phenotypic variation by mediating its expression via cis-regulatory sequence variation. eQTL analysis of the 910 genes within the lignin regulatory network found 163 cis-eQTL and 491 trans-eQTL underlying the three lignin QTL of the A09 chromosome (Supplementary Table S3). There were 14 cis- and 136 trans- eQTL identified for the C08 lignin QTL (Supplementary Table S4). The lignin gene *CCR1* and seed colour genes *TT4*, *TT6*, and *TT8* were identified as cis-eQTLs that colocalized with the lignin QTL on A09. The key lignin genes, *PAL2* and *C4H*, and seed colour genes *TT3*, *TT7*, *TT18*, *TT19*, and *BAN* were found to be trans-eQTLs (Supplementary Tables S3, S4), and none of the cis-eQTLs of lignin and flavonoid genes from C08 QTL were found above the significance threshold value. Gene expression analysis between the two parental lines found that of the genes known to be important to the lignin pathway, only *CCR1* was significantly differentially expressed between the parental lines (Supplementary Table S5). DH lines with the eQTL of *CCR1* (based on the DH12075

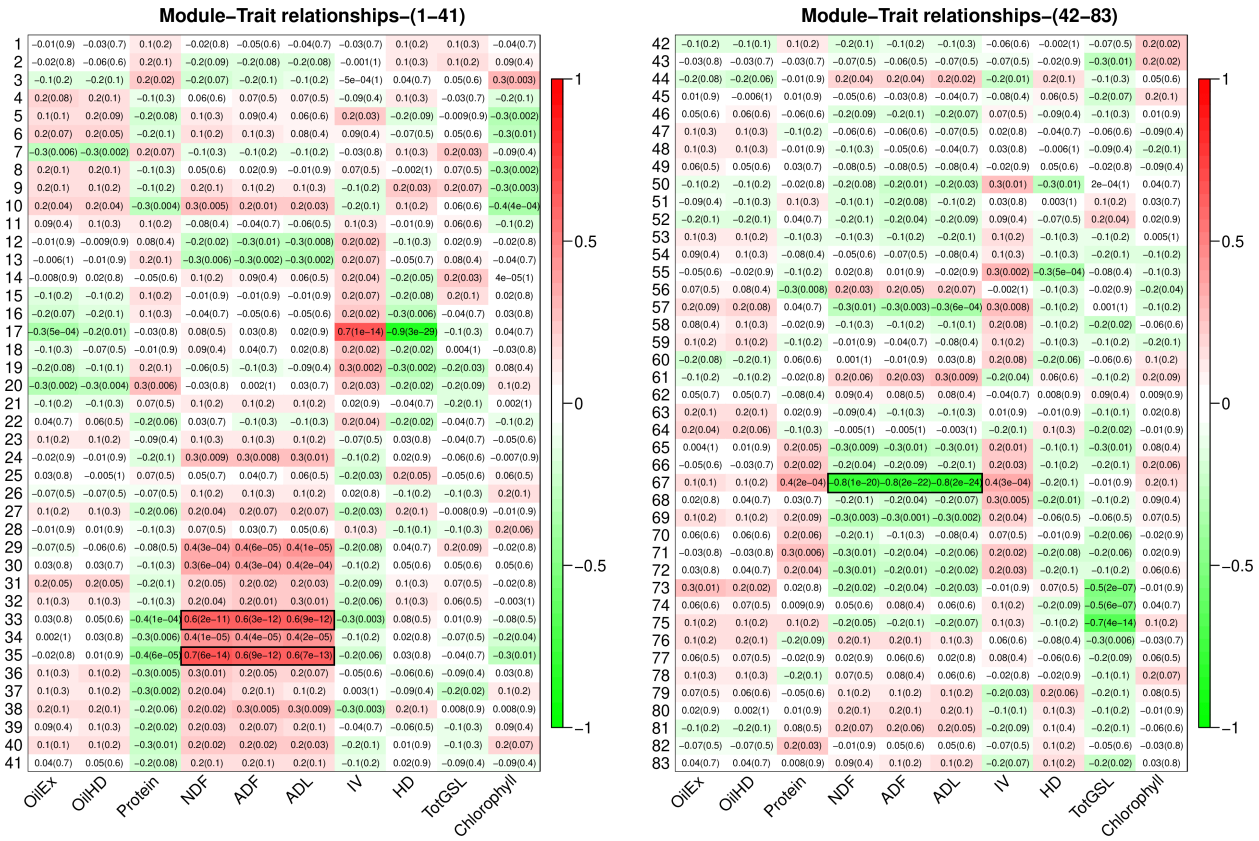


FIGURE 3
Module-trait relationship between 83 modules and seed quality traits. Traits and rows represent the modules. Correlation co-efficient and p-value (in parenthesis) were shown in each cell. Correlation between modules and seed quality traits was presented by colours ranging from red (high positive correlation) to green (high negative correlation). Evaluated seed quality traits include oil content (OilHD, measured by NIR; OilEx, determined by GC-FID), protein, lignin content of NDF (neutral detergent fibre), ADF (acid detergent fibre), ADL (acid detergent lignin), iodine value (IV), hydrogenated density (HD), total glucosinolate (TotGSL), and chlorophyll.

marker type at the eQTL peak BnN9p34992197) have significantly higher group means of gene expression than lines without the eQTL (Supplementary Figure S8). The seed colour genes *TT8*, *TT6* and *TT19* were statistically differential expressed, but with a fold change of gene expression level between 1.5 to 2.0 (Supplementary Table S5). Lignin content was significantly associated with the expression of these genes with cis-QTL, *CCR1*, *TT6*, and *TT8* (Supplementary Figure S9).

4 Discussion

Complex traits or quantitative traits are typically controlled by complex genetic architecture consisting of hundreds to thousands of genes (Kliebenstein, 2009). These genes form a genetic regulatory network that controls the trait, which challenges the single gene or traditional QTL analysis. Therefore, we utilized a systems genetic approach that integrates the quantitative genetic, gene co-regulatory network, network-QTL, and eQTL analysis to uncover the causal regulatory network underlying lignin content in *B. napus* seeds.

Three major QTLs were identified on A09 spanning a physical interval of approximately 13 Mb. All three QTLs conferred stable and relatively large effects on lignin content. QTLs for NDF, ADF, and ADL colocalized at the same genetic locus, indicating these three components of lignin shared the same genetic architecture. Numerous QTLs for seed quality traits including seed colour, oil, protein, and fibre were previously identified on the A09 chromosome (Wittkop et al., 2009; Liu et al., 2012, 2013; Yu et al., 2016; Behnke et al., 2018; Miao et al., 2019; Chao et al., 2022; Yusuf and Möllers, 2024), indicating that A09 is a hotspot that harbours several genes for these seed quality traits. However, this study and all previous research have used a relatively small population size, from ~100 to 300 lines, making it challenging to precisely identify QTLs within this hotspot. Future, larger and advanced-structured genetic populations could be used to dissect this hotspot. In addition, a stable and large effect homologous QTL to the qLignin_A09_2 was identified on C08, indicating both the A and C genome contributed to the variation of lignin content in our population. Currently, limited studies, including Badani et al. (2006) and Chao et al. (2022), have reported the QTL on C08 for seed colour and lignin content. The relative simplicity of the C08 QTL, compared to A09, makes it a better target to uncover genes underlying lignin QTL in the future. Furthermore, since the C08 QTL is homologous to the A09 lignin QTL, uncovering its genetic architecture may also shed light on the A09 hotspot.

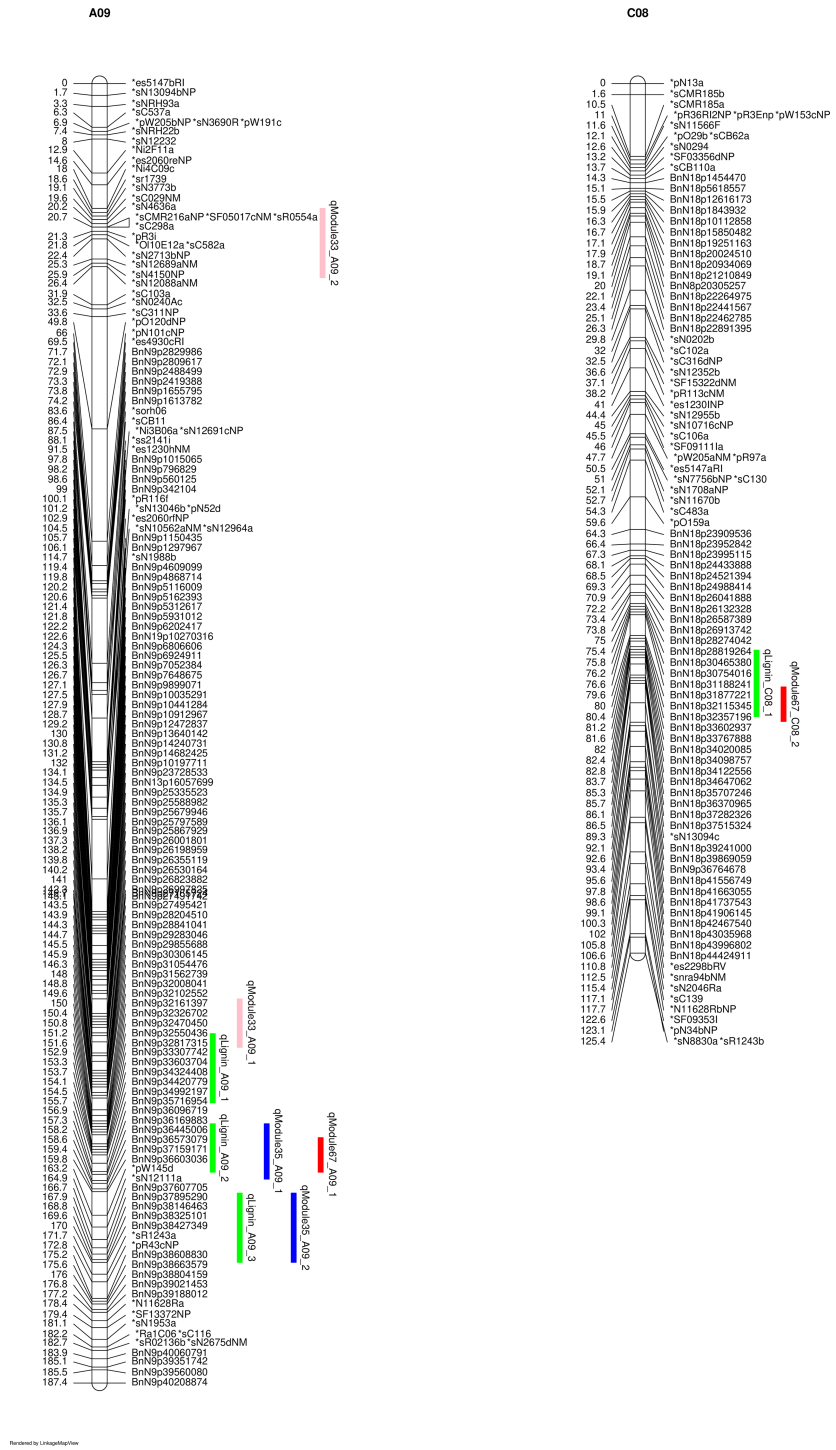


FIGURE 4
Linked lignin phenotype QTLs (pQTLs) to module QTLs (mQTLs). qLignin are lignin QTL, and qModule are module QTL.

The genome-wide transcriptome datasets for developed seeds of the SG population, allowed a seed development biological network to be successfully built (Figure 2). Biological network construction assumes that it will follow the topological scale-free property, in which the distribution of the nodes and edges follows a power-law distribution (Barabási and Oltvai, 2004). Although definitions of the properties of biological networks is still under discussion, the seed developmental regulatory network did exhibit the expected scale-free topology, described as a small number of nodes with a large number of connections, while most nodes have few connections (Figure 2). The distribution of the nodes and edges in our network exhibited a power-law distribution, supporting its topological free property, and was in agreement with the network property discovered in Arabidopsis (Atias et al., 2009; Mao et al., 2009),

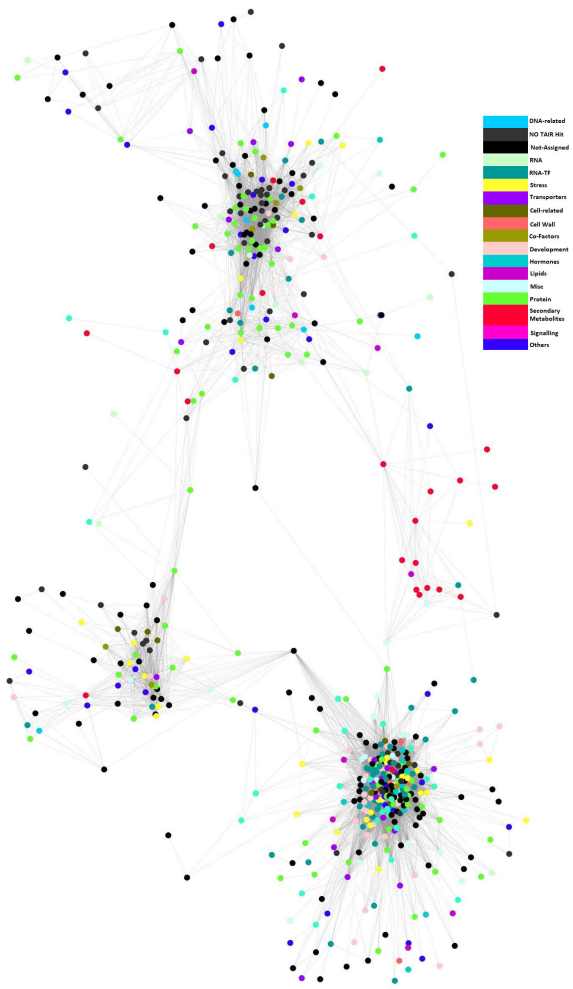


FIGURE 5
Co-regulatory network associated with lignin. Nodes are transcripts identified from modules 33, 35 and 67. Edges indicate significant co-expression between probe sets above a hard threshold. Node colours depict different Mapman ontology terms (legend right).

rice (Ficklin et al., 2010; Ficklin and Feltus, 2011) and maize (Ficklin and Feltus, 2011).

The structure of genome-wide co-expression networks are hierarchical with modular units (Aoki et al., 2007). Given that genes with similar biological functions are typically co-expressed together and formed functional modules, disassembling biological networks into functional modules or clusters is more informative and practical for biologists to interpret the network (Mutwil et al., 2010). Here, the seed network was divided into 83 computational biological modules. If phenotype variation is attributed to the variation of genes within the module, a correlation should be observed between modules and phenotypes. This was the case for two modules (modules 33 and 35) that were significantly positively correlated with lignin, and one (module 67) that was negatively coordinated with lignin. With network QTL analysis, it was found that lignin QTLs (pQTLs) and module QTLs shared the same genetic loci on A09, supporting the suggestion that lignin variation was determined by the interactions of these three functional modules. These findings also support the hypothesis that entire plant networks can be inherited as a phenotypic trait (Marshall-Colón et al., 2019). Large-scale evaluation of 1,671 genes on agronomic performance in maize using a single gene engineering approach found that only one gene encoding a transcription factor, *zmm28*, showed promise, indicating the marginal effect of a single gene for the improvement of complex traits (Simmons et al., 2021). The defined lignin network could enable breeders to target a specific regulatory network and potentially use an engineering approach to rewire and fine-tune the network to improve crop performance.

Since modules 33, 35, and 67 were altogether associated with seed lignin, these three modules were assembled into one large lignin-related functional module, called "lignin causative network", to perform further analysis. This network was enriched with secondary metabolism-related genes. One group of these genes includes *PAL2*, *C4H*, and *CCR1*, encoding three key enzymes controlling critical steps of the lignin pathway (Vogt, 2010). The other group, which mainly consists of flavonoid pathway genes, is involved in the accumulation of flavonoids in the seed coat, such as *TT3*, *TT7*, *TT6*, *TT8*, *TT18*, *TT19*, and *BAN*. All these genes are significantly and positively associated with seed lignin variation. Enrichment of these secondary metabolite genes indicated the high reliability of this causative regulatory network underlying lignin content.

eQTL analysis was performed on genes within this network to identify the causal genes underlying the network for lignin content. The lignin gene, *CCR1*, was detected as a cis-eQTL, while both *PAL* and *C4H* were found as trans-eQTLs. Flavonoid genes including *TT4*, *TT6*, and *TT8* were identified as cis-eQTLs, while the remaining *transparent testa* genes were trans-eQTLs. In addition, all these cis-eQTLs colocalize with the fibre QTL from chromosome A09, indicating they are prioritized as causal genes underlying this network (Kliebenstein, 2009). Liu et al. (2012) pinpointed *CCR1* from chromosome A09 as the candidate gene for ADF in *B. napus* seeds. Zhang et al. (2022) identified a CCR-like gene, *CCRL*, from C07 as one candidate gene for seed content in *B. napus*. *CCR1* has been found to encode the key enzyme for lignin accumulation in stems (Jones et al., 2001). The cis-controlled eQTL impacting the variation of *CCR1*, the co-localization with phenotype QTL, and the unique known gene from the monolignol pathway located on our lignin causative network as well as previous findings, indicated that *CCR1* is one of the causative genes underlying the seed lignin trait. Of the three cis-regulated flavonoid genes, *TT8* was found to be a key regulator for seed oil and colour in Arabidopsis and *B. napus* (Chen et al., 2014; Xu et al., 2014, 2015; Zhai et al., 2020; Zhang et al., 2022) and lignin in *B. napus* (Zhang et al., 2022). In addition, Zhai et al. (2020) found *TT8*, as a BHLH transcription factor, regulated both the flavonoid and lignin pathways. All these findings pinpoint *TT8* as another key regulator in this network. Li et al. (2023) reported that *TT4*, which encodes an enzyme acting at an earlier step of the flavonoid pathway and located on chromosomes A03 and C02, is

key regulator for oil and protein content in *B. napus* seeds, while we identified *TT4* from A09 as playing a key role for lignin content. Therefore, it is hypothesized that multi-homologous and homoeologous *TT4* genes in different genotypes control seed quality traits. In addition, another earlier flavonoid biosynthetic gene *TT6* was identified as a potential new key regulator for lignin content in *B. napus*. Thus, the lignin traits associated with the A09 loci are controlled by a complex regulatory network with multiple genes.

Notably, we did not find any cis-eQTL from the key lignin and flavonoid genes associated with the homologous C08 lignin QTL. Network QTL analysis only found that the QTL of module 67 colocalized with the qLignin_C08 QTL. Genes for the development of *B. napus* seed are spatially and temporally regulated (Zhai et al., 2020). It is possible that the A and C genome coordinately regulate seed lignin content in a temporal manner, while the regulatory network built here, was only from one key time point, 21 days after flowering. We observed, under greenhouse conditions, that the A donor line of PSA12 flowered and matured earliest, followed by DH12075 and PSA12, while the C genome donor line of PSA12 flowered and matured very late (unpublished data), supporting this hypothesis, since positive alleles at the C08 locus were contributed by PSA12. A further spatial-temporal regulatory network built on seed development will be more powerful for dissecting the genetic architecture controlling seed quality traits and should be considered for future studies.

Another interesting finding of the lignin network was that it was enriched with biotic stress genes, including key transcript factors, *WRKY25*, *33* and *SIGMA FACTOR BINDING PROTEIN 1* (*SIB1*), which are involved in triggering resistance to the necrotrophic pathogen *Botrytis cinerea* in Arabidopsis (Andreasson et al., 2005; Zheng et al., 2006; Eulgem and Somssich, 2007; Pandey and Somssich, 2009; Lai et al., 2011). However, all these genes were significantly negatively associated with lignin accumulation, which is contradictory to the positive role of lignin on enhancing the resistance response (Vanholme et al., 2010). This paradoxical role of lignin in the plant defence mechanism compelled us to hypothesize that cell wall integrity plays a key role in this process. Reduced lignin affects the integrity of the wall which mimics disease infection, leading to the release of defence-related elicitors. When a normal plant is infected by disease, cell wall integrity is damaged and in response to this, the cell wall induces lignin biosynthesis to maintain its integrity. This theory is further supported by studies which have shown that cell wall damage directly triggers lignin biosynthesis (Denness et al., 2011).

It is worth noting a few limitations of the current study. First, the population size was small, and may not provide enough statistical power for identifying minor QTL. Second due to the limitation of array technology, many unknown genes within our causative network may be missing from our research. Third, this regulatory network was built only from one stage of seed development, and could not capture a complete temporal picture of development. New advanced structured genetic populations, and genome sequencing tools including long read and pangenome sequence analysis will empower these systems genetic approaches to uncover the genetic mechanism of complex traits.

5 Conclusion

An example of the power of systems genetics that combines quantitative genetic, regulatory network analysis, network QTL, and eQTL analysis to identify causal regulatory networks and causal genes for lignin traits was provided. Further insights were revealed into the complex genetic architecture that regulates lignin content, specifically that controlled by the locus on A09 of *B. napus*. The analyses provide more prioritized genes to target to improve seed quality trait in *B. napus*. Moreover, the research here may also provide a new avenue for crop communities to use regulatory networks to select or design crops with better performance and quality in a changing climate.

Data availability statement

The data presented in the study are deposited in the NCBI data repository, accession number GSE268725.

Author contributions

WZ: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. EH: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – review & editing. SR: Data curation, Methodology, Writing – review & editing. WC: Data curation, Methodology, Writing – review & editing. KB: Formal analysis, Visualization, Writing – review & editing. AS: Conceptualization, Funding acquisition, Writing – review & editing. PF: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing. IP: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the ERA-PG project "Associative expression and systems analysis of complex traits in oilseed rape/canola (ASSYST)" and the Sustainable Protein Production (SPP) program from the National Research Council of Canada (NRC).

Acknowledgments

We thank Lily Tang and Lanette Ehman for field work assistance, Matthew Links and Steve Karcz for bioinformatics assistance, and Daoquan Xiang for critically reviewing the manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1393621/full#supplementary-material

SUPPLEMENTARY TABLE 1
Gene Ontology (GO) Enrichment analysis of the fibre related regulatory network.

SUPPLEMENTARY TABLE 2
Gene function, module, and its correlation with acid detergent fibre (ADF). Significant correlations between gene expression levels and ADF were claimed at p-value < 0.05.

SUPPLEMENTARY TABLE 3
eQTLs overlapping with pQTLs identified from the fibre related network on A09. cis and trans, eQTL defined by B. napus cv. Darmor-bzh reference genome. cis*, cis-eQTL defined by the DH12075 reference genome.

SUPPLEMENTARY TABLE 4
eQTLs overlapping with pQTLs identified from the fibre related network on C08. cis and trans, eQTL defined by B. napus cv. Darmor-bzh reference genome. cis*, cis-eQTL defined by the DH12075 reference genome.

SUPPLEMENTARY TABLE 5
Differential expressed genes within the fibre regulatory network between parental lines DH12075 (DH) and PSA12 (PSA). Up indicated higher gene expression level in DH12075. Corrected p-value: p-value was adjusted with Benjamini-Hochberg's method and controlled at FDR (False Discovery Rate) = 0.05. Genes with more than 2- fold changes and corrected p-value < 0.05 were claimed as significantly differential expressed. Genes highlighted in red have a corrected p-value < 0.05, but a fold change between 2 - 1.5.

SUPPLEMENTARY FIGURE 1
Soft-thresholding powers test to identify the optimized value to fit the approximate scale free topology of the network. The left panel displayed the scale-free topology index (y-axis) with the tested soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) with tested soft-thresholding power (x-axis).

SUPPLEMENTARY FIGURE 2
Distribution of lignin content in the years 2009 **(A-C)** and 2010 **(D-F)**. **(A, D)** acid detergent lignin (ADL); **(B, E)** neutral detergent fiber (NDF); **(C, F)** acid detergent fibre (ADF). Lignin contents of the parental lines PSA12 and DH12075 are highlighted with blue and red dashed lines, respectively. The normality of the distribution was assessed using the Shapiro-Wilk test, which yielded p values of **(A)** p = 0.001, **(B)** p = 0.032, **(C)** p = 0.003, **(D)** p = 0.001, **(E)** p = 0.0099, **(F)** p = 0.001, which all exceed the significance of p < 0.05.

SUPPLEMENTARY FIGURE 3
Synteny analysis of the lignin QTL qLignin_A09_2 and qLignin_C08. Synteny was analyzed and visualized with the Brassica napus Genome Browser (https://www.genoscope.cns.fr/brassicanapus/) with the Darmor-bzh reference genome (Chalhoub et al., 2014).

SUPPLEMENTARY FIGURE 4
Mapped transcripts from fibre network to the major metabolites map. Correlation between transcripts and seed ADF (acid detergent fibre) trait was presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 5
Dissection of transcripts on the major secondary metabolite pathways. Correlation between transcripts and seed ADF (acid detergent fibre) trait was presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 6
Mapped known genes to the lignin and flavonoid pathway. Correlation between transcripts and seed ADF (acid detergent fibre) were presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 7
Identification biotic stress signal pathway coordinated with the changes of lignin. Correlation between transcripts and seed (acid detergent fibre) trait was presented by colors ranging from red (high positive correlation) to blue (high negative correlation).

SUPPLEMENTARY FIGURE 8
Analysis of gene expression of double haploid (DH) lines with cis-eQTL of CCR1. The DH lines were separated into two groups based on the genotype of the eQTL peak marker *BnN9p34992197*: lines with the eQTL (DH12075 marker type) labeled in blue, and lines without the eQTL (PSA12 marker type) labeled with yellow. Student's t-test was used to test the significance of means between groups at the significance level of p < 0.05.

SUPPLEMENTARY FIGURE 9
Pearson correlation analysis of trait-gene expression. The log2 transformed gene expression levels of **(A)**. *cinnamoyl-CoA-reductase (CCR1)*; **(B)**. *TRANSPARENT TESTA 6 (TT6)*; and **(C)**. *TRANSPARENT TESTA 8 (TT8)* are shown in the Y-axis, with acid detergent fibre (ADF) in the X-axis. Parental line DH12075 is indicated in green, PSA12 is indicated in red, and Double haploid (DH) lines are indicated by black dots. R, Pearson Correlation Coefficient; and p, p-value, significance level was claimed at p < 0.05.

References

Andreasson, E., Jenkins, T., Brodersen, P., Thorgrimsen, S., Petersen, N. H., Zhu, S., et al. (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. *EMBO J.* 24, 2579–2589.

Aoki, K., Ogata, Y., and Shibata, D. (2007). Approaches for extracting practical information from gene coexpression networks in plant biology. *Plant Cell Physiol.* 48, 381–390. doi: 10.1093/pcp/pcm013

Atias, O., Chor, B., and Chamovitz, D. A. (2009). Large-scale analysis of Arabidopsis transcription reveals a basal co-regulation network. *BMC Syst. Biol.* 3, 86. doi: 10.1186/1752-0509-3-86

Badani, A. G., Snowdon, R. J., Wittkop, B., Lipsa, F. D., Baetzel, R., Horn, R., et al. (2006). Colocalization of a partially dominant gene for yellow seed colour with a major QTL influencing acid detergent fibre (ADF) content in different crosses of oilseed rape (*Brassica napus*). *Genome* 49, 1499–1509. doi: 10.1139/g06-091

Barabási, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Bassel, G. W., Lan, H., Glaab, E., Gibbs, D. J., Gerjets, T., Krasnogor, N., et al. (2011). Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9709–9714. doi: 10.1073/pnas.1100958108

Basten, C. J., Weir, B. S., and Zeng, Z.-B. (2004). *QTL Cartographer Version 1.17* (Raleigh, NC: North Carolina State University).

Behnke, N., Suprianto, E., and Möllers, C. (2018). A major QTL on chromosome C05 significantly reduces acid detergent lignin (ADL) content and increases seed oil and protein content in oilseed rape (*Brassica napus* L.). *Theor. Appl. Genet.* 131, 2477–2492. doi: 10.1007/s00122-018-3167-6

Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high-density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185

Brem, R. B., and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1572–1577. doi: 10.1073/pnas.0408709102

Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755. doi: 10.1126/science.1069516

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., et al. (2007). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.* 39, 1329–1337. doi: 10.1038/ng.2007.17

Cardinal, A. J., Lee, M., and Moore, K. J. (2003). Genetic mapping and analysis of quantitative trait loci affecting fiber and lignin content in maize. *Theor. Appl. Genet.* 106, 866–874. doi: 10.1007/s00122-002-1136-5

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345, 950–953. doi: 10.1126/science.1253435

Chao, H., Guo, L., Zhao, W., Li, H., and Li, M. (2022). A major yellow-seed QTL on chromosome A09 significantly increases the oil content and reduces the fiber content of seed in *Brassica napus*. *Theoret. Appl. Genet.* 135, 1293–1305. doi: 10.1007/s00122-022-04031-0

Chen, M., Xuan, L., Wang, Z., Zhou, L., Li, Z., Du, X., et al. (2014). TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several seed development regulators in *Arabidopsis*. *Plant Physiol.* 165, 905–916. doi: 10.1104/pp.114.235507

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1093/genetics/138.3.963

Clarke, W. E., Higgins, E. E., Plieske, J., Wieseke, R., Sidebottom, C., Khedikar, Y., et al. (2016). A high-density SNP genotyping array for *Brassica napus* and its ancestral diploid species based on optimised selection of single-locus markers in the allotetraploid genome. *Theor. Appl. Genet.* 129, 1887–1899. doi: 10.1007/s00122-016-2746-7

Denness, L., McKenna, J. F., Segonzac, C., Wormit, A., Madhou, P., Bennett, M., et al. (2011). Cell wall damage-induced lignin biosynthesis is regulated by a ROS- and jasmonic acid dependent process in *Arabidopsis thaliana*. *Plant Physiol.* 156, 1364–1374. doi: 10.1104/pp.111.175737

Dong, J., Keller, W. A., Yan, W., and Georges, F. (2004). Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. *Planta* 218, 483–491. doi: 10.1007/s00425-003-1124-2

Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defence signaling. *Curr. Opin. Plant Biol.* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020

Ficklin, S. P., and Feltus, F. A. (2011). Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. *Plant Physiol.* 156, 1244–1256. doi: 10.1104/pp.111.173047

Ficklin, S. P., Luo, F., and Feltus, F. A. (2010). The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. *Plant Physiol.* 154, 13–24. doi: 10.1104/pp.110.159459

Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., et al. (2007). Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Nat. Acad. Sci. U.S.A.* 104, 6478–6483. doi: 10.1073/pnas.0611629104

Jansen, R. C., and Nap, J. P. (2001). Genetical genomics: the added value from segregation. *Trends Genet.* 17, 388–391. doi: 10.1016/S0168-9525(01)02310-1

Jones, L., Ennos, A. R., and Turner, S. R. (2001). Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. *Plant J.* 26, 205–216. doi: 10.1046/j.1365-313x.2001.01021.x

Keurentjes, J. J., Fu, J., Terpstra, I. R., Garcia, J. M., Van Den Ackerveken, G., Snoek, L. B., et al. (2007). Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1708–1713. doi: 10.1073/pnas.0610429104

Kliebenstein, D. J. (2009). Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annu. Rev. Plant Biol.* 60, 93–114. doi: 10.1146/annurev.arplant.043008.092114

Kliebenstein, D. J. (2020). Using networks to identify and interpret natural variation. *Curr. Opin. Plant Biol.* 54, 122–126. doi: 10.1016/j.pbi.2020.04.005

Lai, Z., Li, Y., Wang, F., Cheng, Y., Fan, B., Yu, J. Q., et al. (2011). Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *Plant Cell* 23, 3824–3841. doi: 10.1105/tpc.111.090571

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Lee, H. S., Wang, J., Tian, L., Jiang, H., Black, M. A., Madlung, A., et al. (2004). Sensitivity of 70-mer oligonucleotides and cDNAs for microarray analysis of gene expression in Arabidopsis and its related species. *Plant Biotechnol. J.* 2, 45–57. doi: 10.1046/j.1467-7652.2003.00048.x

Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Li, L., Tian, Z., Chen, J., Tan, Z., Zhang, Y., Zhao, H., et al. (2023). Characterization of novel loci controlling seed oil content in *Brassica napus* by marker metabolite-based multi-omics analysis. *Genome Biol.* 24, 141. doi: 10.1186/s13059-023-02984-z

Liu, L., Qu, C., Wittkop, B., Yi, B., Xiao, Y., He, Y., et al. (2013). A high-density SNP map for accurate mapping of seed fibre QTL in *Brassica napus* L. *PloS One* 8, e83052. doi: 10.1371/journal.pone.0083052

Liu, L., Stein, A., Wittkop, B., Sarvari, P., Li, J., Yan, X., et al. (2012). A knockout mutation in the lignin biosynthesis gene CCR1 explains a major QTL for acid detergent lignin content in *Brassica napus* seeds. *Theor. Appl. Genet.* 124, 1573–1586. doi: 10.1007/s00122-012-1811-0

Lorieux, M. (2012). MapDisto: fast and efficient computation of genetic linkage maps. *Mol. Breed.* 30, 1231–1235. doi: 10.1007/s11032-012-9706-y

Mackay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612

Mao, L., VanHemert, J. L., Dash, S., and Dickerson, J. A. (2009). Arabidopsis gene coexpression network and its functional modules. *BMC Bioinf.* 10, 346. doi: 10.1186/1471-2105-10-346

Marshall-Colón, A., and Kliebenstein, D. J. (2019). Plant networks as traits and hypotheses: moving beyond description. *Trends Plant Sci.* 24, 840–852. doi: 10.1016/j.tplants.2019.06.003

Mayerhofer, R., Wilde, K., Mayerhofer, M., Lydiate, D., Bansal, V. K., Good, A. G., et al. (2005). Complexities of chromosome landing in a highly duplicated genome: toward map-based cloning of a gene controlling blackleg resistance in *Brassica napus*. *Genetics* 171, 1977–1988. doi: 10.1534/genetics.105.049098

Miao, L., Chao, H., Chen, L., Wang, H., Zhao, W., Li, B., et al. (2019). Stable and novel QTL identification and new insights into the genetic networks affecting seed fiber traits in *Brassica napus*. *Theor. Appl. Genet.* 132, 1761–1775. doi: 10.1007/s00122-019-03313-4

Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöh, O., and Persson, S. (2010). Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. *Plant Physiol.* 152, 29–43. doi: 10.1104/pp.109.145318

O'Hara, P., Slabas, A. R., and Fawcett, T. (2002). Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. *Plant Physiol.* 129, 310–320. doi: 10.1104/pp.010956

Oñate-Sánchez, L., and Vicente-Carbajosa, J. (2008). DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. *BMC Res. Notes* 1, 93. doi: 10.1186/1756-0500-1-93

Ouellette, L. A., Reid, R. W., Blanchard, S. G., and Brouwer, C. R. (2018). LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics* 34, 306–307. doi: 10.1093/bioinformatics/btx576

Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990

Parkin, I. A. P., Clarke, W. E., Sidebottom, C., Zhang, W., Robinson, S. J., Links, M. G., et al. (2010). Towards unambiguous transcript mapping in the allotetraploid *Brassica napus*. *Genome* 53, 929–938. doi: 10.1139/G10-053

Persson, S., Wei, H., Milne, J., Page, G. P., and Somerville, C. R. (2005). Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8633–8638. doi: 10.1073/pnas.0503392102

Potokina, E., Druka, A., Luo, Z., Wise, R., Waugh, R., and Kearsey, M. J. (2007). Gene expression quantitative trait locus analysis of 16 000 barley genes reveals a complex pattern of genome-wide transcriptional regulation. *Plant J.* 53, 90–101. doi: 10.1111/j.1365-313X.2007.03315.x

Raney, J. P., Love, H. K., Rakow, G., and Downey, R. K. (1987). An apparatus for rapid preparation of oil and oil-free meal from Brassica seed. *Lipid Fett.* 89, 235–237. doi: 10.1002/lipi.19870890608

Raney, J. P., and Serblowski, G. (2007). "Inserts for FOSS NIRS 6500 spinning ring cups," in *Proceedings of the 12th International Rapeseed Congress, Wuhan, China*, 26–30 Mar. 2007, Vol. 5. 39–42 (Monmouth Junction, NJ: Science Press USA Inc.).

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Baraba´si, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555. doi: 10.1126/science.1073374

Rockman, M. V., and Kruglyak, L. (2006). Genetics of global gene expression. *Nat. Rev. Genet.* 7, 862–872. doi: 10.1038/nrg1964

Ruuska, S. A., Girke, T., Benning, C., and Ohlrogge, J. B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* 14, 1191–1206. doi: 10.1105/tpc.000877

Schadt, E. E., Monks, S. A., Drake, T. A., Lusis, A. J., Che, N., Colinayo, V., et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302. doi: 10.1038/nature01434

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Simmons, C. R., Lafitte, H. R., Reimann, K. S., Brugie`re, N., Roesler, K., Albertsen, M. C., et al. (2021). Successes and insights of an industry biotech program to enhance maize agronomic traits. *Plant Sci.* 307, 110899. doi: 10.1016/j.plantsci.2021.110899

Smyth, G. K. (2005). "Limma: Linear models for microarray data," in *Bioinformatics and computational biology solutions Using R and Bioconductor*. Eds. R. Gentlemen, V. Carey, S. Dudoit, R. Irizarry and W. Huber (Springer, New York), 397–420.

Snowdon, R. J., Wittkop, B., Rezaidad, A., Hasan, M., Lipsa, F. D., Stein, A., et al. (2010). Regional association analysis delineates a sequenced chromosome region influencing antinutritive seed meal compounds in oilseed rape. *Genome* 53, 917–928. doi: 10.1139/G10-052

Stein, A., Wittkop, B., Liu, L., Obermeier, C., Friedt, W., and Snowdon, R. J. (2013). Dissection of a major QTL for seed colour and fibre content in *Brassica napus* reveals colocalization with candidate genes for phenylpropanoid biosynthesis and flavonoid deposition. *Plant Breed.* 132, 382–389. doi: 10.1111/pbr.12073

Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O. E., Palacios-Rojas, N., et al. (2005). Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. *Plant Physiol.* 138, 1195–1204. doi: 10.1104/pp.105.060459

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

Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J. H., Goeminne, G., et al. (2012). A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *Plant Cell* 24, 3506–3529. doi: 10.1105/tpc.112.102574

Van Soest, P. J., Robertson, J. B., and Lewis, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *J. Dairy. Sci.* 74, 3583–3597. doi: 10.3168/jds.S0022-0302(91)78551-2

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/mp/ssp106

Wang, W., Basten, C. J., and Zeng, Z. B. (2007). *Windows QTL Cartographer 2.5* (Raleigh, NC: Department of Statistics, NCSU).

West, M. A., Kim, K., Kliebenstein, D. J., Van Leeuwen, H., Michelmore, R. W., Doerge, R. W., et al. (2007). Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. *Genetics* 175, 1441–1450. doi: 10.1534/genetics.106.064972

Wittkop, B., Snowdon, R. J., and Friedt, W. (2009). Status and perspectives of breeding for enhanced yield and quality of oilseed crops for Europe. *Euphytica* 170, 131–140. doi: 10.1007/s10681-009-9940-5

Wittkop, B., Snowdon, R. J., and Friedt, W. (2012). New NIRS calibrations for fibre fractions reveal broad genetic in *Brassica napus* seed quality. *J. Agric. Food Chem.* 60, 224–2256. doi: 10.1021/jf204936f

Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20, 176–185. doi: 10.1016/j.tplants.2014.12.001

Xu, W., Grain, D., Bobet, S., Le Gourrierec, J., Thevenin, J., Kelemen, Z., et al. (2014). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. *New Phytol.* 202, 132–144. doi: 10.1111/nph.12620

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., et al. (2002). Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 30, e15. doi: 10.1093/nar/30.4.e15

Yu, B., Boyle, K., Zhang, W., Robinson, S. J., Higgins, E., Ehman, L., et al. (2016). Multi-trait and multi-environment QTL analysis reveals the impact of seed colour on seed composition traits in *Brassica napus*. *Mol. Breed.* 36, 111. doi: 10.1007/s11032-016-0521-8

Yusuf, A. O., and Möllers, C. (2024). Inheritance of cellulose, hemicellulose and lignin content in relation to seed oil and protein content in oilseed rape. *Euphytica* 220, 5. doi: 10.1007/s10681-023-03264-4

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhai, Y., Yu, K., Cai, S., Hu, L., Amoo, O., Xu, L., et al. (2020). Targeted mutagenesis of bnTT8 homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. *Plant Biotechnol. J.* 18, 1153–1168. doi: 10.1111/pbi.13281

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, 17. doi: 10.2202/1544-6115.1128

Zhang, Y., Zhang, H., Zhao, H., Xia, Y., Zheng, X., Fan, R., et al. (2022). Multi-omics analysis dissects the genetic architecture of seed coat content in *Brassica napus*. *Genome Biol.* 23, 86. doi: 10.1186/s13059-022-02647-5

Zheng, Z., Qamar, S. A., Chen, Z., and Mengiste, T. (2006). Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant J.* 48, 592–605. doi: 10.1111/j.1365-313X.2006.02901.x

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 18 July 2024
DOI 10.3389/fpls.2024.1393140

Check for updates

OPEN ACCESS

EDITED BY
Zhaorong Hu,
China Agricultural University, China

REVIEWED BY
Roberto Tuberosa,
University of Bologna, Italy
Tian Li,
Chinese Academy of Agricultural Sciences, China
Huijie Zhai,
Henan Institute of Science and Technology, China

*CORRESPONDENCE
Lingling Jin
lingling.jin@cs.usask.ca
Andrew G. Sharpe
andrew.sharpe@gifs.ca

RECEIVED 28 February 2024
ACCEPTED 04 July 2024
PUBLISHED 18 July 2024

CITATION
Ji Y, Hewavithana T, Sharpe AG and Jin L (2024) Understanding grain development in the Poaceae family by comparing conserved and distinctive pathways through omics studies in wheat and maize.
Front. Plant Sci. 15:1393140.
doi: 10.3389/fpls.2024.1393140

COPYRIGHT
© 2024 Ji, Hewavithana, Sharpe and Jin. 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.

Understanding grain development in the Poaceae family by comparing conserved and distinctive pathways through omics studies in wheat and maize

Yuanyuan Ji[1,2], Thulani Hewavithana[1], Andrew G. Sharpe[2*] and Lingling Jin[1*]

[1]Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada, [2]Global Institute for Food Security, University of Saskatchewan, Saskatoon, SK, Canada

The Poaceae family, commonly known as the grass family, encompasses a diverse group of crops that play an essential role in providing food, fodder, biofuels, environmental conservation, and cultural value for both human and environmental well-being. Crops in Poaceae family are deeply intertwined with human societies, economies, and ecosystems, making it one of the most significant plant families in the world. As the major reservoirs of essential nutrients, seed grain of these crops has garnered substantial attention from researchers. Understanding the molecular and genetic processes that controls seed formation, development and maturation can provide insights for improving crop yield, nutritional quality, and stress tolerance. The diversity in photosynthetic pathways between C3 and C4 plants introduces intriguing variations in their physiological and biochemical processes, potentially affecting seed development. In this review, we explore recent studies performed with omics technologies, such as genomics, transcriptomics, proteomics and metabolomics that shed light on the mechanisms underlying seed development in wheat and maize, as representatives of C3 and C4 plants respectively, providing insights into their unique adaptations and strategies for reproductive success.

KEYWORDS

seed development, omics, C3 plant, C4 plant, wheat, maize, genomics, transcriptome

1 Introduction

Plant seeds are remarkable vessels of life, encapsulating the potential for growth and ensuring the survival and dispersal of the species. They originate from the maturation of ovules within flowering plant, containing an embryo surrounded by a protective outer layer called the seed coat. The genetic repository within the embryo controls the entire life cycle of the plant and the seed coat shields it from environmental stresses and pathogens securing the unfolding of the plant's next generation in interaction with the environment (Kigel, 1995). Seeds stores the vital components of nutrient that nourish the development of a plant until it can photosynthesize on its own. As the seedling germinates, the embryo and cotyledons surrounded by the endosperm start to grow. Endosperm is a nutrient-rich tissue that mainly contains starch and provides essential energy to the germinating seedling. This resource allocation mechanism diverges between monocots and dicots. Monocots, characterized by a single cotyledon which is an embryonic leaf in the germinating seed, tend to retain their endosperm to nourish the growing seedling. Conversely, dicots often transfer nutrients from the endosperm to the cotyledons as the seed matures (Sabelli, 2012; Sreenivasulu and Wobus, 2013).

Approximately 60% of human energy supply is derived from three monocots cereal species — wheat, rice and maize, which all belong to the Poaceae family making Poaceae the most culturally and economically important plant family in the world (Linder et al., 2018). Therefore, seed development in the Poaceae family has been investigated through various approaches, including biochemistry, molecular biology and omics studies. There are studies describing the genotype/trait associations, genetics and transcriptional regulatory network of seeds development in wheat, rice and maize respectively (Dai et al., 2021; Shikha et al., 2021; Chen et al., 2022; Wang and Sun, 2023). However, no comparative analyses of omics data sets between wheat and maize in Poaceae family exist to the best of our knowledge. Therefore, understanding molecular and genetic mechanisms, modeling of biological networks by interrogation of entire pools of genomic, transcriptomic, metabolomic and proteomic data sets from the two crops with extensive comparative analysis remains a substantially important goal to further seed yield improvements.

Furthermore, it is well-established that the C4 pathway involves the incorporation of carbon into 4-carbon metabolites like malate and oxaloacetate, while the C3 photosynthetic pathway fixes CO_2 into 3-carbon metabolites such as 3-phosphoglycerate (PGA) through the Calvin cycle. There has been an ongoing debate surrounding whether wheat, a representative C3 photosynthetic crop, utilizes the C4 photosynthetic pathway during grain development like maize, in recent decades (Hibberd and Furbank, 2016). Omics studies have the potential to provide insights on certain questions regarding to this topic.

In the past decade, high-throughput sequencing technologies have revolutionized entire branches of the life sciences including plant improvement, human disease, pharmaceutical engineering. The term 'omics' is derived from genomics and signifies a holistic approach to the study of biological systems. It involves the comprehensive investigation of entire set of biological molecules and processes on a large scale, providing an in-depth understanding of complex biological systems (Hasin et al., 2017). There are many types of omics, each focusing on a specific aspect of biological data. Several of them have widely applied in plant science as described below.

1. Genomics is a study of the complete genome of a particular species. It focuses on sequencing of an entire genome and identifying genetic variants associated with plant traits or responses. In the plant research field, genome-wide association studies (GWAS) and quantitative trait locus (QTL) analysis are the most popular approaches to discover variants of interest associated with various plant traits (Korte and Farlow, 2013).

2. Transcriptomics is a study of entire RNA molecules transcribed in a cell or organism. It is mainly used for analysis of gene expression levels, alternative splicing, and non-coding RNA. RNA-Seq, a widely used short-read sequencing technology, offers significantly more accurate transcriptome profiling, enabling the detection of specific isoforms with greater precision compared to other methodologies like microarrays (Wang et al., 2009).

3. Proteomics is a comprehensive analysis of the entire protein components in a specific tissue from a particular species. Utilizing mass spectrometry (MS)-based proteomics, it is possible to characterize and quantify thousands of proteins simultaneously, and uncover their post-translational modifications (PTMs) in a parallel manner (Bantscheff et al., 2012; Dupree et al., 2020). Presently, the insights gleaned from proteomic research has substantially enhanced our comprehension of biological complexity. This advancement has improved our understanding of the molecular mechanisms driving plant responses to environmental stimuli and various developmental stages.

4. Metabolomics focuses on the analysis of small molecules (metabolites) within a biological system (cell, tissue, organism, etc). It offers a comprehensive snapshot of a plant's metabolic profile, encompassing the identification and quantification of diverse compounds like sugars, organic acids, and secondary metabolites. Notably, these plant secondary metabolites confer a multitude of advantages, including protective functions for the plants and health-promoting properties for consumers (Zhu et al., 2016; Kumar et al., 2017). By characterizing novel metabolites and examining their dynamic flows, we can significantly deepen our understanding of the pathways through which plants synthesize and regulate these vital compounds. These compounds are crucial not only for plant growth and development but also for their ability to respond effectively to various stressors.

Additionally, a range of omics disciplines, such as epigenomics, phenomics, and lipidomics, play a pivotal role in advancing research on plant seed development. This review is dedicated to offering a thorough overview of the current understanding of the molecular facets of wheat and maize seed development. Specifically, it focuses on delineating both the commonalities and differences between these two vital crops within the context of the aforementioned four omics domains.

2 Seed development in wheat and maize

The development of seeds is a complex process encompassing several maternal and filial tissues with a series of intricate events. These events begin with the fertilization of the ovule where one sperm cell fuses with the egg cell to form a zygote marking the start of the embryonic development. The cellularization events followed after fertilization and lead to the formation of triploid endosperm (two polar nuclei and one sperm cell) with continuous cell division (Kowles and Phillips, 1988; Berger, 1999). Unlike dicot species, that are dominated by the developing embryo, the endosperm in cereals like wheat and maize will continuously accumulates starch and storage proteins leading to a high-calories content. While the endosperms of cereal grains, like wheat, generally achieve full cellularization through ongoing cell division and alveolation (Figure 1), the endosperm of maize displays a somewhat distinct pattern of cellularization. Contrasting with wheat, where maize individual cells form distinct cell walls, the primary endosperm cell undergoes several rounds of nuclear divisions without the formation of cell walls. As a result, the endosperm remains multinucleate, with many nuclei sharing a common cytoplasm (Olsen et al., 1995). Hence, maize demonstrates an irregular final partitioning of the central vacuole at the base of its endosperm, which significantly influences the kernel size and contributes to its distinctive morphology (Leroux et al., 2014) (Figure 1). At the end of mitotic division, the development of the outer and inner layers, known as the pericarp and aleurone layer, respectively, signals the commencement of the grain filling stage. Owing to the reduced mitotic activity and limited cell size, the growth of the endosperm is predominantly driven by cell enlargement, characterized by the accumulation of starch, lipids, and proteins (Randolph, 1936; Kowles and Phillips, 1985; Li et al., 2006; Kong et al., 2015). Studies on the distribution and retention of radioactivity of $^{14}CO_2$ on cereal grains suggested that photosynthesis plays a crucial role in sustaining the majority of the grain weight (Merah and Monneveux, 2015; Yang et al., 2016; Estévez-Geffriaud et al., 2020). However, the sources of reserves remobilized for seed grain development vary between wheat and maize. In maize, although stored carbohydrates are available for seed growth, they are primarily allocated to maintenance processes, especially under significant stress. As a result, maize plants exhibit limited efficiency in utilizing these reserves for grain filling prior to flowering (Kiniry et al., 1992). Instead, these reserves are more effectively used in post-flowering stages, such as during leaf senescence. In contrast, wheat demonstrates a three- to four-fold higher efficiency in the remobilization of assimilates stored before flowering for seed growth, in comparison to maize (Borrás et al., 2004). Initial stages of seed development, including cellularization and grain filling, are characterized by intense transcriptional regulation across various pathways. These pathways encompass primary metabolism, cell division, stress response mechanisms, and protein synthesis and degradation (Sprunck et al., 2005; She et al., 2011; Chen et al., 2017a). However, as seeds advance toward maturity, they develop a suite of crucial physiological characteristics that ensure successful seedling establishment once they are independent of the parent plant. These traits hinge on the ability to undergo complete desiccation without losing viability, a phenomenon referred to as desiccation tolerance (Righetti et al., 2015). Therefore, the focus of transcriptional regulation shifts from primary metabolism pathways to processes like chromatin condensation and nuclear size reduction (Baroux et al., 2007). At this stage, a mature grain is composed of starchy endosperm (accounting for 83–84% of dry weight), embryo (3%), aleurone layer (6.5%), and pericarp (7–8%), with only minor differences observed between wheat and maize (Barron et al., 2007; Sethi et al., 2021).

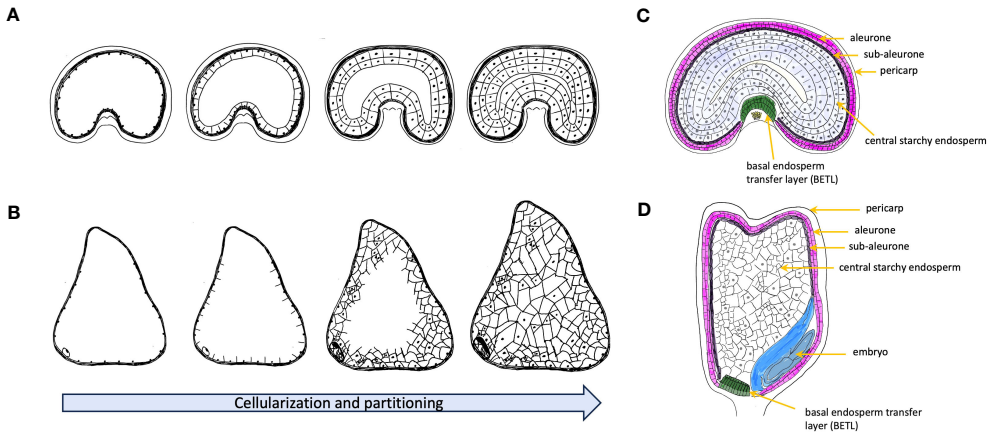


FIGURE 1
Schematic representation of early endosperm cellularization process before accumulation of storage metabolites for wheat **(A)** and maize **(B)**. Nuclear division initially in the coenocyte forms a multinucleate structure and then moves towards to the peripheral part of the multinucleate coenocyte. With formation of the first anticlinal cell walls, nuclear divisions shift towards to the center of multinucleate endosperm. The process continues with further growth moving inwards until cellularization is completed. Wheat embryo develops a uniform partitioning endosperm, and maize forms an irregular distribution of the cell division. **(C)** Schematic representation of cross section of a developing grain of wheat. **(D)** Schematic representation of longitudinal section of a developing grain of maize.

3 GWAS and QTL studies on seeds traits

QTL mapping is a statistical analysis to identify a region of DNA associated with a particular trait that exhibits variability within a population based on a genetic linkage map (Lander and Botstein, 1989). The mapping process requires the population generated between two genetically distinct parents with respect to the trait of interest (Myles and Wayne, 2008). In most of cases, a significant portion of the phenotypic variation within this region can be attributed to a small number of loci with substantial effects (Remington and Purugganan, 2003). The precisely mapping of QTL to a fine genetic position generally requires sufficient map resolution (Dinka et al., 2007).

With the emergence of high-throughput genotyping technologies, millions of markers can be identified at more affordable costs. GWAS has now become one of the most potent tools at the disposal of geneticists (Flint-Garcia et al., 2003). GWAS aims to identify associations between genotypes and traits of interest by leveraging phenotypic variations across diverse populations. These populations consist of unrelated individuals whose ancestry can be traced back to a common progenitor (Uffelmann et al., 2021). Consequently, the genetic distance between these individuals, influenced by numerous recombination events, leads to a faster decay in linkage disequilibrium, requiring extensive marker coverage to sufficiently capture the genome-wide genetic variations (Alqudah et al., 2020).

In recent decades, QTL mapping and GWAS have emerged as successful approaches, significantly advancing our understanding of gene functions in a variety of crops, including wheat, maize, and rice. It has been widely used to investigate a range of biological and physiological traits, as well as to uncover the genetic underpinnings of natural selection and population divergences within a species. There are numerous comprehensive reviews that delve into population design and statistical models used in breeding research (Bordes et al., 2014; Xiao et al., 2017; Shikha et al., 2021; Dwiningsih and Al-Kahtani, 2022; Gupta et al., 2022; Saini et al., 2022). Plant geneticists focus on a wide array of traits, ranging from disease resistance and grain yield to flowering time, biomass, and seed size, among others. Our review of literature from 2011 to the present reveals that disease resistance and grain yield are particularly prominent, accounting for over half of the studies published (773/1004 for wheat, 501/884 for maize). This emphasis highlights the critical importance of these two traits in agricultural research (Kaya and Akcura, 2014; Millet et al., 2019). (Bhatta et al., 2018; Qaseem et al., 2019). (Qaseem et al., 2019). (Muhammad et al., 2020).

Grain yield is a multifaceted trait, intricately linked to both the number of grains per unit area and the thousand-kernel weight (TKW). Furthermore, aspects like grain shape, spike architecture, plant height, and flag leaf size play a crucial role in influencing grain yield by affecting photosynthesis. In wheat research, a multitude of candidate loci have been identified, each associated with grain yield, size, and TKW. In the study conducted by Yu and colleagues, 768 bread wheat accessions were analyzed to pinpoint the genetic loci linked to grain size and weight. Notably, a principal locus related to grain length, designated as qGL-2D, exhibited the most robust association across multiple years and encompasses 125 annotated genes. Through transcriptome analysis, the researchers validated the gene *TraesCS2D02G414800*, which is responsible for encoding an oleosin—a protein crucial in seed maturation and germination—as a candidate gene within the qGL-2D locus (Yu et al., 2022). In a separate study conducted by Tekeu and team, three quantitative trait loci (QTLs) related to grain size were identified on chromosomes 1D, 2D, and 4A. This discovery was made through a GWAS that included 157 international wheat accessions and 71 Canadian wheat accessions. Notably, a candidate gene located on chromosome 2D, *TraesCS2D01G331100*, was identified as an ortholog of the D11 gene in rice. This gene, which encodes a cytochrome P450, displayed high expression levels in the developing embryo, particularly during embryogenesis and grain development in wheat, affecting both endosperm and pericarp tissues (Tekeu et al., 2021).

Although a vast array of candidate genes linked to grain yield have been discovered through GWAS and QTL mapping, the molecular cloning of these genes in wheat is still relatively unexplored (see Tables 1, 2). Unlike wheat, maize is a diploid species, the research on seeds development has greatly benefited from the investigation of mutator-induced mutant collections such as *defective kernel* (*dek*) mutants with affected development in both embryo and endosperm (Neuffer and Sheridan, 1980; McCarty et al., 2005), *empty pericarp* (*emp*) mutants (Scanlon et al., 1994), *defective endosperm* (*de*) or *endosperm-specific* mutant (Consonni et al., 2022) and *embryo specific* (*emb*) mutants (Clark and Sheridan, 1991). Those mutant collections provide abundant material for the molecular genetic analysis on the mechanisms of embryo and endosperm development allowing maize to become the leading system for molecular cloning of functional genes in plants. For instance, numerous *dek* mutants, such as *Dek1* (encoding a membrane protein of the calpain gene superfamily) (Lid et al., 2002), *Dek2*, *Dek35*, *Dek36*, *Dek37*, *Dek39* (encode a pentatricopeptide repeat protein) (Qi et al., 2017; Wang et al., 2017; Chen et al., 2017b; Dai et al., 2018; Li et al., 2018), and *Dek** (encoding AAA-ATPase controlling 60S ribosome export) were characterized. The associated pathways involve cell fate specification in the endosperm (*Dek1*), ribosome use efficiency (*Dek**) and RNA editing (*Dek2, Dek35–39*). However, compared to the mutant collections, the number of genes identified still represents only a small fraction of the entirety of the collections.

With the recent advancements of GWAS and QTL mapping, more natural genetic variations related to maize kernel traits had been identified (Tables 2, 3). By comparing the genes list between wheat and maize, there are some genes orthologs exist in both wheat and maize that effected the grain weight and size, such as *Grain Size 5* (*GS5*) which encodes a serine-type carboxypeptidase. The overexpression of *GS5* in wheat and maize both exhibit a positive regulation of grain weight suggesting a similar pathway controlling the yield (Liu et al., 2015; Ma et al., 2016). Remarkably, a recent GWAS, which examined a Chinese mini-core collection comprising 262 accessions—encompassing over 70% of the genetic diversity in Chinese wheat—uncovered a notable candidate gene, trehalose-6-phosphate phosphatase (*TaTPP-7A*). This gene is significantly

TABLE 1 List of cloned genes associated with seed traits in wheat, this list only covers the genes identified through GWAS and QTL mapping.

Gene	Seed trait	Protein function	Reference
TaGW2	Grain size and weight	encoding a C5HC2-type E3 ubiquitin ligase	Su et al. (2011)
			Li et al. (2023)
			Zhang et al. (2018)
			Zhai et al. (2018)
TaTEF-7A	Kernel number per spike	a member of the transcript elongation factor gene family	Zheng et al. (2014)
TaBT1	Grain weight	transporter for the translocation of small molecules between the mitochondria and cytoplasm	Wang et al. (2019b)
TaTPP-7A	Grain weight	Trehalose 6-phosphate phosphatase	Liu et al. (2023)
TaGS5–3A	Grain weight	serine carboxypeptidase	Ma et al. (2016)
TaDA1	Grain weight	a ubiquitin receptor	Liu et al. (2020)
TaCKX6-D1	Grain weight	a cytokinin oxidase/dehydrogenase	Zhang et al. (2012)
TaGASR7	Grain weight	encoding a protein with a signal peptide that responsive to gibberellic acid	Zhang et al. (2016)
			Dong et al. (2014)
TaAGP-S1–7A	Grain weight	subunit of ADP-glucose pyrophosphorylase	Hou et al. (2017)
TaAGP-L-1B	Grain weight	subunit of ADP-glucose pyrophosphorylase	Hou et al. (2017)
TaGW7	Grain size and weight	encoding a TONNEAU1-recruiting motif (TRM) protein	Wang et al. (2019a)
TaGW6	Grain weight	encoding a novel indole-3-acetic acid-glucose hydrolase	Hu et al. (2016)
TaIAA21	Grain size and weight	negative factor involved in Auxin signaling	Jia et al. (2021)
TaHST1L	Grain yield	encoding a NAD-dependent deacetylase involved in auxin signal	Zhao et al. (2023)

correlated with the TKW in wheat (Liu et al., 2023). Overexpression of *TaTPP-7A* in wheat induced a phenotype featuring higher TKW, longer grains and early maturity, while RNAi or gene-edited lines with silenced *TaTPP-7A* expression exhibited a reversed phenotype, including lower TKW, shorter grains, and late maturity. Further transcriptome analysis revealed that *TaTPP-7A* primarily boosts starch synthesis via the T6P-SnRK1 pathway and the sugar-ABA interaction. Intriguingly, the RAMOSA pathway, first proposed in maize decades ago, consists of three RAMOSA genes, including two transcription factors. The *RA1* gene encodes a Cys2-His2 zinc finger transcription factor, while *RA2* is responsible for a Lateral Organ Boundary (LOB) domain transcription factor. Meanwhile, *RA3* is identified as a gene encoding trehalose-6-phosphate phosphatase (TPP), which is expressed in the axillary inflorescence meristems (Wang et al., 2021). RA3 acts upstream of transcription factor RA1 to regulate inflorescence branching through modulation of treholose or T6P levels. In maize, two *TPP* genes, namely *RA3* and *TPP4*, have been characterized. Mutants of both *ra3* and *tpp4* demonstrated a decrease in crop yield, characterized by abnormal branching and irregular seed row patterns. Notably, the double mutant combining *ra3* and *tpp4* exhibited an even more pronounced phenotype (Satoh-Nagasawa et al., 2006; Claeys et al., 2019). Moreover, the overexpression of a rice *TPP* gene in the female reproductive tissues of maize significantly improved yield under drought conditions during the flowering stage (Nuccio et al., 2015). This, along with molecular evidence from mutants and transgenic lines in both wheat and maize, suggests a conserved pathway in cereal crops related to trehalose-6-phosphate (T6P), impacting grain size and yield.

In addition, increasing evidence has indicated the ubiquitin-proteasome pathway plays a conserved role in controlling the grain size through restricting the cell expansion in plants (Li and Li, 2016; Shi et al., 2019). *GW2* encodes a E3 ubiquitin ligase has been confirmed to be a negative regulators of grain size in rice (Song et al., 2007), maize (Li et al., 2010b) and wheat (Liu et al., 2020). Interestingly, a novel allele of *TaGW2* has been shown to increases grain weight but decrease grain number in wheat, suggesting additional roles of E3 ubiquitin ligase in controlling grain weight and size in wheat (Zhai et al., 2018).

Ortholog-based gene discovery is a valuable strategy, researchers often resort to identify orthologous genes from well-studied model plants. In the realm of cereal research, rice was widely used as a model species. In most cases, leveraging orthologs presents an effective avenue for gene discovery offering insights on conserved pathways and gene functions. However, it's important to acknowledge that relying solely on ortholog-based gene discovery has its limitations. As the evolutionary distance between species can influence the accuracy of orthologs predictions, potentially resulting mis-annotations or missed genes and overlooking species-specific adaptations and unique genetic features (Woodhouse and Hufford, 2019).

Despite the identification of numerous candidate genes through GWAS that exhibit promising influences on grain traits such as ear length and height in maize, the precise roles of these genes remain elusive. For example, the gene *YIGE1*, implicated in sugar and auxin signaling pathways, was discovered via GWAS. CRISPR/Cas9 knockout mutants of *YIGE1* displayed reduced grain yield, accompanied by a decrease in female inflorescence meristem size and the number of kernels per row (Luo et al., 2022). Nevertheless, the specific function of the YIGE1 protein is yet to be fully characterized.

4 Understanding of seed development through transcriptomic, proteomic and metabolomic studies

4.1 Conserved regulatory principles in the grain filling stage during endosperm development

The endosperm, forming the bulk of the seed, acts as a reservoir for essential components, predominantly starch and proteins. Starch, as a primary energy source, accounts for 70% to 80% of the dry weight in cereal seeds, while proteins comprise about 10%. Owing to its crucial role, the endosperm has become a focal point in diverse omics research fields.

In wheat, a significant number of genes, precisely 46,487 out of the total 85,173, were found to be active during endosperm development. This figure is remarkably similar to that observed in maize, where 46,274 genes were expressed (Li et al., 2014; Pfeifer et al., 2014). The variation in endosperm cell differentiation between wheat and maize influences the preferential expression of genes in specific cell types and developmental stages. Nevertheless, within common cell types like the aleurone (AL) and starchy endosperm (SE), there is a notable uniformity in the expression of specific

TABLE 2 Genomic positions of cloned genes associated with seed traits in wheat and maize through GWAS and QTL analysis.

Gene name	Gene ID	Physical position	Genome version
TaDA1	TraesCS2A01G007800	Chr2A:8319781 - 8326143	IWGSC CS RefSeq v2.0
TaGW2	TraesCS6A02G189300	Chr6A:240302888 - 240328295	IWGSC CS RefSeq v2.1
TaEF	TraesCS7A02G108900	Chr7A:69499125 - 69501354	IWGSC CS RefSeq v2.1
TaBT1	TraesCS6A02G175100	Chr6A:192285138 - 192288210	IWGSC CS RefSeq v2.1
TaTPP-7A	TraesCS7A03G0422300	Chr7A:138613354 - 138615932	IWGSC CS RefSeq v2.0
TaGS5–3A	TraesKAR3A01G0118380	Chr3A:188198450 - 188202484	Gramene release 68
TaCKX6-D1	TraesCS3D02G143300	Chr3D:85320853 - 85321389	IWGSC CS RefSeq v2.1
TaGASR7-A	TraesCS7A02G208100	Chr7A:175963414 - 175964402	IWGSC CS RefSeq v2.1
TaAGP-S1–7A	TraesCS7A02G287400	Chr7A:347014423 - 347022505	IWGSC CS RefSeq v2.1
TaAGP-L-1B	TraesCS1D03G0983500	Chr1D:482493243 - 482498148	IWGSC CS RefSeq v2.1
TaGW7–1B	TraesCS2B01G202300	Chr2B:190072020 - 190077273	IWGSC CS RefSeq v2.1
TaGW7–1D	TraesCS2D01G183400	Chr2D:130713880 - 130719328	IWGSC CS RefSeq v2.1
TaGW6-B	TraesCSU02G223800	ChrUnknown:249079212 - 249080545	IWGSC CS RefSeq v2.1
TaIAA21-A	TraesCS7A02G331100	Chr7A:488460835 - 488464693	IWGSC CS RefSeq v2.1
TaIAA21-B	TraesCS7B02G242800	Chr7B:455890694 - 455894566	IWGSC CS RefSeq v2.1
TaIAA21-D	TraesCS7D02G339300	Chr7D:435735509 - 435739416	IWGSC CS RefSeq v2.1
TaHST1L	TraesCS5A02G316400	Chr5A:403990309 - 403991592	IWGSC CS RefSeq v2.1
ZmGS3	GRMZM2G139878	Chr1:70729186 - 70731364	B73 RefGen_v3
ZmYIGE1	GRMZM2G008490	Chr1:50674670 - 50681478	B73 RefGen_v3
ZmVPS29	GRMZM2G068489	Chr4:224200238 - 224211317	B73 RefGen_v3
ZmG6PE	GRMZM2G039588	Chr2:4178376 - 4186062	B73 RefGen_v3
ZmMSH7	GRMZM2G110212	Chr3:9423562 - 9437621	B73 RefGen_v3
ZmGW2	GRMZM2G029690	Chr9:106080135 - 106087638	B73 RefGen_v3
ZmGS5	GRMZM2G123815	Chr3:61240172 - 61246469	B73 RefGen_v3
ZmBZR1	GRMZM6G287292	Chr3:5295292 - 5300354	B73 RefGen_v3
ZmFAD2	GRMZM2G056252	Chr5:203090777 - 203098676	B73 RefGen_v3
ZmNAC78	GRMZM2G406204	Chr1:4240471 - 4247496	B73 RefGen_v3
ZmMADS26	GRMZM2G044408	Chr5:210272259 - 210286975	B73 RefGen_v3

(Continued)

TABLE 2 Continued

Gene name	Gene ID	Physical position	Genome version
ZmTPS9	GRMZM2G312521	Chr4:174990784 - 174997304	B73 RefGen_v3
ZmcrtRB1	GRMZM2G152135	Chr10:136080567 - 136085686	B73 RefGen_v3
ZmKW9	GRMZM2G171994	Chr9:153757502 - 153762626	B73 RefGen_v3
ZmBAM1d	GRMZM2G043584	Chr1:30260694 - 30262694	B73 RefGen_v3

TABLE 3 List of cloned genes associated with seed traits in maize, this list only covers the genes identified through GWAS and QTL mapping.

Gene	Seed trait	Protein function	Reference
ZmGS3	Grain size	a negative regulator of grain size with four functional differential domains	Li et al. (2010b)
ZmYIGE1	Grain weight, Ear length	unknown protein	Luo et al. (2022)
ZmVPS29	Grain shape and size	encoding a retromer complex component involved in auxin biosynthesis	Chen et al. (2020)
ZmG6PE	Grain weight	encoding a glucose-6-phosphate 1-epimerase	Zhang et al. (2023a)
ZmMSH7	Grain weight	a member of MutS-homologous (MSH) family of proteins involved in DNA mismatch repair	Jiang et al. (2024)
ZmGW2	Grain weight and size	encoding a RING-type protein with E3 ubiquitin ligase activity	Li et al. (2010a)
ZmTB1	Grain weight	encoding a member of the TCP family of transcriptional regulators	Clark et al. (2006)
ZmGS5	Grain weight	encoding a protein with serine-type carboxypeptidase activity	Liu et al. (2015)
ZmBZR1	Grain size	transcription factor	Zhang et al. (2020)
ZmFAD2	Grain oleic acid level	encoding a fatty acid desaturase	Beló et al. (2008)
ZmNAC78	Grain iron content	transcription factor	Yan et al. (2023)
ZmMADS26	Grain germination	a MADS-transcription factor 26	Ma et al. (2022)
ZmTPS9	Grain weight, Starch content	encoding a trehalose-6-phosphate synthase in the trehalose pathway	Hu et al. (2021)
ZmcrtRB1	Grain β-carotene content	encoding β-carotene hydroxylase 1	Yan et al. (2010); Yin et al. (2024)
ZmKW9	Grain weight	encoding a DYW motif pentatricopeptide repeat protein	Huang et al. (2020); Zhang et al. (2024)

(Continued)

TABLE 3 Continued

Gene	Seed trait	Protein function	Reference
ZmBAM1d	Grain weight	encoding a CLAVATA1 (CLV1)/BARELY ANY MERISTEM (BAM)-related receptor kinase-like protein	Yang et al. (2019)

genes. Genes exclusive to AL are predominantly linked to lipid metabolism, carbohydrate metabolic processes, and amino acid biosynthesis. Conversely, genes specific to SE primarily focus on carbohydrate and saccharide metabolism (Woo et al., 2001; Gillies et al., 2012; Li et al., 2014; Pfeifer et al., 2014). Additionally, two significant studies centered on wheat grains utilized time-series metabolomic analysis in conjunction with proteomic techniques. These studies successfully uncovered the temporal and spatial distribution of proteins and metabolites within various cell types, providing a more nuanced understanding of their dynamics (Zhang et al., 2021; Zhang et al, 2023b). These studies identified distinct proteins and metabolites in four critical cell types: aleurone (AL), sub-aleurone (SA), starchy endosperm (SE), and endosperm transfer cells (ETC). Notably, during the essential 15-day period post-anthesis (DAA), a pivotal phase in grain development, carbohydrates such as sucrose, glucose, fructose, and myo-inositol showed significant accumulation in SE and ETC. Enzymes involved in carbohydrate interconversion, including beta-fructofuranosidase insoluble isoenzyme 2, 1,2-beta-Fructan 1^F-fructosyltransferase, and sucrose synthase, exhibited a similar trend. This pattern aligns with the transcriptome and metabolome data (Zhang et al., 2023b). The findings from these studies suggest the presence of conserved regulatory principles within the cereal family.

While starch and protein are the primary constituents of cereal crops like maize and wheat, the distribution of these polymers during seed maturation exhibits spatial differences among various plant species. A striking example is evident in a proteomic profiling study of 16 cereal grains, which revealed significant variations in the gluten-enriched fractions of storage proteins between wheat and maize (Colgrave et al., 2015). Wheat varieties, along with barley and rye, predominantly accumulate high molecular weight glutenins, ranging from 35 to 180 kDa. Conversely, in maize, the gluten-enriched fractions chiefly consist of zeins, with molecular weights typically between 9 to 29 kDa. Such divergence in polymer distribution highlights the complexity of the seed maturation process across different cereal species.

4.2 Do wheat seeds employ a C4-type carbon concentrating mechanism like maize?

As a prototypical C4 plant, maize utilizes a specialized cell-type-specific expression pattern for photosynthesis, which allows for the spatial segregation of this process. Key enzymes such as phosphoenolpyruvate carboxylase (PEPC), malate dehydrogenase (MDH), pyruvate orthophosphate dikinase (PPDK), and carbonic anhydrase (CA) are predominantly expressed in the outer mesophyll cells. In contrast, enzymes like ribulose 1,5-bisphosphate carboxylase-oxygenase (Rubisco), dicarboxylic acid transporter (DCT), and NAD-dependent malic enzyme (NAD-ME) are primarily localized in the inner bundle sheath (BS) cells [(Weber and von Caemmerer, 2010; Gowik et al., 2011; Medeiros et al., 2021)]. C4 photosynthesis is designed to concentrate CO_2 in the vicinity of the primary carboxylase, Rubisco, which can bind to CO_2 or O_2 in a reversible reaction. This mechanism significantly enhances photosynthetic efficiency by inhibiting photorespiration, a process that competes with photosynthesis.

In wheat, although the endosperm is incapable of photosynthesis, the surrounding green tissues, such as the pericarp and glumes, are crucial for supporting seed development via photosynthesis. The major contributors to grain filling are photosynthesis in the flag leaf and ear. Despite the pericarp displaying a bright green color during grain filling, it likely re-assimilates CO_2 from respiration rather than direct photosynthesis due to limited stomata. In contrast, the glumes, which envelop the seed, possess a higher density of stomata and feature specialized Kranz cells around their vascular bundles. This morphology suggests a potential role for these cells in a form of C4 photosynthesis (Cochrane and Duffus, 1979).

In the field of omics research, particularly concerning the Poaceae family and the study of C3 and C4 plants, transcriptomic studies are of paramount importance. Recent RNA-Seq experiments spanning various wheat genotypes have yielded a wealth of data. During the mid-development stage of the seed, active photosynthesis in the pericarp is indicated by the expression of related genes, but this activity decreases as the seed matures (Rangan et al., 2017). Additionally, transcriptome analysis of wheat leaves and seeds revealed the expression of a C4-specific form of PEPC, suggesting the involvement of the C4 pathway in the pericarp, possibly to capture carbon released from endosperm respiration. Genes integral to C4 photosynthesis, including PEPC, MDH, NAD-ME, and PPDK, were identified, with all transcripts accumulating in the pericarp (Rangan et al., 2016). Consequently, it is hypothesized that C4 photosynthesis may play a role in grain filling in wheat, particularly via the pericarp.

Typically, assessments of mRNA abundance are common methodologies used to infer the proteotype. However, it has been clearly established that there is a tenuous correlation between mRNA levels and protein abundance (Petricka et al., 2012; Vogel and Marcotte, 2012). This is particularly true for metabolic enzymes, where protein functionality is complex and may vary due to amino acid substitutions affecting substrate and product relationships. Accurately characterizing different isoenzymes necessitates more in-depth carbon flux analyses and metabolomic studies, especially focusing on the pericarp. For instance, a previous experiment using radio-labeled $^{14}CO_2$ to trace carbon flux in wheat grains revealed that approximately 10% of the $^{14}CO_2$ was incorporated into the C4 acids malate and aspartate (Bort et al., 1995). Therefore, a more thorough analysis of carbon flux specifically from the pericarp would be particularly revealing.

To comprehensively understand the role of photosynthesis in grain filling within cereal crops, an integrated multi-omics approach that synthesizes insights from genomics, transcriptomics, proteomics, and metabolomics is essential. The current debate regarding the contributions of photosynthetic metabolism in wheat ears highlights the imperative for ongoing research. This research is necessary to unravel the complex regulatory mechanisms that govern these vital processes in cereal crops.

5 Conclusion

This review presents an in-depth exploration of the complex mechanisms involved in seed development in wheat and maize, with a particular focus on the molecular dimensions as revealed through genomics, transcriptomics, proteomics, and metabolomics. Despite considerable research on these individual species, there is a notable deficiency in comprehensive comparative studies between C3 and C4 plants within the Poaceae family. The conclusion emphasizes the essential need for a multidisciplinary approach to decode the complexities of photosynthetic contributions to grain filling in cereal crops. It draws attention to the ongoing debates within this field and underlines the critical importance of sustained research. Future research should aim to clarify the regulatory principles, particularly through an integrated lens of genomics, transcriptomics, proteomics and metabolomics.

Author contributions

YJ: Conceptualization, Writing – original draft, Writing – review & editing. TH: Writing – original draft, Writing – review & editing. AS: Supervision, Writing – review & editing. LJ: Conceptualization, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This project (Project number: 20210933) was funded by the Agriculture Development Fund (ADF) under the Ministry of Agriculture of Saskatchewan.

Acknowledgments

We thank Nadeem Khan for critical reading of the manuscript.

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

Alqudah, A. M., Sallam, A., Baenziger, P. S., and Börner, A. (2020). GWAS: fast-forwarding gene identification and characterization in temperate cereals: lessons from barley–a review. *J. Adv. Res.* 22, 119–135. doi: 10.1016/j.jare.2019.10.013

Bantscheff, M., Lemeer, S., Savitski, M. M., and Kuster, B. (2012). Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 404, 939–965. doi: 10.1007/s00216-012-6203-4

Baroux, C., Pien, S., and Grossniklaus, U. (2007). Chromatin modification and remodeling during early seed development. *Curr. Opin. Genet. Dev.* 17, 473–479. doi: 10.1016/j.gde.2007.09.004

Barron, C., Surget, A., and Rouau, X. (2007). Relative amounts of tissues in mature wheat (*Triticum aestivum L.*) grain and their carbohydrate and phenolic acid composition. *J. Cereal Sci.* 45, 88–96. doi: 10.1016/j.jcs.2006.07.004

Beló, A., Zheng, P., Luck, S., Shen, B., Meyer, D. J., Li, B., et al. (2008). Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize. *Mol. Genet. Genomics* 279, 1–10. doi: 10.1007/s00438-007-0289-y

Berger, F. (1999). Endosperm development. *Curr. Opin. Plant Biol.* 2, 28–32. doi: 10.1016/S1369-5266(99)80006-5

Bhatta, M., Morgounov, A., Belamkar, V., and Baenziger, P. S. (2018). Genome-wide association study reveals novel genomic regions for grain yield and yield-related traits in drought-stressed synthetic hexaploid wheat. *Int. J. Mol. Sci.* 19, 3011. doi: 10.3390/ijms19103011

Bordes, J., Goudemand, E., Duchalais, L., Chevarin, L., Oury, F. X., Heumez, E., et al. (2014). Genome-wide association mapping of three important traits using bread wheat elite breeding populations. *Mol. Breed.* 33, 755–768. doi: 10.1007/s11032-013-0004-0

Borrás, L., Slafer, G. A., and Otegui, M. E. (2004). Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal. *Field Crops Res.* 86, 131–146. doi: 10.1016/j.fcr.2003.08.002

Bort, J., Brown, R., and Araus, J. (1995). Lack of C4 photosynthetic metabolism in ears of C3 cereals. *Plant Cell Environ.* 18, 697–702. doi: 10.1111/j.1365-3040.1995.tb00571.x

Chen, J., Strieder, N., Krohn, N. G., Cyprys, P., Sprunck, S., Engelmann, J. C., et al. (2017a). Zygotic genome activation occurs shortly after fertilization in maize. *Plant Cell* 29, 2106–2125. doi: 10.1105/tpc.17.00099

Chen, L., Li, Y.-X., Li, C., Shi, Y., Song, Y., Zhang, D., et al. (2020). The retromer protein ZmVPS29 regulates maize kernel morphology likely through an auxin-dependent process (es). *Plant Biotechnol. J.* 18, 1004–1014. doi: 10.1111/pbi.13267

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

Chen, X., Feng, F., Qi, W., Xu, L., Yao, D., Wang, Q., et al. (2017b). Dek35 encodes a PPR protein that affects cis-splicing of mitochondrial nad4 intron 1 and seed development in maize. *Mol. Plant* 10, 427–441. doi: 10.1016/j.molp.2016.08.008

Claeys, H., Vi, S. L., Xu, X., Satoh-Nagasawa, N., Eveland, A. L., Goldshmidt, A., et al. (2019). Control of meristem determinacy by trehalose 6-phosphate phosphatases is uncoupled from enzymatic activity. *Nat. Plants* 5, 352–357. doi: 10.1038/s41477-019-0394-z

Clark, J. K., and Sheridan, W. F. (1991). Isolation and characterization of 51 embryo-specific mutations of maize. *Plant Cell* 3, 935–951. doi: 10.2307/3869156

Clark, R. M., Wagler, T. N., Quijada, P., and Doebley, J. (2006). A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* 38, 594–597. doi: 10.1038/ng1784

Cochrane, M., and Duffus, C. (1979). Morphology and ultrastructure of immature cereal grains in relation to transport. *Ann. Bot.* 44, 67–72. doi: 10.1093/oxfordjournals.aob.a085707

Colgrave, M. L., Goswami, H., Byrne, K., Blundell, M., and Howitt, C. A. (2015). Proteomic profiling of 16 cereal grains and the application of targeted proteomics to detect wheat contamination. *J. J. Proteome Res.* 14, 2659–2668. doi: 10.1021/acs.jproteome.5b00187

Consonni, G., Castorina, G., and Varotto, S. (2022). The Italian research on the molecular characterization of maize kernel development. *Int. J. Mol. Sci.* 23, 11383. doi: 10.3390/ijms231911383

Dai, D., Luan, S., Chen, X., Wang, Q., Feng, Y., Zhu, C., et al. (2018). Maize Dek37 encodes a P-type PPR protein that affects cis-splicing of mitochondrial nad2 intron 1 and seed development. *Genetics* 208, 1069–1082. doi: 10.1534/genetics.117.300602

Dai, D., Ma, Z., and Song, R. (2021). Maize endosperm development. *J. Integr. Plant Biol.* 63, 613–627. doi: 10.1111/jipb.13069

Dinka, S. J., Campbell, M. A., Demers, T., and Raizada, M. N. (2007). Predicting the size of the progeny mapping population required to positionally clone a gene. *Genetics* 176, 2035–2054. doi: 10.1534/genetics.107.074377

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

Dupree, E. J., Jayathirtha, M., Yorkey, H., Mihasan, M., Petre, B. A., and Darie, C. C. (2020). A critical review of bottom-up proteomics: the good, the bad, and the future of this field. *Proteomes* 8, 14. doi: 10.3390/proteomes8030014

Dwiningsih, Y., and Al-Kahtani, J. (2022). Genome-wide association study of complex traits in maize detects genomic regions and genes for increasing grain yield and grain quality. *Adv. Sustain. Sci. Eng. Technol.* 4, 0220209. doi: 10.26877/asset.v4i2.12678

Estévez-Geffriaud, V., Vicente, R., Vergara-Díaz, O., Narváez Reinaldo, J. J., and Trillas, M. I. (2020). Application of *Trichoderma asperellum* T34 on maize (*Zea mays*) seeds protects against drought stress. *Planta* 252, 1–12. doi: 10.1007/s00425-020-03404-3

Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, IV, E. S. (2003). Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907

Gillies, S. A., Futardo, A., and Henry, R. J. (2012). Gene expression in the developing aleurone and starchy endosperm of wheat. *Plant Biotechnol. J.* 10, 668–679. doi: 10.1111/j.1467-7652.2012.00705.x

Gowik, U., Bräutigam, A., Weber, K. L., Weber, A. P., and Westhoff, P. (2011). Evolution of C4 photosynthesis in the genus Flaveria: how many and which genes does it take to make C4? *Plant Cell* 23, 2087–2105. doi: 10.1105/tpc.111.086264

Gupta, O. P., Singh, A. K., Singh, A., Singh, G. P., Bansal, K. C., and Datta, S. K. (2022). Wheat biofortification: Utilizing natural genetic diversity, genome-wide association mapping, genomic selection, and genome editing technologies. *Front. Nutr.* 9, 826131. doi: 10.3389/fnut.2022.826131

Hasin, Y., Seldin, M., and Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol.* 18, 1–15. doi: 10.1186/s13059-017-1215-1

Hibberd, J. M., and Furbank, R. T. (2016). Wheat genomics: Seeds of C4 photosynthesis. *Nat. Plants* 2, 1–2. doi: 10.1038/nplants.2016.172

Hou, J., Li, T., Wang, Y., Hao, C., Liu, H., and Zhang, X. (2017). ADP-glucose pyrophosphorylase genes, associated with kernel weight, underwent selection during

wheat domestication and breeding. *Plant Biotechnol. J.* 15, 1533–1543. doi: 10.1111/pbi.12735

Hu, M.-J., Zhang, H.-P., Cao, J.-J., Zhu, X.-F., Wang, S.-X., Jiang, H., et al. (2016). Characterization of an IAA-glucose hydrolase gene *TaTGW6* associated with grain weight in common wheat (*Triticum aestivum L.*). *Mol. Breed.* 36, 1–11. doi: 10.1007/s11032-016-0449-z

Hu, S., Wang, M., Zhang, X., Chen, W., Song, X., Fu, X., et al. (2021). Genetic basis of kernel starch content decoded in a maize multi-parent population. *Plant Biotechnol. J.* 19, 2192–2205. doi: 10.1111/pbi.13645

Huang, J., Lu, G., Liu, L., Raihan, M. S., Xu, J., Jian, L., et al. (2020). The kernel size-related quantitative trait locus *qKW9* encodes a pentatricopeptide repeat protein that aaffects photosynthesis and grain filling. *Plant Physiol.* 183, 1696–1709. doi: 10.1104/pp.20.00374

Jia, M., Li, Y., Wang, Z., Tao, S., Sun, G., Kong, X., et al. (2021). *TaIAA21* represses *TaARF25*-mediated expression of *TaERFs* required for grain size and weight development in wheat. *Plant J.* 108, 1754–1767. doi: 10.1111/tpj.15541

Jiang, L., Guo, T., Song, X., Jiang, H., Lu, M., Luo, J., et al. (2024). MSH7 confers quantitative variation in pollen fertility and boosts grain yield in maize. *Plant Biotechnol. J.* doi: 10.1111/pbi.14272

Kaya, Y., and Akcura, M. (2014). Effects of genotype and environment on grain yield and quality traits in bread wheat (*T. aestivum L.*). *Food Sci. Technol.* 34, 386–393. doi: 10.1590/fst.2014.0041

Kigel, J. (1995). *Seed development and germination* Vol. 41 (New York: CRC press).

Kiniry, J., Tischler, C., Rosenthal, W., and Gerik, T. (1992). Nonstructural carbohydrate utilization by sorghum and maize shaded during grain growth. *Crop Sci.* 32, 131–137. doi: 10.2135/cropsci1992.0011183X003200010029x

Kong, L., Guo, H., and Sun, M. (2015). Signal transduction during wheat grain development. *Planta* 241, 789–801. doi: 10.1007/s00425-015-2260-1

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9, 1–9. doi: 10.1186/1746-4811-9-29

Kowles, R. V., and Phillips, R. L. (1985). DNA amplification patterns in maize endosperm nuclei during kernel development. *Proc. Natl. Acad. Sci.* 82, 7010–7014. doi: 10.1073/pnas.82.20.7010

Kowles, R. V., and Phillips, R. L. (1988). Endosperm development in maize. *Int. Rev. Cytology* 112, 97–136. doi: 10.1016/S0074-7696(08)62007-0

Kumar, R., Bohra, A., Pandey, A. K., Pandey, M. K., and Kumar, A. (2017). Metabolomics for plant improvement: status and prospects. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01302

Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199. doi: 10.1093/genetics/121.1.185

Leroux, B. M., Goodyke, A. J., Schumacher, K. I., Abbott, C. P., Clore, A. M., Yadegari, R., et al. (2014). Maize early endosperm growth and development: from fertilization through cell type differentiation. *Am. J. Bot.* 101, 1259–1274. doi: 10.3732/ajb.1400083

Li, G., Wang, D., Yang, R., Logan, K., Chen, H., Zhang, S., et al. (2014). Temporal patterns of gene expression in developing maize endosperm identified through transcriptome sequencing. *Proc. Natl. Acad. Sci.* 111, 7582–7587. doi: 10.1073/pnas.1406383111

Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33, 23–32. doi: 10.1016/j.pbi.2016.05.008

Li, Q., Li, L., Yang, X., Warburton, M. L., Bai, G., Dai, J., et al. (2010a). Relationship, evolutionary fate and function of two maize co-orthologs of rice *GW2* associated with kernel size and weight. *BMC Plant Biol.* 10, 1–15. doi: 10.1186/1471-2229-10-143

Li, Q., Yang, X., Bai, G., Warburton, M. L., Mahuku, G., Gore, M., et al. (2010b). Cloning and characterization of a putative *GS3* ortholog involved in maize kernel development. *Theor. Appl. Genet.* 120, 753–763. doi: 10.1007/s00122-009-1196-x

Li, S., Zhang, Y., Liu, Y., Zhang, P., Wang, X., Chen, B., et al. (2023). The E3 ligase *TaGW2* mediates transcription factor *TaARR12* degradation to promote drought resistance in wheat. *Plant Cell* 36, 605–625. doi: 10.1093/plcell/koad307

Li, X., Gu, W., Sun, S., Chen, Z., Chen, J., Song, W., et al. (2018). Defective kernel 39 encodes a PPR protein required for seed development in maize. *J. Integr. Plant Biol.* 60, 45–64. doi: 10.1111/jipb.v60.1

Li, X., Wang, H., Li, H., Zhang, L., Teng, N., Lin, Q., et al. (2006). Awns play a dominant role in carbohydrate production during the grain-filling stages in wheat (*Triticum aestivum*). *Physiologia plantarum* 127, 701–709. doi: 10.1111/j.1399-3054.2006.00679.x

Lid, S. E., Gruis, D., Jung, R., Lorentzen, J. A., Ananiev, E., Chamberlin, M., et al. (2002). The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. *Proc. Natl. Acad. Sci.* 99, 5460–5465. doi: 10.1073/pnas.042098799

Linder, H. P., Lehmann, C. E., Archibald, S., Osborne, C. P., and Richardson, D. M. (2018). Global grass (Poaceae) success underpinned by traits facilitating colonization, persistence and habitat transformation. *Biol. Rev.* 93, 1125–1144. doi: 10.1111/brv.12388

Liu, J., Deng, M., Guo, H., Raihan, S., Luo, J., Xu, Y., et al. (2015). Maize orthologs of rice *GS5* and their trans-regulator are associated with kernel development. *J. Integr. Plant Biol.* 57, 943–953. doi: 10.1111/jipb.12421

Liu, H., Li, H., Hao, C., Wang, K., Wang, Y., Qin, L., et al. (2020). *TaDA1*, a conserved negative regulator of kernel size, has an additive effect with *TaGW2* in common wheat (*Triticum aestivum* l.). *Plant Biotechnol. J.* 18, 1330–1342. doi: 10.1111/pbi.13298

Liu, H., Si, X., Wang, Z., Cao, L., Gao, L., Zhou, X., et al. (2023). *TaTPP-7A* positively feedback regulates grain filling and wheat grain yield through T6P-SnRK1 signalling pathway and sugar–ABA interaction. *Plant Biotechnol. J.* doi: 10.1111/pbi.14025

Luo, Y., Zhang, M., Liu, Y., Liu, J., Li, W., Chen, G., et al. (2022). Genetic variation in *YIGE1* contributes to ear length and grain yield in maize. *New Phytol.* 234, 513–526. doi: 10.1111/nph.17882

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, L., Wang, C., Hu, Y., Dai, W., Liang, Z., Zou, C., et al. (2022). GWAS and transcriptome analysis reveal *MADS26* involved in seed germination ability in maize. *Theor. Appl. Genet.* 135, 1717–1730. doi: 10.1007/s00122-022-04065-4

McCarty, D. R., Mark Settles, A., Suzuki, M., Tan, B. C., Latshaw, S., Porch, T., et al. (2005). Steady-state transposon mutagenesis in inbred maize. *Plant J.* 44, 52–61. doi: 10.1111/j.1365-313X.2005.02509.x

Medeiros, D. B., Brotman, Y., and Fernie, A. R. (2021). The utility of metabolomics as a tool to inform maize biology. *Plant Commun.* 2. doi: 10.1016/j.xplc.2021.100187

Merah, O., and Monneveux, P. (2015). Contribution of different organs to grain filling in durum wheat under Mediterranean conditions i. contribution of post-anthesis photosynthesis and remobilization. *J. Agron. Crop Sci.* 201, 344–352. doi: 10.1111/jac.12109

Millet, E. J., Kruijer, W., Coupel-Ledru, A., Alvarez Prado, S., Cabrera-Bosquet, L., Lacube, S., et al. (2019). Genomic prediction of maize yield across European environmental conditions. *Nat. Genet.* 51, 952–956. doi: 10.1038/s41588-019-0414-y

Muhammad, A., Hu, W., Li, Z., Li, J., Xie, G., Wang, J., et al. (2020). Appraising the genetic architecture of kernel traits in hexaploid wheat using GWAS. *Int. J. Mol. Sci.* 21, 5649. doi: 10.3390/ijms21165649

Myles, C., and Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nat. Educ.* 1, 208.

Neuffer, M., and Sheridan, W. F. (1980). Defective kernel mutants of maize. I. genetic and lethality studies. *Genetics* 95, 929–944. doi: 10.1093/genetics/95.4.929

Nuccio, M. L., Wu, J., Mowers, R., Zhou, H.-P., Meghji, M., Primavesi, L. F., et al. (2015). Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. *Nat. Biotechnol.* 33, 862–869. doi: 10.1038/nbt.3277

Olsen, O.-A., Brown, R., and Lemmon, B. (1995). Pattern and process of wall formation in developing endosperm. *BioEssays* 17, 803–812. doi: 10.1002/bies.950170910

Petricka, J. J., Schauer, M. A., Megraw, M., Breakfield, N. W., Thompson, J. W., Georgiev, S., et al. (2012). The protein expression landscape of the Arabidopsis root. *Proc. Natl. Acad. Sci.* 109, 6811–6818. doi: 10.1073/pnas.1202546109

Pfeifer, M., Kugler, K. G., Sandve, S. R., Zhan, B., Rudi, H., Hvidsten, T. R., et al. (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345, 1250091. doi: 10.1126/science.1250091

Qaseem, M. F., Qureshi, R., Shaheen, H., and Shafqat, N. (2019). Genome-wide association analyses for yield and yield-related traits in bread wheat (*Triticum aestivum L.*) under pre-anthesis combined heat and drought stress in field conditions. *PloS One* 14, e0213407. doi: 10.1371/journal.pone.0213407

Qi, W., Yang, Y., Feng, X., Zhang, M., and Song, R. (2017). Mitochondrial function and maize kernel development requires Dek2, a pentatricopeptide repeat protein involved in nad1 mrna splicing. *Genetics* 205, 239–249. doi: 10.1534/genetics.116.196105

Randolph, L. F. (1936). *Developmental morphology of the caryopsis in maize*. (Washington, D.C.: United States Department of Agriculture).

Rangan, P., Furtado, A., and Henry, R. J. (2016). New evidence for grain specific C4 photosynthesis in wheat. *Sci. Rep.* 6, 31721. doi: 10.1038/srep31721

Rangan, P., Furtado, A., and Henry, R. J. (2017). The transcriptome of the developing grain: a resource for understanding seed development and the molecular control of the functional and nutritional properties of wheat. *BMC Genomics* 18, 1–9. doi: 10.1186/s12864-017-4154-z

Remington, D. L., and Purugganan, M. D. (2003). Candidate genes, quantitative trait loci, and functional trait evolution in plants. *Int. J. Plant Sci.* 164, S7–S20. doi: 10.1086/367812

Righetti, K., Vu, J. L., Pelletier, S., Vu, B. L., Glaab, E., Lalanne, D., et al. (2015). Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. *Plant Cell* 27, 2692–2708. doi: 10.1105/tpc.15.00632

Sabelli, P. A. (2012). "Seed development: a comparative overview on biology of morphology, physiology, and biochemistry between monocot and dicot plants," in *Seed development: OMICS technologies toward improvement of seed quality and crop yield*. (Dordrecht, Netherlands: Springer Dordrecht), 3–25.

Saini, D. K., Chopra, Y., Singh, J., Sandhu, K. S., Kumar, A., Bazzer, S., et al. (2022). Comprehensive evaluation of mapping complex traits in wheat using genome-wide association studies. *Mol. Breed.* 42, 1–52. doi: 10.1007/s11032-021-01272-7

Satoh-Nagasawa, N., Nagasawa, N., Malcomber, S., Sakai, H., and Jackson, D. (2006). A trehalose metabolic enzyme controls inflorescence architecture in maize. *Nature* 441, 227–230. doi: 10.1038/nature04725

Scanlon, M. J., Stinard, P. S., James, M. G., Myers, A. M., and Robertson, D. S. (1994). Genetic analysis of 63 mutations affecting maize kernel development isolated from mutator stocks. *Genetics* 136, 281–294. doi: 10.1093/genetics/136.1.281

Sethi, M., Singh, A., Kaur, H., Phagna, R. K., Rakshit, S., and Chaudhary, D. P. (2021). Expression profile of protein fractions in the developing kernel of normal, Opaque-2 and quality protein maize. *Sci. Rep.* 11, 2469. doi: 10.1038/s41598-021-81906-0

She, M., Ye, X., Yan, Y., Howit, C., Belgard, M., and Ma, W. (2011). Gene networks in the synthesis and deposition of protein polymers during grain development of wheat. *Funct. Integr. Genomics* 11, 23–35. doi: 10.1007/s10142-010-0196-x

Shi, C., Ren, Y., Liu, L., Wang, F., Zhang, H., Tian, P., et al. (2019). Ubiquitin specific protease 15 has an important role in regulating grain width and size in rice. *Plant Physiol.* 180, 381–391. doi: 10.1104/pp.19.00065

Shikha, K., Shahi, J., Vinayan, M., Zaidi, P., Singh, A., and Sinha, B. (2021). Genome-wide association mapping in maize: status and prospects. *3 Biotech.* 11, 244. doi: 10.1007/s13205-021-02799-4

Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z., and Lin, H.-X. (2007). A qtl for rice grain width and weight encodes a previously unknown ring-type e3 ubiquitin ligase. *Nat. Genet.* 39, 623–630. doi: 10.1038/ng2014

Sprunck, S., Baumann, U., Edwards, K., Langridge, P., and Dresselhaus, T. (2005). The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum L.*). *Plant J.* 41, 660–672. doi: 10.1111/j.1365-313X.2005.02332.x

Sreenivasulu, N., and Wobus, U. (2013). Seed-development programs: a systems biology–based comparison between dicots and monocots. *Annu. Rev. Plant Biol.* 64, 189–217. doi: 10.1146/annurev-arplant-050312-120215

Su, Z., Hao, C., Wang, L., Dong, Y., and Zhang, X. (2011). Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat (*Triticum aestivum L.*). *Theor. Appl. Genet.* 122, 211–223. doi: 10.1007/s00122-010-1437-z

Tekeu, H., Ngonkeu, E. L., Bélanger, S., Djocgoué, P. F., Abed, A., Torkamaneh, D., et al. (2021). GWAS identifies an ortholog of the rice *D11* gene as a candidate gene for grain size in an international collection of hexaploid wheat. *Sci. Rep.* 11, 19483. doi: 10.1038/s41598-021-98626-0

Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Primers* 1, 59. doi: 10.1038/s43586-021-00056-9

Vogel, C., and Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. doi: 10.1038/nrg3185

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, Y., Hou, J., Liu, H., Li, T., Wang, K., Hao, C., et al. (2019b). *TaBT1*, affecting starch synthesis and thousand kernel weight, underwent strong selection during wheat improvement. *J. Exp. Bot.* 70, 1497–1511. doi: 10.1093/jxb/erz032

Wang, W., Pan, Q., Tian, B., He, F., Chen, Y., Bai, G., et al. (2019a). Gene editing of the wheat homologs of TONNEAU 1-recruiting motif encoding gene affects grain shape and weight in wheat. *Plant J.* 100, 251–264. doi: 10.1111/tpj.14440

Wang, Y., and Sun, G. (2023). Molecular prospective on the wheat grain development. *Crit. Rev. Biotechnol.* 43, 38–49. doi: 10.1080/07388551.2021.2001784

Wang, C., Yang, X., and Li, G. (2021). Molecular insights into inflorescence meristem specification for yield potential in cereal crops. *Int. J. Mol. Sci.* 22, 3508. doi: 10.3390/ijms22073508

Wang, G., Zhong, M., Shuai, B., Song, J., Zhang, J., Han, L., et al. (2017). E+ subgroup PPR protein defective kernel 36 is required for multiple mitochondrial transcripts editing and seed development in maize and arabidopsis. *New Phytol.* 214, 1563–1578. doi: 10.1111/nph.14507

Weber, A. P., and von Caemmerer, S. (2010). Plastid transport and metabolism of C3 and C4 plants—comparative analysis and possible biotechnological exploitation. *Curr. Opin. Plant Biol.* 13, 256–264. doi: 10.1016/j.pbi.2010.01.007

Woo, Y.-M., Hu, D. W.-N., Larkins, B. A., and Jung, R. (2001). Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *Plant Cell* 13, 2297–2317. doi: 10.1105/tpc.010240

Woodhouse, M., and Hufford, M. (2019). Parallelism and convergence in post-domestication adaptation in cereal grasses. *Philos. Trans. R. Soc. B* 374, 20180245. doi: 10.1098/rstb.2018.0245

Xiao, Y., Liu, H., Wu, L., Warburton, M., and Yan, J. (2017). Genome-wide association studies in maize: praise and stargaze. *Mol. Plant* 10, 359–374. doi: 10.1016/j.molp.2016.12.008

Yan, P., Du, Q., Chen, H., Guo, Z., Wang, Z., Tang, J., et al. (2023). Biofortification of iron content by regulating a NAC transcription factor in maize. *Science* 382, 1159–1165. doi: 10.1126/science.adf3256

Yan, J., Kandianis, C. B., Harjes, C. E., Bai, L., Kim, E.-H., Yang, X., et al. (2010). Rare genetic variation at *Zea mays crtRB1* increases β-carotene in maize grain. *Nat. Genet.* 42, 322–327. doi: 10.1038/ng.551

Yang, L., Guo, S., Chen, Q., Chen, F., Yuan, L., and Mi, G. (2016). Use of the stable nitrogen isotope to reveal the source-sink regulation of nitrogen uptake and remobilization during grain filling phase in maize. *PloS One* 11, e0162201. doi: 10.1371/journal.pone.0162201

Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., et al. (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* 51, 1052–1059. doi: 10.1038/s41588-019-0427-6

Yin, P., Fu, X., Feng, H., Yang, Y., Xu, J., Zhang, X., et al. (2024). Linkage and association mapping in multi-parental populations reveal the genetic basis of carotenoid variation in maize kernels. *Plant Biotechnol. J.* doi: 10.1111/pbi.14346

Yu, H., Hao, Y., Li, M., Dong, L., Che, N., Wang, L., et al. (2022). Genetic architecture and candidate gene identification for grain size in bread wheat by GWAS. *Front. Plant Sci.* 13, 1072904. doi: 10.3389/fpls.2022.1072904

Zhai, H., Feng, Z., Du, X., Song, Y., Liu, X., Qi, Z., et al. (2018). A novel allele of *TaGW2-A1* is located in a finely mapped QTL that increases grain weight but decreases grain number in wheat (*Triticum aestivum L.*). *Theor. Appl. Genet.* 131, 539–553. doi: 10.1007/s00122-017-3017-y

Zhang, L., Fu, M., Li, W., Dong, Y., Zhou, Q., Wang, Q., et al. (2024). Genetic variation in *ZmKW1* contributes to kernel weight and size in dent corn and popcorn. *Plant Biotechnol. J.* doi: 10.1111/pbi.14279

Zhang, S., Ghatak, A., Bazargani, M. M., Bajaj, P., Varshney, R. K., Chaturvedi, P., et al. (2021). Spatial distribution of proteins and metabolites in developing wheat grain and their differential regulatory response during the grain filling process. *Plant J.* 107, 669–687. doi: 10.1111/tpj.15410

Zhang, S., Ghatak, A., Mohammad Bazargani, M., Kramml, H., Zang, F., Gao, S., et al. (2023b). Cell-type proteomic and metabolomic resolution of early and late grain filling stages of wheat endosperm. *Plant Biotechnol. J.* doi: 10.1111/pbi.14203

Zhang, X., Guo, W., Du, D., Pu, L., and Zhang, C. (2020). Overexpression of a maize BR transcription factor *ZmBZR1* in Arabidopsis enlarges organ and seed size of the transgenic plants. *Plant Sci.* 292, 110378. doi: 10.1016/j.plantsci.2019.110378

Zhang, Y., Li, D., Zhang, D., Zhao, X., Cao, X., Dong, L., et al. (2018). 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., 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, H., Luo, B., Liu, J., Jin, X., Zhang, H., Zhong, H., et al. (2023a). Functional analysis of *ZmG6PE* reveals its role in responses to low-phosphorus stress and regulation of grain yield in maize. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1286699

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, L., Zheng, Y., Wang, Y., Wang, S., Wang, T., Wang, C., et al. (2023). A *HST1-like* gene controls tiller angle through regulating endogenous auxin in common wheat. *Plant Biotechnol. J.* 21, 122–135. doi: 10.1111/pbi.13930

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

Zhu, M., Liu, T., and Guo, M. (2016). Current advances in the metabolomics study on lotus seeds. *Front. Plant Sci.* 7, 891. doi: 10.3389/fpls.2016.00891

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 July 2024
DOI 10.3389/fpls.2024.1425651

Check for updates

OPEN ACCESS

EDITED BY
Zhaorong Hu,
China Agricultural University, China

REVIEWED BY
Qibin Wu,
Chinese Academy of Tropical Agricultural Sciences, China
Yang Yang,
Shanxi Agricultural University, China
Rajesh Kumar Pathak,
Chung-Ang University, Republic of Korea

*CORRESPONDENCE
Xiaofang Xie
xxf317@fafu.edu.cn

RECEIVED 30 April 2024
ACCEPTED 12 July 2024
PUBLISHED 30 July 2024

CITATION
Chen R, Gu G, Zhang B, Du C, Lin X, Cai W, Zheng Y, Li T, Wang R and Xie X (2024) Genome-wide identification and expression analysis of the U-box E3 ubiquitin ligase gene family related to bacterial wilt resistance in tobacco (*Nicotiana tabacum* L.) and eggplant (*Solanum melongena* L.).
Front. Plant Sci. 15:1425651.
doi: 10.3389/fpls.2024.1425651

COPYRIGHT
© 2024 Chen, Gu, Zhang, Du, Lin, Cai, Zheng, Li, Wang and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Genome-wide identification and expression analysis of the U-box E3 ubiquitin ligase gene family related to bacterial wilt resistance in tobacco (*Nicotiana tabacum* L.) and eggplant (*Solanum melongena* L.)

Rui Chen[1], Gang Gu[2], Binghui Zhang[2], Chaofan Du[3], Xiaolu Lin[3], Weiwei Cai[3], Yan Zheng[1], Tong Li[1], Ruiqi Wang[1] and Xiaofang Xie[1,4]*

[1]College of Life Sciences, Fujian Agriculture & Forestry University, Fuzhou, China, [2]Institute of Tobacco Science, Fujian Provincial Tobacco Company, Fuzhou, China, [3]Longyan Tobacco Company, Longyan, China, [4]Fujian Key Laboratory of Crop Breeding by Design, Fujian Agriculture & Forestry University, Fuzhou, China

The E3 enzyme in the UPS pathway is a crucial factor for inhibiting substrate specificity. In Solanaceae, the U-box E3 ubiquitin ligase has a complex relationship with plant growth and development, and plays a pivotal role in responding to various biotic and abiotic stresses. The analysis of the *U-box* gene family in Solanaceae and its expression profile under different stresses holds significant implications. A total of 116 tobacco *NtU-boxs* and 56 eggplant *SmU-boxs* were identified based on their respective genome sequences. Phylogenetic analysis of *U-box* genes in tobacco, eggplant, tomato, *Arabidopsis*, pepper, and potato revealed five distinct subgroups (I-V). Gene structure and protein motifs analysis found a high degree of conservation in both exon/intron organization and protein motifs among tobacco and eggplant *U-box* genes especially the members within the same subfamily. A total of 15 pairs of segmental duplication and 1 gene pair of tandem duplication were identified in tobacco based on the analysis of gene duplication events, while 10 pairs of segmental duplication in eggplant. It is speculated that segmental duplication events are the primary driver for the expansion of the *U-box* gene family in both tobacco and eggplant. The promoters of *NtU-box* and *SmU-box* genes contained *cis*-regulatory elements associated with cellular development, phytohormones, environment stress, and photoresponsive elements. Transcriptomic data

Abbreviations: *Ras*, *Ralstonia solanacearum* L.; MW, Molecular weight; pI, Isoelectric points; ML, maximum likelihood; *NtU-box*, *U-box* genes of *Nicotiana tabacum*; *SmU-box*, *U-box* genes of *Solanum melongena*; FPKM, Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced; qRT-PCR, Quantitative real-time PCR.

analysis shows that the expression levels of the tobacco and eggplant *U-box* genes in different tissues and various abiotic stress conditions. Using cultivar Hongda of tobacco and cultivar Yanzhi of eggplant as materials, qRT-PCR analysis has revealed that 15 selected *NtU-box* genes and 8 *SmU-box* may play important roles in response to pathogen *Ras* invasion both in tobacco and eggplant.

KEYWORDS

***U-box*, biotic stress, expression analysis, phylogenetic analysis, *Nicotiana tabacum* L., *Solanum melongena* L.**

Introduction

The Ubiquitin-Proteasome System (UPS) is considered to be a major pathway of protein-specific degradation and plays an important role in the signal pathway of regulating environmental stresses in the post-translation stage of protein (Serrano et al., 2018; Stone, 2019; Sharma and Taganna, 2020). The system involves the coordinated catalytic activities of three types of enzymes, namely a large number of ubiquitin ligases (E3), together with a few ubiquitin activating enzymes (E1) and ubiquitin-conjugating enzymes (E2) (Yan et al., 2003; March and Farrona, 2018). In the process of ubiquitination, when ATP supplies energy, E1 activates the ubiquitin molecules and transmits them to E2, and E3 connects the ubiquitin binding E2 to the target protein, and finally achieves ubiquitination of the target protein (Finley and Chau, 1991; Pickart, 2001; Trenner et al., 2022). Ubiquitin ligases are crucial in the ubiquitin pathway as they specifically recognize the target proteins for ubiquitination, and it is also the most abundant enzyme in terms of quantity. Studies have shown that the ubiquitin ligases are classified into different families according to their structure, function, and substrate specificity (Kim et al., 2021a), but functional domains of four families (HECT, RING, U-box, and cullin) are common. U-box protein contains a 70-amino acid U-box domain, a single protein widely distributed in yeast, plants, and animals (Shu and Yang, 2017). The first plant elucidated U-box protein (PUB) family was in *Arabidopsis*, which contains 64 members (Azevedo et al., 2001; Trujillo, 2018). Subsequent studies have successively identified 82 members in wild emmer wheat (Yang et al., 2021a), 70 in *Salvia miltiorrhiza* (Pan et al., 2022), 77 in rice (Kim et al., 2021b), 121 in *Phyllostachys edulis* (Zhou et al., 2021), 67 in barley (Ryu et al., 2019) and 62 in tomato (Sharma and Taganna, 2020).

Numerous studies have shown that U-box proteins are involved in the regulation of plant hormone signal transduction, abiotic and biotic stress responses (Zeng et al., 2008). For example, *AtU-box18* and *AtU-box19* were found to coordinately function as regulatory components in development and stress response in *Arabidopsis* (Bergler and Hoth, 2011). Furthermore, *NtACRE276* was confirmed to have the E3 ligase activity and involved in cell death and defense signaling, and its ortholog in *Arabidopsis* (*AtU-box17*) and canola (*BnARC1*) showed similar biological function (Yang et al., 2006). In wheat, *TaPUB1* plays a key role in regulating the antioxidant capacity of diploid wheat under drought stress (Zhang et al., 2017), enhancing its resistance to powdery mildew fungi by controlling U-box proteins of CMPG1-V (Zhu et al., 2015). Similarly, in tomato, *SIU-box13* and *SIU-box40* were found to confer resistance against tomato yellow leaf curl virus (Sharma and Taganna, 2020).

Tobacco and eggplant are important crops. Extensive studies have demonstrated that *U-box* genes play vital roles in regulating diverse developmental processes and stress signaling in plants (Azevedo et al., 2001). Current research on members of the *U-box* gene family in tobacco and eggplant is limited, and their exact function is still unknown. Therefore, it is of great significance to systematically analyze the *U-box* gene family in tobacco and eggplant. The purpose of this study was to comprehensively analyze the *U-box* gene family by integrating transcriptome data of tobacco and eggplant and study the expression patterns of *U-box* gene family members under pathogen *Ralstonia solanacearum* L. *(Ras)* infection conditions. The results of this study lay an important foundation for further analysis of the function and trait improvement of the *U-box* gene family in tobacco and eggplant.

Materials and methods

Genome identification of *U-box* gene family members in two Solanaceae species

The genome sequence and annotation data of Solanaceae species, including tomato (ITAG2.4), eggplant (SME-HQ) and tobacco (Nitab-v4.5) were downloaded from the Sol Genomics Network (https://solgenomics.net/) (Fernandez-Pozo et al., 2015; Edwards et al., 2017). The local protein database of tobacco and eggplant was constructed by command 'makeblastdb' of the local BLAST tool (BLAST+ 2.13.0) and a total of 62 known tomato U-box protein sequences were used as seed sequences to align with the tobacco and eggplant protein sequences by BLASTP program

(Sharma and Taganna, 2020) U-box domain (PF04564) was obtained from the Pfam database (version 37.0) (http://pfam.xfam.org/) (Mistry et al., 2021).Sequences with the similarity ≥30% and E≤ $1e^{-10}$ were considered as the candidate proteins. Subsequently, the candidate protein sequences were further analyzed for the presence of a U-box structure domain (PF04564) using the CDD program of NCBI (https://www.ncbi.nlm.nih.gov/cdd/) and SMART tool (http://smart.embl-heidelberg.de/). The candidate protein containing the U-box conserved domain was confirmed as the final U-box protein. These *U-box* genes of tobacco and eggplant were renamed as *NtU-boxs* and *SmU-boxs*, respectively. The physicochemical properties of the tobacco and eggplant U-box proteins were predicted and analyzed using the ExPASy software (https://www.expasy.org) (Wilkins et al., 1999) and the subcellular location of each U-box protein analysis was based on the Cell PLoc 2.0 (Chou and Shen, 2008). The gene structure, conserved motif, phylogenetic tree, chromosomal localization, and synteny were analyzed, and corresponding flow chart was provided in Supplementary Figure S1.

Gene structure and conserved motif analysis

The GFF format file of gene structure for tobacco and eggplant was obtained from the Solanaceae genome database (https://solgenomics.net/) (Fernandez-Pozo et al., 2015; Edwards et al., 2017).The intron-exon gene structures of *NtU-box* and *SmU-box* genes were displayed using TBtools (version 2.097) (Chen et al., 2020) based on the gff3 files of the tobacco and eggplant genome. The conserved motifs of NtU-box and SmU-box proteins were analyzed using the online program MEME (https://meme-suite.org/meme/tools/meme) (Ma et al., 2014; Bailey et al., 2015) and the parameters were as follows: the number of motifs was set to 20, and the width range of motifs was set to be 5-200 amino acids respectively. Motif annotation was identified using the Pfam online tool (http://pfam-legacy.xfam.org/). *U-box* genes were submitted to the PlantCARE online program (Lescot et al., 2002) (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) for *cis*-acting elements prediction.

Multiple sequence alignment and phylogenetic analysis

To explore the evolutionary relationship of the *U-box* gene family in plants, the full protein sequences of U-box from tobacco, eggplant, pepper, potato, tomato, and *Arabidopsis* were analyzed. The Clustal X software (Thompson et al., 1997) was used to perform multiple sequence alignment. The phylogenetic tree was constructed by the MEGA-11 (Tamura et al., 2021) tool using the maximum likelihood method (ML) with a bootstrap of 1000 replications. The ML is a significant statistical method for parameter estimation. The ITOL (version 6.0) tool (https://itol.embl.de/) was used to edit the phylogenetic tree of NtU-box proteins and SmU-box proteins.

Chromosomal localization and synteny analysis

Based on the annotation information and the full genome protein sequences of tobacco (Nitab-v4.5), eggplant (SME-HQ), tomato (ITAG2.4) and *Arabidopsis* (TAIR10), the MCScanX software (Wang et al., 2012) with default parameters was employed to analyze the possible segmental duplication, tandem duplication events, intra-genomic syntenic and inter-genomic collinearity blocks (Edwards et al., 2017) and the TBtools software (version 2.097) (Chen et al., 2020) was used for visualization.

Expression analysis of *NtU-box* and *SmU-box* genes

To investigate the expression patterns *U-box* genes across different tissues and under various abiotic stresses, we analyzed the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values of *NtU-box* genes in different tissues (GSE233199) (Mo et al., 2023) under drought conditions (GSE214048) (Hu et al., 2022), as well as *SmU-box* genes in different tissues (PRJNA328564) (Barchi et al., 2019) and under high temperature conditions (Liu et al., 2023), which were download from NCBI(https://www.ncbi.nlm.nih.gov/geo/). A map was generated using the heatmap function of the R gplots package (Walter et al., 2015).

The Hongda variety of tobacco and Yanzhi variety of eggplant were cultivated using conventional cultivation methods. The seedlings were managed until the 3-5 leaf stage. A total of 100 tobacco seedlings and 100 eggplant seedlings were selected and inoculated with a highly efficient strain of *Ralstonia solanacearum* L. (*Ras*) that had been isolated and maintained by our laboratory (Gao et al., 2019). These plants were cultured in a greenhouse with high-humidity and high-temperature (approximately 80% humidity, 28~30°C, 14 h light exposure; 10 h dark environment). The seedlings of tobacco and eggplant were collected at 0 h, 12 h, 24 h, 48 h and 96 h after inoculation, with each biological sample consisting of 5 plants and a total of 3 replicates. For sampling, seedlings were uprooted and their roots were quickly washed with sterile water to remove any attached soil and pathogens. The cDNA synthesis was carried out using the SMART Kit (Takara). To evaluate the expression levels of the *NtU-box* and *SmU-box* genes, real-time quantitative PCR (qRT-PCR) was conducted using SYBR Green qPCR Premix (Universal), and the relative expression levels were calculated using the $2^{-\Delta\Delta t}$ method (Livak and Schmittgen, 2001). Three technical replicates were performed for each sample.

The actin genes of both tobacco and eggplant were used as the internal reference gene, and the primers of *NtU-box* and *SmU-box* genes (Supplementary Table S1) were designed using primer3 software (https://bioinfo.ut.ee/primer3-0.4.0/).

Results

Characterization and distribution of *U-box* genes in tobacco and eggplant genomes

In this study, a total of 116 *U-box* genes were identified in tobacco, while 56 *U-box* genes were in eggplant. These genes were renamed from *NtU-box1* to *NtU-box116* and *SmU-box1* to *SmU-box56*, respectively. To characterize these *NtU-box* and *SmU-box* genes, gene ID and protein molecular weight (MW), theoretical isoelectric point (pI), subcellular localization, number of exon, and CDS sequences were analyzed (Supplementary Table S2). In tobacco, *NtU-box* genes contained 1 to 19 exons, and the relative molecular weight of their corresponding proteins varied greatly from 20890.72 Da (*NtU-box112*) to 194534.46 Da (*NtU-box31*), and the theoretical isoelectric point ranged from 5.12 (*NtU-box47*) to 9.47 (*NtU-box110*). In eggplant, *SmU-box* genes contained 1 to 18 exons. The predicted molecular weight ranged from 31784.29 Da *(SmU-box53)* to 166473.85 Da *(SmU-box43)*, while the theoretical isoelectric point varied from 4.92 *(SmU-box51)* to 9.23 *(SmU-box9)*. Subcellular localization analysis of NtU-box and SmU-box proteins showed that these proteins were mainly present in the nucleus. Among the 116 NtU-box proteins, 5 were located on the cytoplasm, 4 on the cell membrane, and 1 on the chloroplast, while other NtU-box proteins were located on the nucleus. Among the 56 SmU-box proteins, 3 were located in the cytoplasm, 1 was located in the chloroplast, and the other SmU-box proteins were located in the nucleus (Supplementary Table S2).

Chromosome localization and collinearity analysis of *NtU-box* and *SmU-box* genes

The analysis of chromosomal localization showed that some *NtU-box* genes could not acquire the precise location information due to the incomplete sequencing of the tobacco genome. Among the 116 *NtU-box* genes, a total of 64 genes were unevenly distributed on 24 chromosomes of tobacco, while the remaining 52 *NtU-box* genes were mapped to unassigned scaffolds (Figure 1A). Notably, chromosome 19 contained the largest number of *NtU-boxs* (8 genes), followed by chromosome 04 with 7 *NtU-box* genes, and chromosomes 13 and 14 with 6 *NtU-box* genes each. In contrast, chromosomes 03, 05, 11, 20, 21, and 24 each contained only one *NtU-box* gene, while chromosomes 01, 02, 08 and 23 had no *NtU-box* genes detected (Figure 1A). In addition, 1 pair of tandem duplication genes on chromosome 19 (*NtU-box14/10*) and 15 pairs of segmental duplication genes were identified in the tobacco *NtU-box* gene family (Figure 1A, Supplementary Table S3). The result of chromosomal location analysis revealed that 55 out of 56 *SmU-box* genes were unevenly distributed among the 12 chromosomes of eggplant, with only 1 *SmU-box* gene located on an unattributed scaffold. Chromosome 01 contained the largest number of *SmU-boxs* (12 genes), followed by chromosomes 03 and 11 with 7 *SmU-box* genes, and 5 *SmU-box* genes for chromosomes 04, 05, 09 and 12. However, chromosomes 02, 06, 07, and 10 contained 4, 3, 1, and 1 *SmU-box* genes, respectively, while none of the *SmU-box* genes were detected on chromosome 08 (Figure 1B). In the eggplant *SmU-box* gene family, 10 pairs of fragment duplication genes were identified, but no tandem repeats were identified (Figure 1B; Supplementary Table S3).

A total of 22 orthologous genes were identified between tobacco and *Arabidopsis* based on the interspecies syntenic analysis, while there were 65 syntenic counterparts between tobacco and tomato (Figures 2A, B). A total of 68 orthologous genes were identified

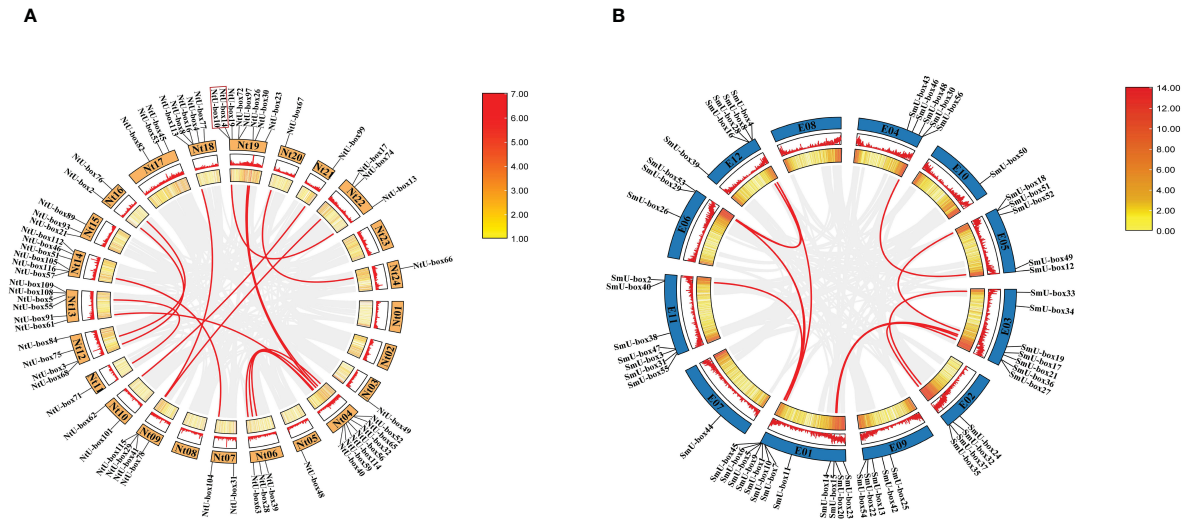


FIGURE 1
Gene distribution and duplication **(A)** *NtU-box* genes on the 24 chromosomes in tobacco **(B)** *SmU-box* genes on the 12 chromosomes in eggplant. Tandem-duplicated gene pairs of Nt19 are marked with red box, and segmental duplication genes are connected by red line.

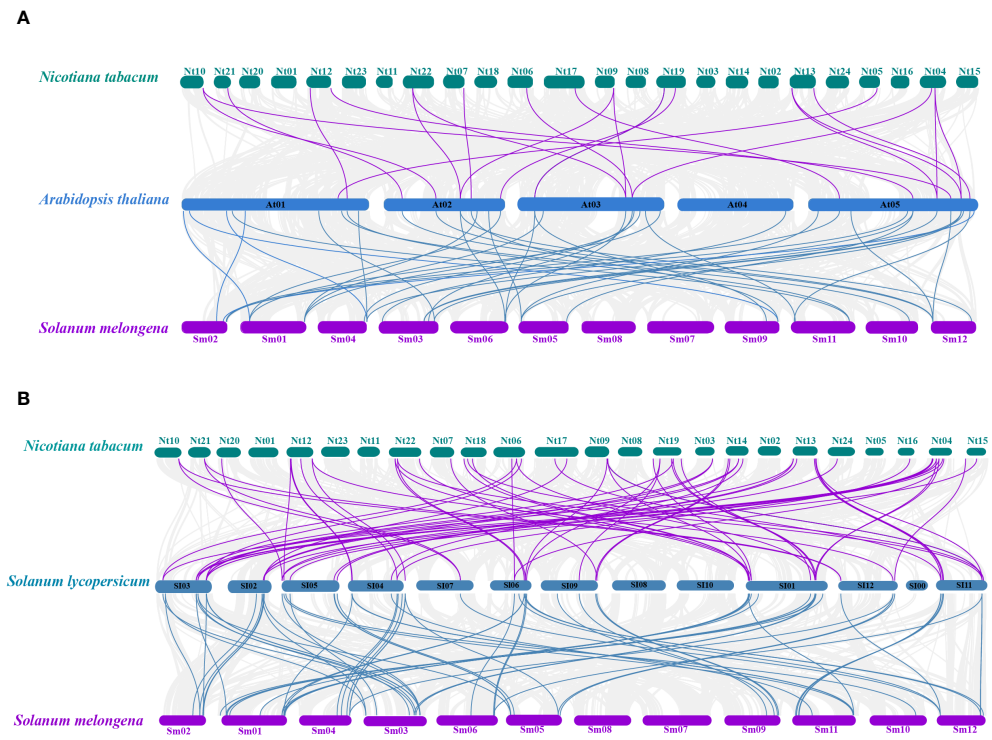


FIGURE 2
Collinearity analyses of *U-box* genes among tobacco, eggplant tomato and *Arabidopsis*. **(A)** Collinearity analyses among *Arabidopsis thaliana*, *Nicotiana tabacum* and *Solanum melongena*. **(B)** Collinearity analyses among *Solanum Lycopersicon*, *Nicotiana tabacum* and *Solanum melongena*. The gray line represents the co-collinearity of all genes among the three species, and the purple and blue line represents the collinearity among members of the *U-box* gene family.

between eggplant and *Arabidopsis* based on the interspecies syntenic analysis, while there are 41 syntenic counterparts between eggplant and tomato (Figures 2A, B). The genomic regions around *NtU-box41/61/62/74*, *SmU-box9/15/29/39/40* showed strong syntenic relationships with their counterparts in both *Arabidopsis* and tomato (Supplementary Table S4). Notably, good collinearity was detected among the *U-box* genes of four distinct species, even after undergoing speciation and long-term evolution, and the result suggested that these genes might have originated before Solanaceae species diversification and retained conserved functional roles.

Phylogenetic and gene structure analysis of *NtU-box* and *SmU-box* genes

To investigate the evolutionary relationship between *NtU-box* and *SmU-box* genes, a phylogenetic tree was constructed (Figures 3A, 4A). In tobacco, *NtU-box* genes were divided into 5 subgroups (I ~ V), with the largest members (41 members) in subgroup I. The subgroup represented more than 35.3% of the total *NtU-box* members. In contrast, subgroups II, IV, and V had only 6, 9 and 21 members, respectively. In eggplant, the largest members of eggplant (18 members) found in the subgroup I and this subgroup represented more than 32.1% of the total *SmU-box* members. In contrast, subgroups II, III, and V had only contained 3, 5 and 10 members, respectively. Gene structure of *NtU-boxs* were found that the number of exons varied from 1 (*NtU-box1*) to 19 (*NtU-box105*) (Figure 3B).

The number of exons *SmU-box* gene ranging from 1 to 18. Among 116 *U-box* genes in tobacco, *NtU-box105* and *NtU-box106* contained the greatest number of exons (19), while 40 *NtU-box* genes (34.5%) only contained one exon (Figure 3B). Among 56 *U-box* genes in eggplant, 21 *SmU-box* genes (37.5%) contained one exon (Figure 4B). In addition, the *U-box* genes with similar gene structures were clustered into the same sub-clade. For example, most tobacco members of Group II only housed five exons. This result indicated that the members of the same groups exhibited similar gene structures.

Domain and motif analysis of the NtU-box and SmU-box proteins

A total of 20 conserved motifs have been identified in the 116 *NtU-box* and 56 *SmU-box* genes. The lengths and conserved sequences of each motif are listed in Supplementary Table S5. Among them, Motif 7 and Motif 1 were prevalent across most genes in all five groups of tobacco NtU-box proteins (Figure 5). Similarly, in eggplant SmU-box proteins Motif 1, Motif 4, Motif 6, and Motif 5 were found in most genes of the five groups, indicating their high conservation in U-box proteins (Figure 6). The protein sequences of the 20 motifs were uploaded to the CDD program for further domain analysis. Motif 1, Motif 2, and Motif 7 were

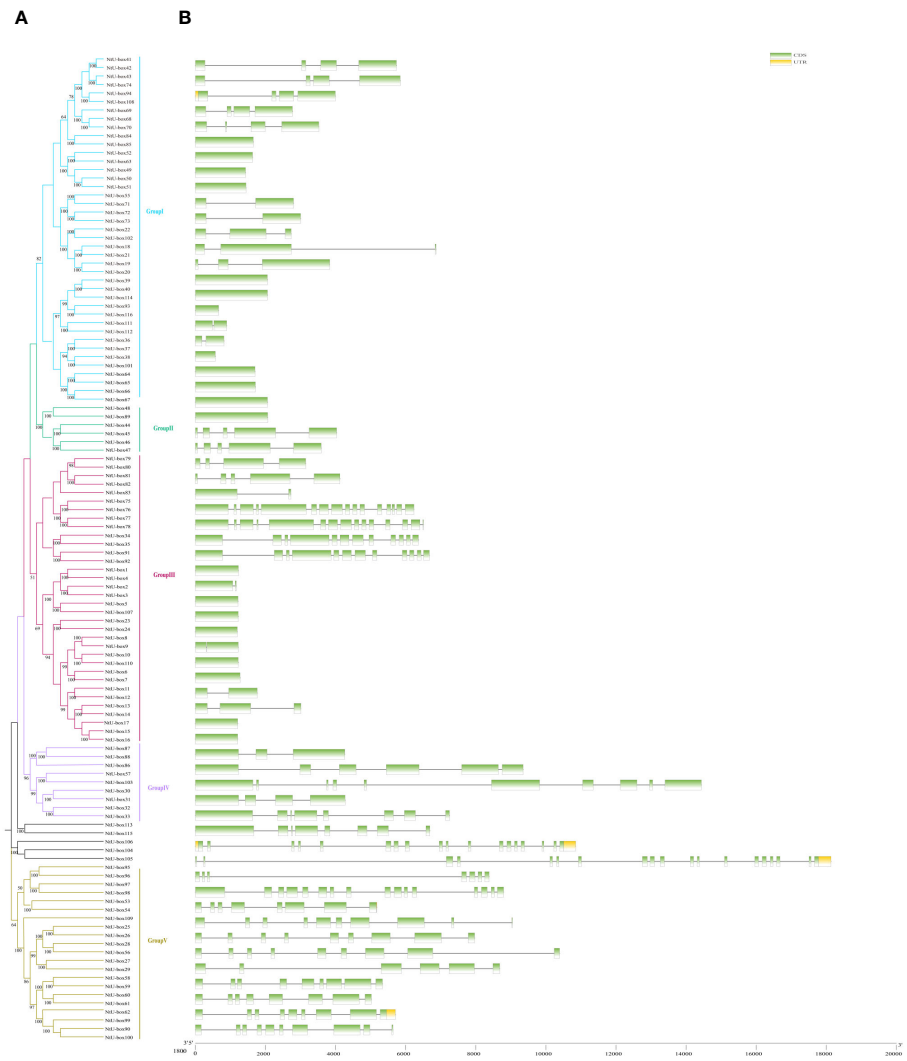


FIGURE 3
Gene structure and evolution of U-box family in tobacco **(A)** Phylogenetic relationships of *NtU-boxs*. Different subgroups were marked with different colors. **(B)** Intron-exon structure of *NtU-boxs*. UTR and CDS are represented by different colors. The scale bar of bottom demonstrates the length of exons and introns.

annotated as components of the conserved U-box domain sequences, essential for maintaining the structural integrity of the U-box and facilitating ubiquitin linkage activity. Additionally, Motif 3 was annotated as part of the ARM conserved domain, which was the most common type in the U-box family. Moreover, genes in the same group on the phylogenetic tree exhibited similar conserved motifs. For example, all *NtU-box* genes in Group II had the same 7 motifs (Motifs 1, 2, 7, 9, 11, 15, and 20), suggesting potential functional similarities.

Phylogenetic analysis and functional prediction of the *U-box* gene family in Solananceae species

To explore the evolution of the *U-box* gene family in plant, 428 *U-box* gene members from 6 species were selected to construct a phylogenetic tree (Figure 7), including tobacco (116), eggplant (56), potato (66), tomato (62), *Arabidopsis thaliana* (64), and pepper (64) (Supplementary Table S6). Based on previous study, 428 *U-box* genes were divided into 5 different subfamilies (Wang et al., 2021). Based on the phylogenetic tree, a total of 138 sister pairs of homologous proteins were identified, including 71 pairs of orthologous genes and 67 pairs of paralogous genes (Supplementary Table S7). Specifically, there were 41 paralogous pairs from tobacco, 17 pairs from *Arabidopsis thaliana*, 6 pairs from pepper, 2 pairs from potato, and 1 pair from eggplant.

Similarity in gene expression patterns implies similar functions, especially for homologous genes (Yang et al., 2021b). Based on published transcriptomic data, the expression patterns of *U-box* genes in major tissues were compared in tobacco (116), eggplant (56), and *Arabidopsis thaliana* (64). These *NtU-box* and *SmU-box* genes were classified into 7 categories according to their normalized expression levels (Supplementary Figure S2A). Among them, the *U-*

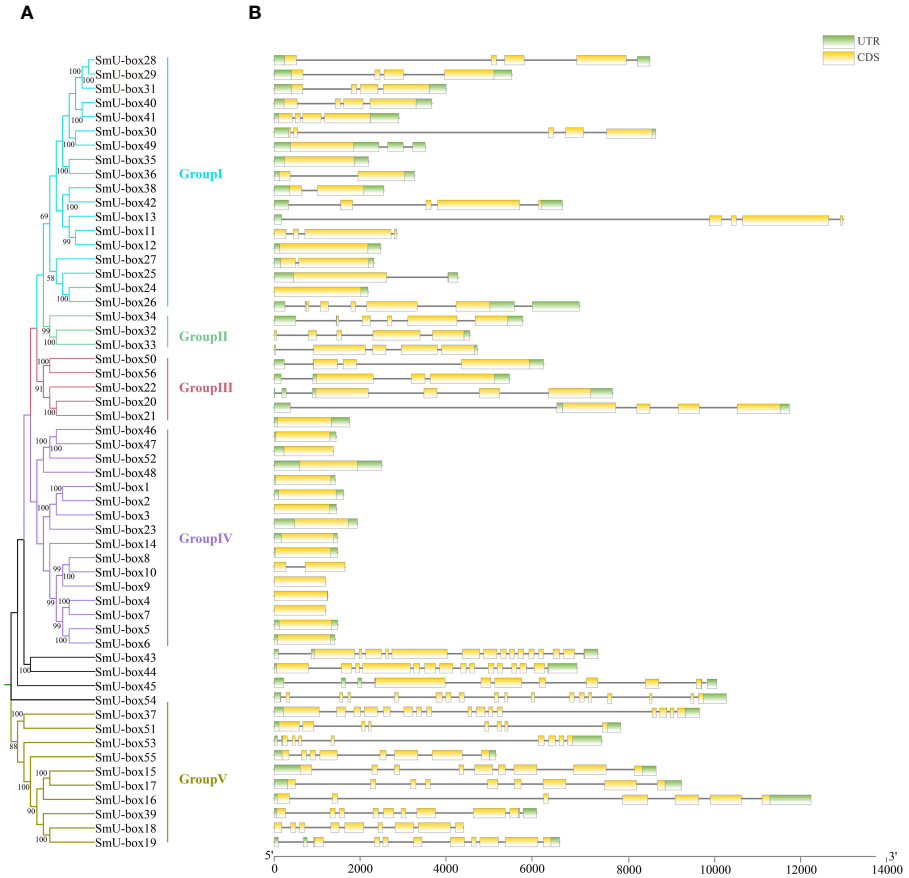


FIGURE 4
Gene structure and evolution of U-box family in eggplant. **(A)** Phylogenetic relationships of *SmU-boxs*. Different subgroups were marked with different colors. **(B)** Intron-exon structure of *SmU-boxs*. UTR and CDS are represented by different colors. The scale bar of bottom demonstrates the length of exons and introns.

box genes of group I specifically expressed in leaf with a high level; the *U-box* genes of group II expressed specifically high in all tissue; *U-box* genes in Cluster III mostly expressed in stem; *U-box* in group IV and V highly expressed in roots. Based on the tissue expression clustering characteristics, phylogenetic relationship of multispecies *U-box* genes were predicted (Supplementary Table S8).

In general, the functions of some *NtU-box* and *SmU-box* genes were mainly divided into the following four categories: 1. Play a role in formation of organs and regulator of flowering time. 2. Play a role in regulating root development, pollen tapetum development and ROS induced chloroplast degradation. 3. Play combinatory roles in response to drought stress. 4. Encodes a U-box domain-containing E3 ubiquitin ligase with central Ser/Thr protein kinase domain, and its expression is responsive to both phosphate (Pi) and phosphite (Phi) in both roots and shoots (Supplementary Table S8). Only those *NtU-box* and *SmU-box* genes that were homologous to the reported *AtU-box* genes, had highly similar expression patterns, and shared similar functions. Fox example, *AtPUB18* and *AtPUB19* function as regulators in the drought stress response (Liu et al., 2011) and their homologous genes *NtU-box65/66/67* genes and *SmU-box27* genes, which had similar expression patterns, were predicted to have similar functions (Supplementary Figure S2B). Transcriptome data showed that *NtU-box65/66/67* genes and *SmU-box27* genes showed a positive regulatory expression pattern under heat stress (Supplementary Table S9). *AtPUB14* plays a role in organ formation and flowering time regulation (Feke et al., 2020), and its homologous gene *SmU-box28/29/31* has a similar expression pattern in eggplant (Supplementary Figure S2C). Transcriptome data showed that *SmU-box28/29/31* has a high expression level in eggplant flowers, which is expected to have similar functions with *AtPUB14*.

Cis-regulatory elements analysis and tissue expression patterns of *U-box* genes

To investigate the potential function of *U-box* genes during plant development and upon exposure to various stresses, the *cis*-elements within the promoter regions of *U-box* genes were analyzed. A total of 36 *cis*-regulatory genes were identified in the promoter region of *NtU-box* genes, while 34 *cis*-regulatory genes were identified in the promoter region of *SmU-box* genes (Figure 8, Supplementary Table S10). These *cis*-regulatory genes can be divided into four categories, specifically cell development, plant hormones, environmental stress, and photoresponse elements (Supplementary Table S10). The 5 *cis*-acting elements related to

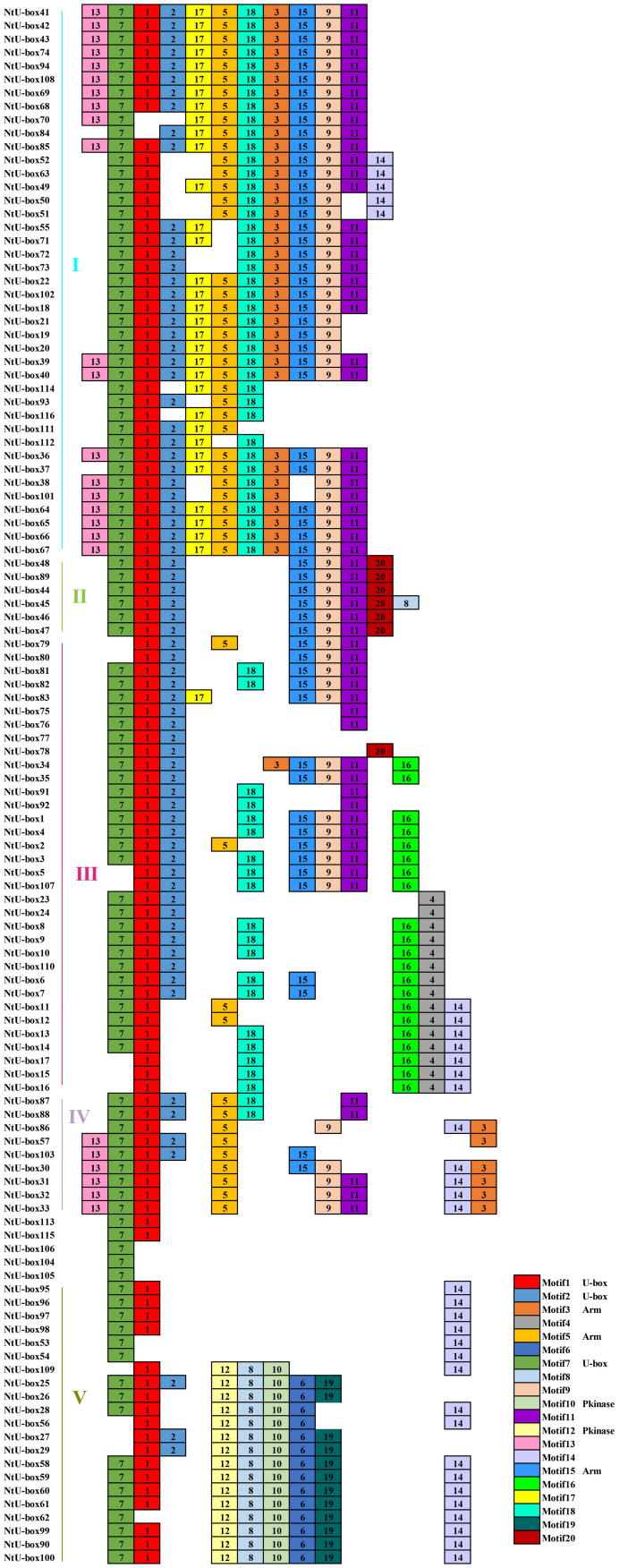


FIGURE 5
Conserved motifs for U-box proteins in tobacco. Different motifs are showed with different colored boxes and numbers (1-20).

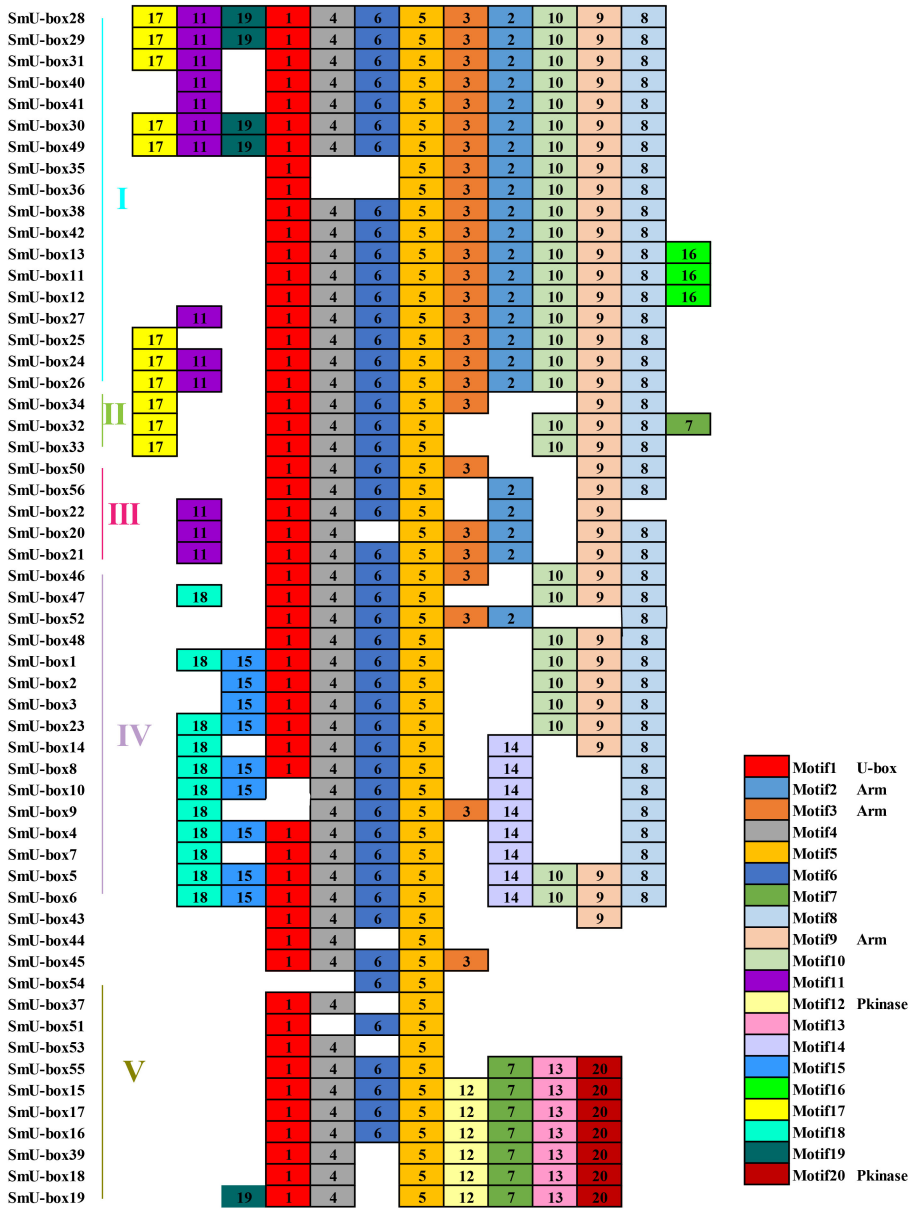


FIGURE 6
Conserved motifs for U-box proteins in eggplant. Different motifs are showed with different colored boxes and numbers (1-20).

cell development include CAT-box, MSA-like, CCAAT-box, MBSI, and HD-Zip 1. For phytohormone-responsive elements, the *cis*-acting elements include TGACG-motif, ABRE, P-box, TGA-element, TCA-element, AuxRR-core, TATC-box, GARE-motif, AuxRE, A-box, and O2-site. Additionally, ten light responsive elements were identified, including GT1-motif, G-Box, Box4, MRE, ATC-motif, Sp1, ATCT-motif, ACE, 3-AF1 binding site, and AAAC-motif. The expression of these genes might be regulated by phytohormones, various light-responsiveness *cis*-elements, defense signaling transduction, and abiotic stresses during the growth of tobacco and eggplant.

Transcriptome analysis revealed diverse expression patterns of *NtU-box* and *SmU-box* genes in different tissues, which were clustered into three groups (Figure 9) (Yang et al., 2022). In tobacco (Figure 9A), group I comprised 41 *NtU-box* genes, with the majority showing high levels of expression in vegetative tissue, particularly in root. Conversely, *NtU-box* genes in groups II and III exhibited low levels of expression. A total of 5 *SmU-box* genes (*SmU-box* 54/42/53/34/41) in group II showed high levels of expression in all investigated tissues, including stem, leaf, radicle, cotyledons, root, and flower (Figure 9B), and 31*SmU-box* genes in group I exhibited moderate levels of expression. Conversely, *SmU-*

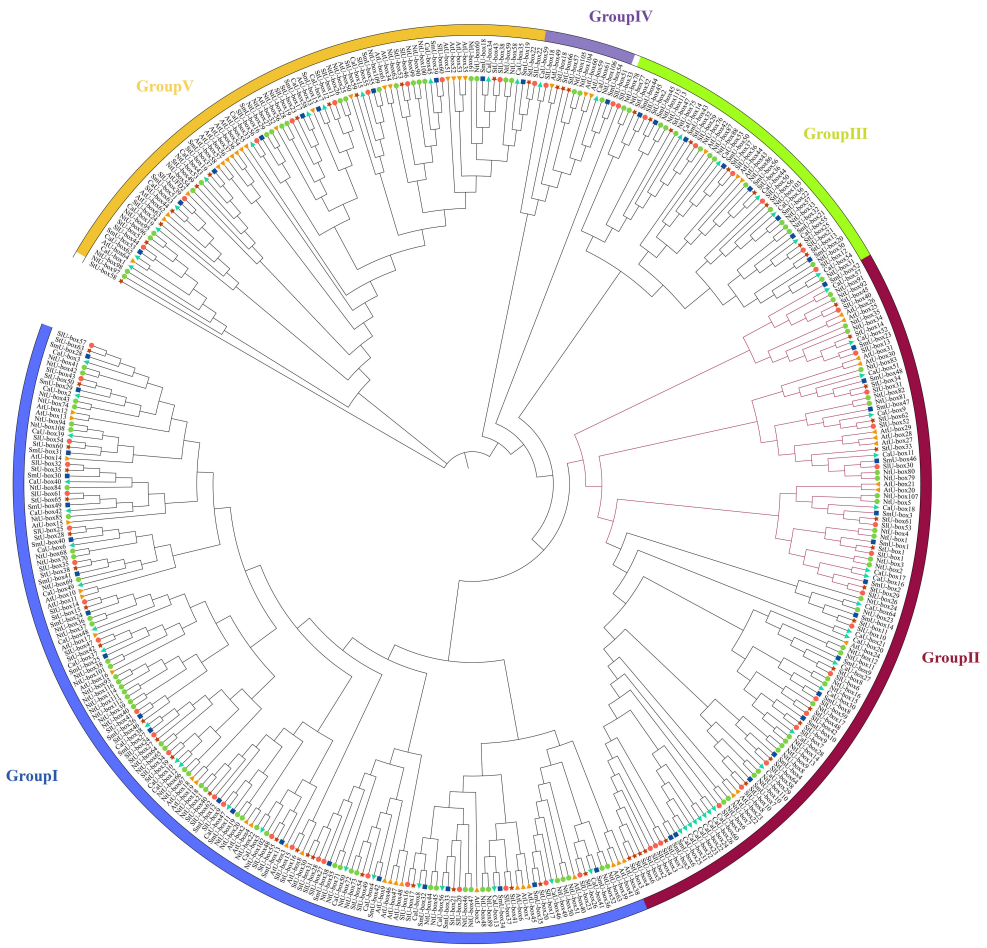


FIGURE 7
Phylogenetic relationship of 116 NtU-box and 56 SmU-box proteins, along with another 256 published U-box proteins. The phylogenetic relationships were generated by using MEGA-11 using the Maximum Likelihood (ML) method (1000 bootstrap replicates), and visualized with ITOL software. U-box proteins were classified into five distinct groups, as indicated by the different colors.

box genes in group III showed nearly negligible expression, exhibited lower expression levels.

The expression profiles of tobacco and eggplant *U-box* genes under various abiotic and biotic stresses

To determine the expression profiles of *U-box* family genes under abiotic stress, the FPKM data of tobacco *NtU-box* genes under drought stress (Hu et al., 2022) and eggplant *SmU-box* genes under high temperature (Liu et al., 2023) were downloaded. The expression profiles of 62 *NtU-box* genes and 26 *SmU-box* genes were analyzed (Supplementary Table S9). The results revealed distinct expression patterns among the *NtU-box* and *SmU-box* gene members under various stress conditions (Figure 10). Specifically, the 62 *NtU-box* genes and 56 *SmU-box* genes were each classified into three groups (I ~ III). In tobacco, a total of 15 *NtU-box* genes were included in group I, and the high expression levels of these genes in the five drought stress stages mean that these genes can play an important role in the drought stress process, while the 32 *NtU-box* genes clustered in group II showed low or no expression in the whole drought stress process. In eggplant, a total of 14 *SmU-box* genes were included in group III, and the expression profiles of these genes exhibited a pattern of initial decrease followed by an increase across the three high-temperature stress stages, indicating their potential significance in responding to high-temperature stress. Conversely, the 6 *SmU-box* genes in group I showed either minimal or no expression throughout the high-temperature stress process. It is worth noting that the expression levels of *NtU-box59* and *NtU-box46* genes gradually decreased with the increase of drought time. In terms of high-temperature stress (Figure 10B), the expression level of the *SmU-box27* gene in group II gradually increased with the increase of high-temperature time, while the majority of *SmU-box* genes (*SmU-box54/32/22/38/39*) reached the peak expression level at the 12 h of high-temperature stress. These results indicated the functional diversity of *U-box* members among different species.

No significant changes were observed in the seedlings of both tobacco and eggplant at the initial stage after being infected by *Ralstonia solanacearum* L. (*Ras*). In tobacco, the primary symptoms induced by *Ras* infection manifested in the seedling at 96 h

FIGURE 8
(A) Predicted *cis*-elements in *NtU-box* promoters. **(B)** Predicted *cis*-elements in *SmU-box* promoters. Different shapes and colors represent the different types of *cis*-elements. Annotations of *cis*-elements were listed in Supplementary Table S10.

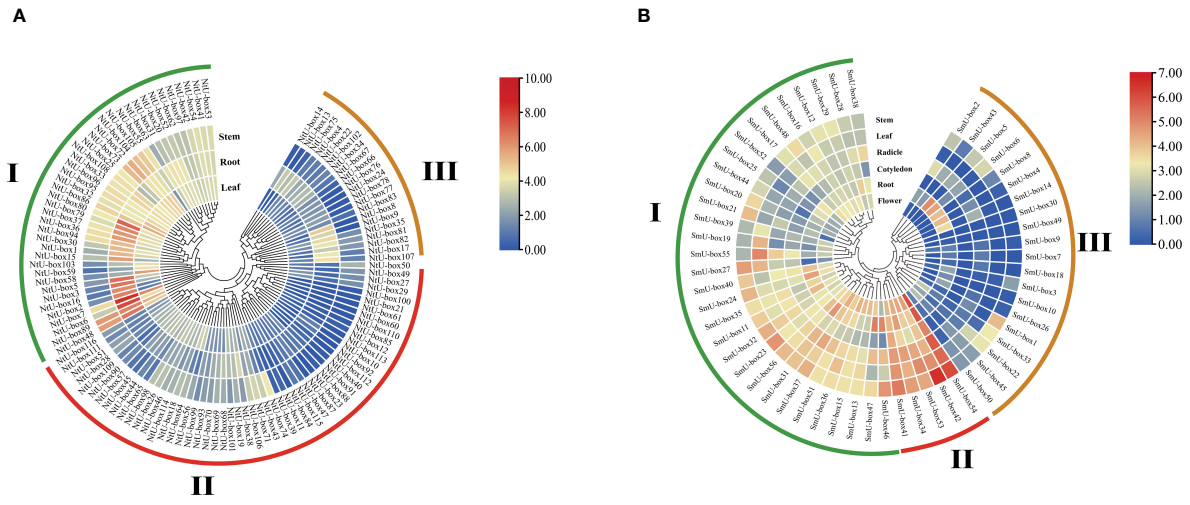


FIGURE 9
The expression patterns of tobacco and eggplant *U-box* genes in different tissues. **(A)** The expression patterns of *NtU-box* genes **(B)** The expression patterns of *SmU-box* genes. FPKM values for *U-box* genes were transformed by log2 (n+1).

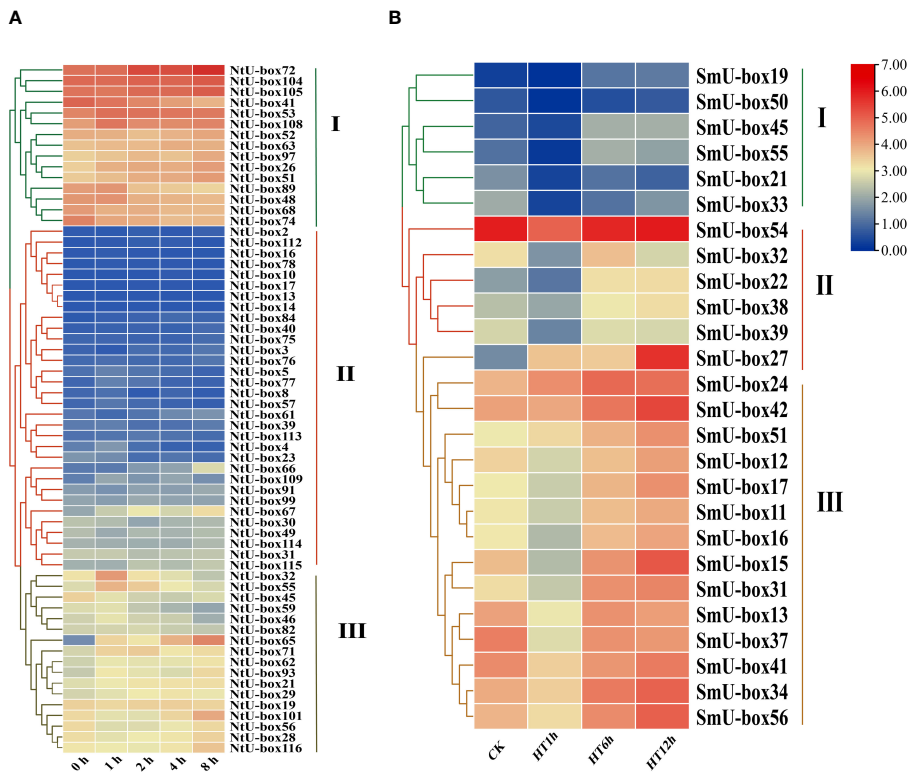


FIGURE 10
The expression profiles of tobacco and eggplant *U-box* genes under abiotic stresses. **(A)** The expression patterns of *NtU-box* genes under drought stress. **(B)** The expression patterns of *SmU-box* genes under high temperature. FPKM values for *U-box* genes were transformed by log2 (n+1).

(Figure 11A). During this stage, the seedling exhibiting leaf wilting and stem necrosis, with the roots turning yellow and exhibiting necrosis. In contrast, symptoms in the eggplant appeared later than those in tobacco, with notable symptoms appeared at 120 h (Figure 11B). The leaves appeared withered, and the basal part of the stem showed severe necrosis and turned black. It was reported that *SIU-box13* and *SIU-box40* in tomato, play a crucial role in the regulation of pathogen invasion (Sharma and Taganna, 2020). To further explore the potential functions of the *U-box* genes in tobacco and eggplant, a total of 15 *NtU-box* and 8 *SmU-box* genes that clustered with *SIU-box13* and *SIU-box40* in II subgroups of the phylogenetic tree (Figure 7) were selected for qRT-PCR analysis under *Ras* infection. The qRT-PCR analysis found that the majority of the selected genes showed significant response to *Ras* infection. In tobacco (Figure 12A), the expression levels of 7 *NtU-box* genes (*NtU-box1/3/79/81/82/83/91*) exhibited a trend of initial increased followed by decreased with the extension of time after inoculation. Among these genes, the expression levels of 5 genes (*NtU-box3/79/81/82/83*) were significantly up-regulated at 12 h after inoculation, while 2 genes (*NtU-box2/4*) displayed significant down-regulated and 6 genes (*NtU-box5/34/35/80/92/107*) showed a pattern of decrease followed by a slight increase. In eggplant (Figure 12B), all the 8 selected genes displayed up-regulated expression in response to the infection compared to the initial stage (0 h), and 2 genes (*SmU-box1/2*) exhibited significant up-regulation exceeding a 4-fold increase at 24 h post-inoculation.

Discussion

U-box gene family members are present in almost all eukaryotes. The advent of genome sequencing has facilitated comprehensive analyses of the *U-box* gene family in numerous species, such as *Arabidopsis thaliana* (64), rice (77), soybeans (125), and tomato (62) (Wiborg et al., 2008; Zeng et al., 2008; Wang et al., 2016; Sharma and Taganna, 2020). In this study, a total of 116 *U-box* genes were identified in tobacco, which were larger than those in *Arabidopsis thaliana* and rice, but less than in soybean. As an allotetraploid, the genome size of *Nicotiana tabacum* is 4.5Gb, while that of Arabidopsis, rice and soybean is 125Mb, 430Mb and 1.025Gb, respectively (Kaul S et al., 2000; Burr, 2002; Schnable et al., 2009; Schmutz et al., 2010). In this case, it seems that there is no direct correlation between the number of *U-box* genes and genome size in these plants. The number of exons observed between *NtU-box* genes varies significantly, of which 64 genes contain two or more exons, 13 genes have more than 10 exons, and 39 genes possess only one exon. Similarly, for *SmU-box* genes, differences are notable: 32 genes contain two or more exons and 4 genes have more than 10 exons, and however, 20 genes possess only one exon. This suggested that distinct RNA splicing processes that could modulate the proportion of isoforms (active and non-active), crucial for stress regulation (Tang et al., 2021). Many *SmU-box* genes are either intron-less or have two introns, consistent with reports in tomato (Sharma and Taganna, 2020). *U-box* genes with multiple introns

FIGURE 11
Disease symptoms of tobacco and eggplant seedlings. **(A)** Tobacco seedlings at 0 h and 96 h post-inoculation with *Ras*. **(B)** Eggplant seedlings at 0 h and 120 h post inoculation with *Ras*. The basal parts of stems were magnified and shown in the circles.

may act as a mutational buffer, protecting coding sequences from deleterious mutations. The presence of the intron-less genes underscores structural integrity within the *U-box* gene family, suggesting irregular distribution of U-box protein in different species (Zeng et al., 2008). The phylogenetic analysis of the *U-box* gene family showed a great similarity among all the five classes due to the presence of the core U-box domain in all the members (Figure 7), suggesting a shared ancestor predating the divergence of these species. Moreover, a single tandem duplication gene pair and 15 segmental duplication gene pairs were detected in tobacco (Figure 1A), while 10 segmental duplication gene pairs were identified in eggplant (Figure 1B), highlighting segmental duplication events as the primary driver for the expansion of the *U-box* gene family in both tobacco and eggplant. In addition, a good collinearity was detected among the *U-box* genes of four distinct species, even after undergoing speciation and long-term evolution. Syntenic analysis revealed a higher number of orthologous gene pairs between tobacco and tomato compared to those between tobacco and *Arabidopsis* (Figures 2A, B). Conversely, the count of orthologous gene pairs between eggplant and *Arabidopsis* exceeded those between eggplant and tomato (Figures 2A, B). The results suggested that species with close evolutionary relationships tend to exhibit greater similarity, higher homology, and increased conservation of the *U-box* genes.

In general, the evolution of gene family is predominantly determined by the organization of gene structures, whereas within a gene family, members of the same subfamily typically exhibit high conservation in both structure and function, reflecting their evolutionary relatedness. In this study, a total of 20 conserved motifs were identified in the *U-box* gene families of both tobacco and eggplant. Interestingly, despite belonging to the same or different subfamilies, *U-box* members exhibited variations in motif types and quantities. However, the differences in the same subfamily were notably smaller, indicating a higher level of conservation in motif composition within closely related members. These observations highlight the complexity of the tobacco and eggplant genomes and the differentiation and diversity of the function within the *U-box* gene family.

U-box genes play important roles in plant responses to both abiotic and biotic stresses (Adler et al., 2017; Sharma and Taganna, 2020; Tang et al., 2021). Functional analysis has revealed the molecular mechanisms involving U-box proteins involving U-box proteins in theses stress response (Cho et al., 2017). However, the role of the *U-box* gene family in tobacco and eggplant remains unclear. Stress refers to the phenomenon that plants are exposed to adverse conditions in their environment, causing their physiological processes to be negatively affected. Stress is usually divided into two categories: abiotic stress and biotic stress. Abiotic stress mainly refers to the unfavorable conditions caused by environmental factors, such as temperature, light, humidity, drought, salt and alkali. Biotic stresses include a range of biological factors that are harmful to plant survival and development, often stemming from infections and competition, including diseases, pests, and weeds (Atkinson and Urwin, 2012). These stress factors significantly affect crop growth and production. Therefore, mining excellent resistance genes in plants has become one of the main strategies to cope with various stress challenges. In this study, 116 *NtU-boxs* were identified in tobacco and 56 *SmU-boxs* were identified in eggplant, and their expressions under abiotic and biotic stresses were further analyzed. Transcriptomic data analysis revealed the different responses of

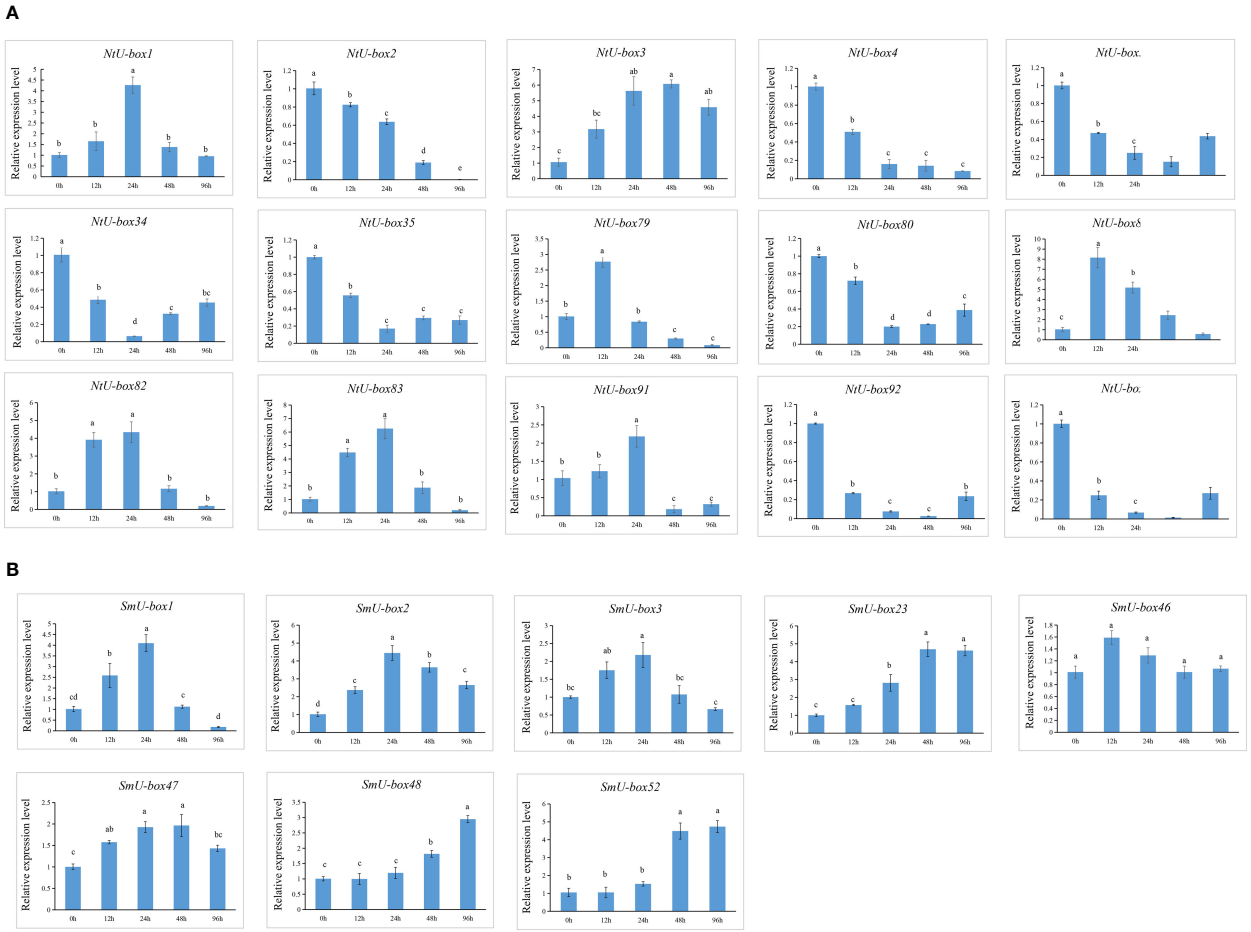


FIGURE 12
Expression profiles of *U-box* genes in response to *Ras* infection. **(A)** Relative expression level of 15 *NtU-boxs* in response to *Ras* infection. **(B)** Relative expression level of 8 *SmU-boxs* in response to *Ras* infection. Error bars are standard deviations of three biological replicates. Diverse letters on top of the bars indicate significant differences at $P<0.05$.

plant *U-box* gene family members to abiotic stress, indicating functional differences within the family. In addition, *SIU-box13* and *SIU-box40* of tomato have been shown to play a key role in regulating pathogen invasion (Sharma and Taganna, 2020). Therefore, it can be inferred that the *NtU-box* and *SmU-box* genes clustered in the same subfamily with *SIU-box13* and *SIU-box40* may have similar functions. In this study, the expression patterns of 15 selected *NtU-box* genes and 8 *SmU-box* genes under *Ras* infection highlighting their important role in *Ras* resistance. Among the 15 *NtU-box* genes, the expression of 7 *NtU-box* genes (*NtU-box1/3/79/81/82/83/91*) was significantly up-regulated, and the *NtU-box81* gene showed a strong response to *Ras* invasion, which increased more than 8 times at 12 h post-inoculation, while the remaining 8 *NtU-box* genes showed down-regulated expression patterns under *Ras* infection (Figure 12A). It is worth noting that certain *NtU-box* genes exhibit contrasting responses to bacterial and drought stresses. For instance, three genes, *NtU-box3*, *NtU-box82* and *NtU-box91* were up-regulated in response to *Ras* infection; whereas they showed down-regulated under drought conditions (Figures 10A, 12A; Supplementary Table S9). In eggplant, a similar phenomenon was observed. For example, all 8 selected *SmU-box* genes (*SmU-box1/2/3/23/46/47/48/52*) exhibited up-regulated in response to *Ras* infection (Figure 12B), while no significant changes were detected under high temperature stress condition (Figures 10B, 12B; Supplementary Table S9). This result suggested a functional differentiation in the tobacco *NtU-box* gene family in their responses to different stresses. Further study of these genes would greatly enhance the understanding of their functions in both tobacco and eggplant.

Conclusions

In this study, a total of 116 *NtU-box* genes and 56 *SmU-box* genes were identified in the genome of tobacco and eggplant, which were categorized into 5 subfamilies, respectively. These *NtU-box* genes and *SmU-box* genes were randomly distributed on the 24 chromosomes of tobacco and the 12 chromosomes of eggplant.

Phylogenetic analysis suggested a shared ancestor predating the divergence of six species (tobacco, eggplant, potato, tomato, *Arabidopsis thaliana*, and pepper), and segmental duplication event was the primary driver for the expansion of the *U-box* gene family in both tobacco and eggplant. The promoters of *NtU-box* and *SmU-box* genes contained *cis*-regulatory elements associated with cell development, plant hormone response, photoresponsive elements, and stress response. The expression levels of the tobacco and eggplant *U-box* genes varied under various abiotic stress conditions. qRT-PCR analysis revealed that 15 selected *NtU-box* and 8 *SmU-box* genes play important roles in response to pathogen *Ras* invasion in tobacco and eggplant. Our results provided valuable information for further functional study of *U-box* genes in both tobacco and eggplant.

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

RC: Writing – original draft. BZ: Writing – original draft, Data curation. GG: Writing – original draft, Resources. CD: Writing – review & editing, Resources. XL: Writing – review & editing, Resources. WC: Writing – review & editing, Resources. YZ: Writing – review & editing, Resources. TL: Writing – review & editing, Data curation. RW: Writing – review & editing, Data curation. XX: Writing – review & editing, Validation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was financially supported by China Tobacco Company (110202201028 (LS-12) and Longyan Tobacco Company (LK-2022Y02; LK-2022Y06). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1425651/full#supplementary-material

References

Adler, G., Konrad, Z., Zamir, L., Mishra, A. K., Raveh, D., and Bar-Zvi, D. (2017). The Arabidopsis paralogs, PUB46 and PUB48, encoding U-box E3 ubiquitin ligases, are essential for plant response to drought stress. *BMC Plant Biol.* 17, 8. doi: 10.1186/s12870-016-0963-5

Atkinson, N. J., and Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot.* 63, 3523–3543. doi: 10.1093/jxb/ers100

Azevedo, C., Santos-Rosa, M. J., and Shirasu, K. (2001). The U-box protein family in plants. *Trends Plant Sci.* 6, 354–358. doi: 10.1016/S1360-1385(01)01960-4

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., et al. (2019). A chromosome-anchored eggplant genome sequence reveals key events in Solanaceae evolution. *Sci. Rep.* 9(1), 11769. doi: 10.1038/s41598-019-47985-w

Bergler, J., and Hoth, S. (2011). Plant U-box armadillo repeat proteins AtPUB18 and AtPUB19 are involved in salt inhibition of germination in Arabidopsis. *Plant Biol. (Stuttg)* 13, 725–730. doi: 10.1111/j.1438-8677.2010.00431.x

Burr, B. (2002). Mapping and sequencing the rice genome. *Plant Cell* 14, 521–523. doi: 10.1105/tpc.140310

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Cho, S. K., Ryu, M. Y., Kim, J. H., Hong, J. S., Oh, T. R., Kim, W. T., et al. (2017). RING E3 ligases: key regulatory elements are involved in abiotic stress responses in plants. *BMB Rep.* 50, 393–400. doi: 10.5483/BMBRep.2017.50.8.128

Chou, K. C., and Shen, H. B. (2008). Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat. Protoc.* 3, 153–162. doi: 10.1038/nprot.2007.494

Edwards, K. D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A. D., Bombarely, A., et al. (2017). A reference genome for Nicotiana tabacum enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics* 18, 448. doi: 10.1186/s12864-017-3791-6

Feke, A. M., Hong, J., Liu, W., and Gendron, J. M. (2020). A decoy library uncovers U-box E3 ubiquitin ligases that regulate flowering time in arabidopsis. *Genetics* 215, 699–712. doi: 10.1534/genetics.120.303199

Fernandez-Pozo, N., Menda, N., Edwards, J. D., Saha, S., Tecle, I. Y., Strickler, S. R., et al. (2015). The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. *Nucleic Acids Res.* 43, D1036–D1041. doi: 10.1093/nar/gku1195

Finley, D., and Chau, V. (1991). Ubiquitination. *Annu. Rev. Cell Biol.* 7, 25–69. doi: 10.1146/annurev.cb.07.110191.000325

Gao, W., Chen, R., Pan, M., Tang, W., Lan, T., Huang, L., et al. (2019). Early transcriptional response of seedling roots to Ralstonia solanacearum in tobacco (Nicotiana tabacum L.) *European J. Plant Pathol* 155, 527–536. doi: 10.1007/s10658-019-01788-x

Hu, Z., He, Z., Li, Y., Wang, Q., Yi, P., Yang, J., et al. (2022). Transcriptomic and metabolic regulatory network characterization of drought responses in tobacco. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1067076

Kaul, S., Koo, H. L., Jenkins, J., Rizzo, M, Rooney, T., Tallon, L. J., et al. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408, 796–815. doi: 10.1038/35048692

Kim, M.-S., Kang, K.-K., and Cho, Y.-G. (2021b). Molecular and functional analysis of U-box E3 ubiquitin ligase gene family in rice (Oryza sativa). *Int. J. Mol. Sci.* 22 (5), 12088. doi: 10.3390/ijms222112088

Kim, D. Y., Lee, Y. J., Hong, M. J., Kim, J. H., and Seo, Y. W. (2021a). Genome wide analysis of U-box E3 ubiquitin ligases in wheat (Triticum aestivum L.). *Int. J. Mol. Sci.* 22, 2699. doi: 10.3390/ijms22052699

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Liu, R., Shu, B., Wang, Y., Yu, B., Wang, Y., Gan, Y., et al. (2023). Transcriptome analysis reveals key genes involved in the eggplant response to high-temperature stress. *Environ. Exp. Bot.* 211, 105369. doi: 10.1016/j.envexpbot.2023.105369

Liu, Y. C., Wu, Y. R., Huang, X. H., Sun, J., and Xie, Q. (2011). AtPUB19, a U-box E3 ubiquitin ligase, negatively regulates abscisic acid and drought responses in Arabidopsis thaliana. *Mol. Plant* 4, 938–946. doi: 10.1093/mp/ssr030

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Ma, W., Noble, W. S., and Bailey, T. L. (2014). Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat. Protoc.* 9, 1428–1450. doi: 10.1038/nprot.2014.083

March, E., and Farrona, S. (2018). Plant deubiquitinases and their role in the control of gene expression through modification of histones. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02274

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–d419. doi: 10.1093/nar/gkaa913

Mo, Z., Huang, Y., Pu, T., Duan, L., Pi, K., Luo, J., et al. (2023). Genome-wide identification and characterization of Glutathione S-Transferases (GSTs) and their expression profile under abiotic stresses in tobacco (Nicotiana tabacum L.). *BMC Genomics* 24, 341. doi: 10.1186/s12864-023-09450-x

Pan, J., Li, X. I. N., Li, J., Liu, H. U. I., Hou, Z., Yang, Z., et al. (2022). Genome-wide identification of U-box gene family and expression analysis under abiotic stresses in Salvia miltiorrhiza. *Biocell* 46, 1751–1762. doi: 10.32604/biocell.2022.019455

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* 70, 503–533. doi: 10.1146/annurev.biochem.70.1.503

Ryu, M. Y., Cho, S. K., Hong, Y., Kim, J., Kim, J. H., Kim, G. M., et al. (2019). Classification of barley U-box E3 ligases and their expression patterns in response to drought and pathogen stresses. *BMC Genomics* 20, 326. doi: 10.1186/s12864-019-5696-z

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534

Serrano, I., Campos, L., and Rivas, S. (2018). Roles of E3 ubiquitin-ligases in nuclear protein homeostasis during plant stress responses. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00139

Sharma, B., and Taganna, J. (2020). Genome-wide analysis of the U-box E3 ubiquitin ligase enzyme gene family in tomato. *Sci. Rep.* 10, 9581. doi: 10.1038/s41598-020-66553-1

Shu, K., and Yang, W. (2017). E3 ubiquitin ligases: ubiquitous actors in plant development and abiotic stress responses. *Plant Cell Physiol.* 58, 1461–1476. doi: 10.1093/pcp/pcx071

Stone, S. L. (2019). Role of the ubiquitin proteasome system in plant response to abiotic stress. *Int. Rev. Cell Mol. Biol.* 343, 65–110. doi: 10.1016/bs.ircmb.2018.05.012

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tang, X., Ghimire, S., Liu, W., Fu, X., Zhang, H., Sun, F., et al. (2021). Genome-wide identification of U-box genes and protein ubiquitination under PEG-induced drought stress in potato. *Physiologia Plantarum* 174 (1), e13475. doi: 10.1111/ppl.13475

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876

Trenner, J., Monaghan, J., Saeed, B., Quint, M., Shabek, N., and Trujillo, M. (2022). Evolution and functions of plant U-box proteins: from protein quality control to signaling. *Annu. Rev. Plant Biol.* 73, 93–121. doi: 10.1146/annurev-arplant-102720-012310

Trujillo, M. (2018). News from the PUB: plant U-box type E3 ubiquitin ligases. *J. Exp. Bot.* 69, 371–384. doi: 10.1093/jxb/erx411

Walter, W., Sánchez-Cabo, F., and Ricote, M. (2015). GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* 31, 2912–2914. doi: 10.1093/bioinformatics/btv300

Wang, N., Liu, Y., Cong, Y., Wang, T., Zhong, X., Yang, S., et al. (2016). Genome-wide identification of soybean U-box E3 ubiquitin ligases and roles of gmPUB8 in negative regulation of drought stress response in arabidopsis. *Plant Cell Physiol.* 57, 1189–1209. doi: 10.1093/pcp/pcw068

Wang, C., Song, B., Dai, Y., Zhang, S., and Huang, X. (2021). Genome-wide identification and functional analysis of U-box E3 ubiquitin ligases gene family related to drought stress response in Chinese white pear (Pyrus bretschneideri). *BMC Plant Biol.* 21 (1), 235. doi: 10.1186/s12870-021-03024-3

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wiborg, J., O'Shea, C., and Skriver, K. (2008). Biochemical function of typical and variant Arabidopsis thaliana U-box E3 ubiquitin-protein ligases. *Biochem. J.* 413, 447–457. doi: 10.1042/BJ20071568

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., et al. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552. doi: 10.1385/1-59259-584-7:531

Yan, J., Wang, J., Li, Q., Hwang, J. R., Patterson, C., and Zhang, H. (2003). AtCHIP, a U-box-containing E3 ubiquitin ligase, plays a critical role in temperature stress tolerance in Arabidopsis. *Plant Physiol.* 132, 861–869. doi: 10.1104/pp.103.020800

Yang, Y., Chai, Y., Liu, J., Zheng, J., Zhao, Z., Amo, A., et al. (2021b). Amino acid transporter (AAT) gene family in foxtail millet (Setaria italica L.): widespread family expansion, functional differentiation, roles in quality formation and response to abiotic stresses. *BMC Genomics* 22 (1), 519. doi: 10.1186/s12864-021-07779-9

Yang, C. W., González-Lamothe, R., Ewan, R. A., Rowland, O., Yoshioka, H., Shenton, M., et al. (2006). The E3 ubiquitin ligase activity of arabidopsis PLANT U-BOX17 and its functional tobacco homolog ACRE276 are required for cell death and defense. *Plant Cell* 18, 1084–1098. doi: 10.1105/tpc.105.039198

Yang, Y., Wang, X., Zheng, J., Men, Y., Zhang, Y., Liu, L., et al. (2022). Amino acid transporter (AAT) gene family in Tartary buckwheat (Fagopyrum tataricum L. Gaertn.): Characterization, expression analysis and functional prediction. *Int. J. Biol. Macromol* 217, 330–344. doi: 10.1016/j.ijbiomac.2022.07.059

Yang, G., Ying, G., Wang, Z., Pan, W., Linghu, B., Pan, Y., et al. (2021a). Genome-wide identification and expression analysis of U-box gene family in wild emmer wheat (Triticum turgidum L. ssp. dicoccoides). *Gene* 799, 145840. doi: 10.1016/j.gene.2021.145840

Zeng, L. R., Park, C. H., Venu, R. C., Gough, J., and Wang, G. L. (2008). Classification, expression pattern, and E3 ligase activity assay of rice U-box-containing proteins. *Mol. Plant* 1, 800–815. doi: 10.1093/mp/ssn044

Zhang, G., Zhang, M., Zhao, Z., Ren, Y., Li, Q., and Wang, W. (2017). Wheat TaPUB1 modulates plant drought stress resistance by improving antioxidant capability. *Sci. Rep.* 7, 7549. doi: 10.1038/s41598-017-08181-w

Zhou, J., Hu, Y., Li, J., Yu, Z., and Guo, Q. (2021). Genome-wide identification and expression analysis of the plant U-box protein gene family in phyllostachys edulis. *Front. Genet.* 12. doi: 10.3389/fgene.2021.710113

Zhu, Y., Li, Y., Fei, F., Wang, Z., Wang, W., Cao, A., et al. (2015). E3 ubiquitin ligase gene CMPG1-V from Haynaldia villosa L. contributes to powdery mildew resistance in common wheat (Triticum aestivum L.). *Plant J.* 84, 154–168. doi: 10.1111/tpj.12966

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 06 August 2024
DOI 10.3389/fpls.2024.1416216

Check for updates

OPEN ACCESS

EDITED BY
Raju Datla,
Global Institute for Food Security (GIFS),
Canada

REVIEWED BY
Sarah J. Whitcomb,
Agricultural Research Service (USDA),
United States
Zhihua Hua,
Ohio University, United States

*CORRESPONDENCE
Liang Song
liang.song@botany.ubc.ca

RECEIVED 12 April 2024
ACCEPTED 20 June 2024
PUBLISHED 06 August 2024

CITATION
Go D, Lu B, Alizadeh M, Gazzarrini S and Song L (2024) Voice from both sides: a molecular dialogue between transcriptional activators and repressors in seed-to-seedling transition and crop adaptation.
Front. Plant Sci. 15:1416216.
doi: 10.3389/fpls.2024.1416216

COPYRIGHT
© 2024 Go, Lu, Alizadeh, Gazzarrini and Song. 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.

Voice from both sides: a molecular dialogue between transcriptional activators and repressors in seed-to-seedling transition and crop adaptation

Dongeun Go[1], Bailan Lu[1], Milad Alizadeh[1], Sonia Gazzarrini[2,3] and Liang Song[1*]

[1]Department of Botany, University of British Columbia, Vancouver, BC, Canada, [2]Department of Biological Science, University of Toronto Scarborough, Toronto, ON, Canada, [3]Department of Cell and Systems Biology, University of Toronto, Toronto, ON, Canada

High-quality seeds provide valuable nutrients to human society and ensure successful seedling establishment. During maturation, seeds accumulate storage compounds that are required to sustain seedling growth during germination. This review focuses on the epigenetic repression of the embryonic and seed maturation programs in seedlings. We begin with an extensive overview of mutants affecting these processes, illustrating the roles of core proteins and accessory components in the epigenetic machinery by comparing mutants at both phenotypic and molecular levels. We highlight how omics assays help uncover target-specific functional specialization and coordination among various epigenetic mechanisms. Furthermore, we provide an in-depth discussion on the Seed dormancy 4 (Sdr4) transcriptional corepressor family, comparing and contrasting their regulation of seed germination in the dicotyledonous species Arabidopsis and two monocotyledonous crops, rice and wheat. Finally, we compare the similarities in the activation and repression of the embryonic and seed maturation programs through a shared set of *cis*-regulatory elements and discuss the challenges in applying knowledge largely gained in model species to crops.

KEYWORDS
climate, crop adaptation, epigenetic regulation, HDAC, LAFL, PKL, PRC, Sdr4

Introduction

Seeds contributed greatly to the successful colonization of dry land by spermatophytes (seed plants) by allowing reproduction in the absence of water and dispersal under unfavorable growth conditions. During mid-to-late development, seeds enter the maturation phase, during which the seed accumulates storage compounds (lipids,

proteins, or polysaccharides depending on the species) that are needed during germination to sustain post-germinative growth. Seeds can contribute up to 70% of our caloric intake, as food and livestock feed, and therefore play a fundamental role in human nutrition. For these reasons, there is a great interest in understanding the gene regulatory networks (GRNs) controlling seed development and germination to improve qualitative and quantitative traits associated with these processes.

GRN depicts transcriptional regulators and their target genes as nodes and the regulatory relationships as directed edges (Barabási and Oltvai, 2004). GRN analysis often employs one or more of model-, information theory-, and machine learning-based methods (Zhao et al., 2021). Although barebone GRNs can be inferred from transcription profiles of transcription factors (TFs) and their potential target genes based on co-expression in bulk tissues, higher resolution of transcriptome profiles and data examining additional regulatory features will provide valuable information for more accurate inference and a higher understanding of context-specific regulation. However, data availability differs by biological systems. In animals, single-cell (sc) omics data are widely available. These datasets are increasingly paired, which means various types of omics data are collected from the same cells. Equally importantly, a plethora of tools, developed and validated with multi-omics data in animals, are available to infer GRNs (Badia-i-Mompel et al., 2023). By contrast, although a growing amount of plant sc transcriptome and ATAC-seq profiles became available in recent years, sc datasets are still relatively scarce in number and type in plants, and multi-omics assays are usually generated from separate samples, resulting in reduced data resolution, increased noise, and fewer features for multimodal predictions of important biological processes. To date, plant sc transcriptome profiles of vegetative tissues have provided excellent insights into plant development (Shahan et al., 2022), evolution (Guillotin et al., 2023), and response to environmental cues (Wendrich et al., 2020). TF binding sites are widely used together with sc profiles of transcriptome or chromatin accessibility in integrative analysis pipelines such as MINI-EX and MINI-AC to infer cell type-specific GRN (Ferrari et al., 2022; Manosalva Pérez et al., 2024).

There are several challenges in studying transcriptional repression in seeds. First, high-resolution data are limited, especially during late seed maturation. Second, the regulation of gene repression is arguably more complex. For instance, the inference of gene activation is relatively straightforward based on open chromatin, binding (sites) of transcriptional activators, and elevated transcript abundance. By contrast, lack of transcription could result from either active repression or a lack of activation, which may not be distinguishable solely by chromatin accessibility assays. Additionally, both computational analysis (Brooks et al., 2021) and experimental evidence (Zhu et al., 2024) show that TFs may play a dual role of activation and repression depending on their interacting partners, and such characterizations are relatively limited in seeds. In this review, we highlight various omics datasets useful for GRN inference in seeds. We place a special emphasis on the repressive machinery by detailing the genetic and functional genomic characterizations to provide context for their mode of action. Considering complex biological networks are hierarchical and scale-free (Basso et al., 2005), we focus on the master LAFL TFs in the examples because these hubs are likely to capture extensive regulation.

The function of LAFL in seed development and maturation has been extensively studied in Arabidopsis (Lepiniec et al., 2018; Alizadeh et al., 2021; Gazzarrini and Song, 2024). *LEC1, FUS3*, and *LEC2* transcripts accumulate since the zygote or pre-globular stage of embryogenesis, and the expression of ABI3 initiates later at the globular stage. These TFs play a prominent role during seed maturation, where they are required for the accumulation of seed storage compounds, such as triacylglycerols (TAGs), seed storage proteins (SSPs), oleosins (OLEO), and stress proteins (LEA) (Giraudat et al., 1992; Meinke, 1992; Meinke et al., 1994; West et al., 1994; Roscoe et al., 2015). LAFL also promote seed dormancy and suppress precocious germination of immature seeds, known as vivipary or pre-harvest sprouting (PHS) in cereals, by inhibiting cell division, the activity of the shoot and root meristems, and the differentiation of the cotyledon epidermis (trichome development) and vascular system (xylem) (Meinke, 1992; Keith et al., 1994; Nambara et al., 1995; Raz et al., 2001; Bryant et al., 2019). LAFL's functions in seed development and germination are partly mediated by the hormones auxin, ABA, GA, ethylene, brassinosteroids, and jasmonate (Parcy et al., 1994; Curaba et al., 2004; Gazzarrini et al., 2004; Braybrook et al., 2006; Stone et al., 2008; Lumba et al., 2012; Ryu et al., 2014; Chiu et al., 2016; Pan et al., 2020). The spatiotemporal expression patterns of *LAFL* are tightly controlled at the transcriptional and epigenetic levels. Epigenetic silencing of *LAFL* is required to promote post-embryonic development (Lepiniec et al., 2018; Alizadeh et al., 2021; Gazzarrini and Song, 2024).

Seed quality is a holistic term that includes seed viability, moisture and nutrient content, depth of dormancy, longevity, and vigor. Some of these traits, such as seed vigor, are assessed during the transition from seed to seedlings (Finch-Savage and Bassel, 2016). Therefore, we focus on the regulation and performance of seed-to-seedling transition by using this process as a proxy of seed quality in this review. Recent advances made possible by omics tools provide a rich resource to compare and contrast various epigenetic machinery that regulate the transition from seed to seedlings. The information will also provide mechanistic context for GRN inference at this developmental stage. Using a novel transcriptional cofactor, Seed dormancy 4 (Sdr4), as an example, we summarize and discuss the role of this corepressor and its orthologs in Arabidopsis, rice, and wheat. We also discuss cis-regulatory elements (CREs) important for seed maturations as well as a potential regulatory symmetry exerted on these CREs, which together facilitate the activation and repression of the seed maturation program.

Phenotypic and omics resemblance of genetic and epigenetic regulators

Functional characterization of players in seed-to-seedling transition by phenotypic resemblance and genetic evidence

The importance of shutting down the embryonic program and silencing *LAFL* at the end of seed development is clearly shown by *LAFL* overexpression/ectopic expression (OE) phenotypes, which

include delayed germination, growth and flowering (*ABI3, FUS3*, and *LEC2*), development of cotyledon-like organs (*LEC1*, *FUS3*, and *LEC2*), somatic embryos, and callus-like structures (*LEC1* and *LEC2*) (Table 1). Ectopic expression of early-acting genes (*LEC1* and *LEC2*) shows the most dramatic phenotypes, such as somatic embryos and development of callus-like structures. These embryonic structures accumulate seed storage lipids (TAGs), and proteins (2S and 12S), as a result of ectopic expression of the seed maturation program (Table 1). Robust repression of *LAFL* and the embryonic program during vegetative development is orchestrated by a suite of epigenetic regulators, including Polycomb group (PcG) repressive complex 1 (PRC1) and PRC2, Trithorax group (TrxG), chromatin remodeling factors and other repressive proteins (Figure 1) (Xiao et al., 2017a; Lepiniec et al., 2018; Gazzarrini and Song, 2024). PRC1 catalyzes monoubiquitylation of lysine (K) 121 on histone 2A (H2AK121ub). PRC1 core components include BMI1A, BMI1B, BMI1C, RING1A, and RING1B (Figure 1), all of which have E3 ligase activity (Mozgova and Henning, 2015; Baile et al., 2022). PRC2 promotes trimethylation of H3 on K27 (H3K27me3), catalyzed by the histone methyltransferases SWN, CLF, and MEA. Arabidopsis has three core PRC2 complexes: EMF-PRC2, composed of EMF2, CLF/SWN, FIE, and MSI1; VRN-PRC2, composed of VRN2, CLF/SWN, FIE, and MSI1; and FIS-PRC2, which includes FIS2, MEA, FIE, and MSI1 (Figure 1). These PRC2 complexes have overlapping as well as specific functions throughout development (Bieluszewski et al., 2021; Baile et al., 2022). Several PRC1 and PRC2 accessory proteins that physically interact with the PRC core components to repress transcription, such as VAL1/2, HDAC, LHP1/TFL2, AL6, AL7, EMF1, NDX, and VRN5/VIL1, as well as TrxG factors such as H3K4me3 methyltransferases ATX and ULT1, chromatin remodeling factors such as PKL and PKR2, and corepressors such as TPL, TPR, and AtSDR4L, aid in repression of *LAFL* and the embryonic program during vegetative development (Gazzarrini and Song, 2024).

Loss-of-function mutants in PRC1 and PRC2 components mimic *LAFL* OE phenotypes, supporting a role for PRCs in

TABLE 1 Genetic and epigenetic regulators in seed-to-seedling transition.

OE/ mutant	Phenotypes	Genetic and genomic insights related to seeds and seedlings
LEC1 OE	**Storage compounds:** Ectopic lipids and starch in *lec1-d* ^{tnp} (Casson and Lindsey, 2006). Ectopic cruciferin and starch in *35S:LEC1-GR* seedlings induced 0–3 days after imbibition (DAI) (Junker et al., 2012). **Callus and somatic embryos (SE):** Cotyledon-like organs and somatic embryos in *35S:LEC1* seedlings (Lotan et al., 1998). Swollen hypocotyl, but no embryonic callus in *lec1-d* ^{tnp} (dominant mutant). Phenotype enhanced by auxin, sugar or GA inhibitor (paclobutrazol), but not rescued by GA (Casson and Lindsey, 2006). *35S:LEC1-GR* seedlings show different phenotypes dependent on time of induction: most show cotyledon-like organs, swollen and green roots, and callus and somatic embryos, and 10%–40% show arrested roots (0–2 DAI, days after imbibition); long hypocotyls with an apical hook (3 DAI); no phenotype if induced at 4 DAI, but induction at 4DAI+ABA resulted in cotyledon-like leaves that express *CRU* (Junker et al., 2012).	**Selected marker genes:** Ectopic expression (*in situ* hyb) of seed storage proteins: 12S cruciferins (*CRA*), S3 oleosin (*OLEO*) in *35S:LEC1* seedlings (Lotan et al., 1998). Ectopic expression of *LEC2, FUS3, ABI3*, and maturation genes (*2S* albumin, *CRC*) in *35S:LEC1-GR* seedlings (Kagaya et al., 2005b). Ectopic expression of *CRU, LEC2, FUS3*, and *ABI3* in *35S:LEC1-GR* seedlings induced 0–3 DAI (Junker et al., 2012) (Junker et al., 2012). *35S:LEC1-GR* associates with *LAFL*, *WRI*, and seed maturation genes in ChIP-chip in seedlings (Pelletier et al., 2017). **Omics datasets:** GSE22352 (ChIP-chip of LEC1 of 2-week-old *35S:LEC1:GR* seedlings treated by DEX or mock for 24 h), GSE22173 (ATH1 microarray of 2-week-old *35S:LEC1-GR* treated by DEX or mock for 8 h with and without ABA) (Junker et al., 2012). GSE99528 (microarray of 8-day-old *35S:LEC1:GR* treated with and without DEX for 1 h), GSE99529 (ChIP-chip of LEC1 in 8-day-old *35S:LEC1-GR* seedlings either grown on DEX plates or treated by DEX for 4 h), GSE99587 (ChIP-seq of LEC1 in *LEC1:LEC1-GFP*: *LEC1/lec1–1* seeds at the bent cotyledon-stage) (Pelletier et al., 2017).
LEC2 OE	**Storage compounds:** Ectopic lipids and starch accumulation in *35S:LEC2-GR* ovules (Stone et al., 2008). Ectopic accumulation of seed-specific lipids and triacylglycerol (TAGs) in leaves of 35S:LEC2-GR induced after 2 weeks (Santos Mendoza et al., 2005) **Callus and SE:** Somatic embryos in *35S:LEC2* (Stone et al., 2001).	**Selected marker genes:** Ectopic expression of *CRA* (*12S*) and *OLEO* detected by *in situ* hybridization in *35S:LEC2* seedlings (Stone et al., 2001). Ectopic expression (RT-PCR) of *LEC1*, *FUS3*, *2S, CRA1*, and *OLEO* in *35S:LEC2-GR* seedlings (Stone et al., 2001). Ectopic expression of *S3* (*OLEO*), *2S3* (albumin), and *LAFL* in leaves of *35S: LEC2-GR* (Santos Mendoza et al., 2005). Ectopic expression (microarray and/or RT-PCR) of *LEC1*, *FUS3*, *2S*, *CRA1*, and *OLEO* in *35S:LEC2-GR*. LEC2 associates (ChIP) with *2S3* and *OLEO* in *35S:LEC2-GR* seedlings (Stone et al., 2008). **Omics datasets:** GSE3959 (ATH1 microarray of 8-day-old *35S:LEC2-GR* seedlings treated by DEX for 1 and 4 h) (Braybrook et al., 2006).
FUS3 OE	**Dormancy and germination:** Delayed germination, vegetative growth, and flowering of *fus3 ML1:FUS3*. Strong lines are arrested at the seedling stage (Gazzarrini et al., 2004; Tsai and Gazzarrini, 2012). *ML1:FUS3* seeds hypersensitive to ABA, sorbitol, and glucose during germination (Tsai and Gazzarrini, 2012). **Storage compounds:** Ectopic accumulation of seed storage proteins (2S, 12S) in *ML1:FUS3* leaves	**Selected marker genes:** Ectopic expression of *2S3* (enhanced by *ABA*). Ectopic expression of *CRC* only in +ABA in 7 DAI seedlings of DEX inducible *GRpro:FUS3* (Kagaya et al., 2005a). Repression of GA biosynthesis genes in *fus3 ML1:FUS3* seedlings (*GA20ox* and *GA3ox*) (Gazzarrini et al., 2004). Ectopic expression of *OLEO*, *2S3*, *CRU*, and *WRI* in *ESTpro: FUS3* (Zhang et al., 2016).

(Continued)

TABLE 1 Continued

OE/ mutant	Phenotypes	Genetic and genomic insights related to seeds and seedlings
	(Gazzarrini et al., 2004). Increased TAGs content in estradiol-inducible *XVEpro: FUS3* (Zhang et al., 2016). **Callus and SE:** Development of cotyledon-like organs, arrested seedlings, but no somatic embryos in *fus3 ML1:FUS3* strong lines. Delayed flowering and cotyledon-like leaves are partially rescued by GA (Gazzarrini et al., 2004).	**Genomics datasets:** GSE43291 (ChIP-chip of *FUS3:FUS3-myc/fus3–3* embryonic culture) (Wang and Perry, 2013) GSE80360 (ATH1 microarray of 8 day-old *XVEpro: FUS3* seedlings ± sucrose) (Zhang et al., 2016).
ABI3 OE	**Dormancy and germination:** Increased sensitivity to ABA during seed germination (Zhang et al., 2005). Increased sensitivity to ABA in inhibition of root elongation (Parcy et al., 1994). **Storage compounds:** Increased TAGs content in *XVEpro: ABI3* (Yang et al., 2022b).	**Selected marker genes:** Ectopic expression of seed maturation genes *2S3*, *Em1*, *Em6*, and *CRC* in *35S:ABI3* in response to ABA (Parcy et al., 1994). Ectopic expression of *2S3*, and ectopic expression of *Em* and *CRC* in response to ABA in 7-day-old seedlings of *GRpro: ABI3* (Kagaya et al., 2005a). Ectopic depression of *FUS3*, *OLEO*, and *WRI* in *XVEpro: ABI3* (Yang et al., 2022b). **Omics datasets:** GSE150561 (microarray of wild-type and *abi3–5* seeds harvested at 15–16 days after flowering with and without placement on moist blotter for 1 day; ChIP-chip of *ABI3:ABI3-myc/abi3–5* embryonic culture) (Tian et al., 2020) PRJNA678646 (RNA-Seq of XVEpro: *ABI3* using fifth to eighth rosette leaves treated with β-estradiol or mock for 4 days) (Yang et al., 2022b).
atbmi1a atbmi1b	**Dormancy and germination:** Delayed germination both in unstressed condition and under salt or osmotic stress (Molitor et al., 2014) **Storage compounds:** Low penetrance of the pickle-root trait (~8%) (Chen et al., 2010a). **Callus and SE:** Low penetrance of embryonic callus (~18%) (Chen et al., 2010a). The same double mutant exhibits high penetrance (>75%) of callus and somatic embryos in another study (Bratzel et al., 2010).	**Selected marker genes:** Increased expression of *STM*, *WOX5*, *WUS*, *LEC1*, and *FUS3* (Bratzel et al., 2010). Derepression of *LAFL* in *atbmi1a atbmi1b* examined by RT-qPCR in 2-week-old seedlings (Chen et al., 2010a). Reduced H2AK121ub of *LEC1*, *FUS3*, *ABI3*, *WUS*, and *BBM* (Yang et al., 2013). Increased expression of *ABI3, DOG1, CRU1, CUR3, PER1*, and *CHO1*. Altered histone marks (H3K4me3, H3K27me3) (Molitor et al., 2014). **Omics datasets:** GSE67322 (RNA-seq and H3K27me3 ChIP-seq of 2-week-old *atbmi1a atbmi1b*, *atring1a atring1b*, *lhp1*, *clf*, and *clf swn* seedlings) showed that H3K27me3 and differential expression of seed maturation genes are similarity regulated in *clf swn*, *atring1a atring1b*, and *atbmi1a atbmi1b* (Wang et al., 2016). GSE89358 (RNA-seq of 7-day-old *atbmi1abc* and wild-type seedlings; H3K27me3 and H2AK121ub ChIP-seq of *atbmi1abc*, *clf swn*, *lhp1*, and wild-type seedlings) (Zhou et al., 2017). GSE83568 (RNA-seq of 10-day-old single, double, and triple mutants of *atbmi* and wild-type seedlings). PRJE 52473 (Hi-C and ChIP-seq of BMI1B-FLAG and H3K4me3 using 10-day-old seedlings) (Yin et al., 2023)
atring1a atring1b	**Dormancy and germination:** Hypersensitive to ABA in seedling establishment (Zhu et al., 2020). **Storage compounds:** moderate penetrance of the pickle-root phenotype (~50%) (Chen et al., 2010a). **Callus and SE:** Low penetrance of embryonic callus (~17%). Embryonic and pickle-root traits are attenuated by auxin transport inhibitor NPA (Chen et al., 2010a).	**Selected marker genes:** RT-qPCR showed that *LAFL* are derepressed in *atring1a atring1b* 2-week-old and 1-month-old seedlings (Chen et al., 2010a). *ABI3* is upregulated in *atring1a atring1b* (Zhu et al., 2020). **Omics datasets:** GSE67322 (see description in the *atbmi1a atbmi1b* row in this table). GSE155378 (ATAC-seq of *atbmi1abc*, *atring1a atring1b*, *clf swn*, *emf1*, and *lhp1*; H2AK121ub ChIP-seq of *emf1* and *atring1a atring1b*; H3K27me3 ChIP-seq of *atring1a atring1b* and *lhp1*; RNA-seq of *clf swn*, *atring1a atring1b* and *emf1*; all assays were carried out using 10-day-old whole seedlings and include wild-type Col-0 controls) (Yin et al., 2021).
atring1a atring1b clf	**Storage compounds:** *clf* slightly exacerbates the pickle-root phenotype of *atring1a atring1b* (Chen et al., 2010a). **Callus and SE:** *clf* substantially exacerbates the embryonic callus phenotype of *atring1a atring1b* (Chen et al., 2010a).	**Selected marker genes:** RT-PCR showed upregulation of *LEC1, LEC2*, and *FUS3* is further increased in *atring1a atring1b clf* compared to *atring1a atring1b* (Chen et al., 2010a).
al6 al7	**Dormancy and germination:** Delayed germination; enhanced under salt (NaCl) or osmotic (mannitol) stress (Molitor et al., 2014). **Storage compounds:** Increased level of *CRU1* and *CRU3* (3 DAI). Tissue-level defects not observed.	**Selected marker genes:** Increased expression of *ABI3, DOG1, CRU1, CUR3, PER1*, and *CHO1*, but much lower than in *atbmi1a atbmi1b*. Altered histone marks at *ABI3* and *DOG1* in 3 DAG (increased H3K4m3, decreased H3K27me3) albeit less than in *atbmi1a atbmi1b*. AL6 shows similar binding to LHP1 at the *ABI3* and *DOG1* loci (Molitor et al., 2014)

(Continued)

TABLE 1 Continued

OE/ mutant	Phenotypes	Genetic and genomic insights related to seeds and seedlings
atbmi1a atbmi1b al6 al7	**Dormancy and germination:** Further delayed germination on regular plates and under salt (NaCl) or osmotic (mannitol) compared to *atbmi1a atbmi1b* and *al6 al7* double mutants (Molitor et al., 2014).	
ndx	**Dormancy and germination:** Hypersensitive to ABA during seedling establishment (greening) and root growth (Zhu et al., 2020).	**Selected marker genes:** *ABI4* and *ABI5* upregulation in *ndx*, and *ABI3* and *ABI4* upregulation in *ndx* + ABA; RNA-seq and RT-qPCR showed that NDX binds to *ABI3*, *ABI4*, and *ABI5*, but associates only with *ABI4* in ChIP-qPCR. Upregulation of these genes is much stronger in *atring1a atring1b*. ChIP-qPCR showed that levels of H2AK121ub are reduced at *ABI4, Em1*, and *SUT4*, and slightly at *ABI3* in *ndx-5* and to a similar level to *atring1a atring1b* (Zhu et al., 2020). **Omics datasets:** PRJNA556351 (RNA-seq of 7-day-old seedlings *ndx*, *atring1a atring1b*, and Col-0 ± ABA) (Zhu et al., 2020). GSE201841 (RNA-seq, BS-seq, and sRNA-seq DRIP-seq of *ndx1–4* and wild-type control. ChIP-seq of flag-NDX and NDX-GFP. Ten-day-old seedlings were used for the sequencing assays) (Karányi et al., 2022).
lhp1 (*tfl2*)	**Dormancy and germination:** Elevated expression of *DOG1*. Delayed germination on ABA plates, possibly mediated by ANAC019 and ANAC055 (Ramirez-Prado et al., 2019).	**Selected marker genes:** LHP1 binds to *ABI3* and *DOG1* (Molitor et al., 2014). Modest upregulation of *DOG1* in *lhp1* mutant (Chen et al., 2020). No deregulation of *FUS3* or *ABI3* in 14-day-old *lhp1* seedlings (Ramirez-Prado et al., 2019). **Omics datasets:** DamID-chip, an *E. coli* Dam fused with LHP1 followed by tiling microarray profiling, and A-MEXP-602 (chip-chip of LHP1 and H3K27me3 in 10-day-old wild-type and *lhp1* seedlings) showed that LHP1 colocalizes genome-wide with H3K27me3 (Turck et al., 2007; Zhang et al., 2007). GSE76571 (ChIP-seq of LHP1 and H3K27me3 in 14-day-old wild-type, *lhp1*, and *clf* seedlings) showed LHP1 is involved in the spreading of H3K27me3 and shaping chromatin topology (Veluchamy et al., 2016). GSE67322 (see description in the *atbmi1a atbmi1b* row in this table). GSE89358 (see description in the *atbmi1a atbmi1b* row in this table). GSE155378 (see description in the *atring1a atring1b* row in this table).
emf1	**Callus and SE:** *emf1–2* forms oval-shaped, petiole-less cotyledons that develop into carpeloid, and no leaf primordia and do not produce vegetative rosettes (Sung et al., 1992; Chen et al., 1997). A small % of *emf1–2* and *emf1–2 emf2–1* plants form callus after 1 month of culture (Calonje et al., 2008).	**Selected marker genes:** Increased *ABI3* and *At2S3* and decreased *LEC1* and *LEC2* transcript abundance in 14-day-old seedlings (Xu et al., 2018). **Genomics datasets:** GSE155378 (see description in the *atring1a atring1b* row in this table).
emf1 atx, *emf1 ult*, *emf1 atx1 ult1*	**Storage compounds:** Increased storage lipids in pickle-root regions of *emf1 atx, emf1 ult*, and *emf1 atx ult* (Xu et al., 2018). **Callus and SE:** *emf1 atx (4%), emf1 ult (11%)*, and *emf1 atx ult (22%)* show embryo- and callus-like structures arising from cotyledons, hypocotyls, and roots; roots are arrested and swollen, pkl-like. The pickle-root phenotype is enhanced by a GA biosynthesis inhibitor, PAC. *atx*, *ult*, or *atx ult* does not show any of these phenotypes (Xu et al., 2018).	**Selected marker genes:** Upregulation of *ABI3*, *FUS3*, *LEC2*, seed maturation genes (*2S*, *OLEO, CRU*, and *LEA*), dormancy (*DOG1*), GA catabolism (*GA2ox1*) and GA signaling repressor (*RGL1*), and downregulation of GA synthesis (*GA3ox1*) in 14-day-old *emf1 atx1 ult1* seedlings by qRT-PCR and/or RNA-seq (*ABI3*, *FUS3*, *LEC2*, and 2S3 are also upregulated in *emf1 ult* but not in *atx ult*, by qRT-PRC). *LEC1* is downregulated in *atx, ult, emf1*, *atx ult*, and *ult atx emf*. ChIP-seq and ChIP-qPCR showed that ULT and ATX are associated with *ABI3*, *LEC2*, and *2S3*, by ChIP-seq and ChIP-qPCR. Decreased H3K27me3 marks at *ABI3*, *LEC2*, seed maturation genes (*CRU* and *OLEO*) and dormancy (*DOG1*) in *emf1 atx1 ult1* (Xu et al., 2018).
fie	**Dormancy and germination:** Delayed germination and 40% dormant seeds, delayed cotyledon greening similar to WT germinated on ABA; these phenotypes are not rescued by GA (Bouyer et al., 2011). **Storage compounds:** Sugar-enhanceable accumulation of storage reserves at the root tip and in the aerial part (Bouyer et al., 2011). **Callus and SE:** Development of somatic embryos and callus-like structures in seedlings (Makarevich et al., 2006; Bouyer et al., 2011).	**Selected marker genes:** Strong decrease in H3K27me3 levels and *ABI3, FUS3*, and *LEC2* derepression. Upregulation of seed maturation (*CRU3*, *CRA1*, *2S1*, *2S2*, *OLEO*, and *LEA*), dormancy (*DOG1* and *SOM*), and ABA signaling (*ABI4*) (Bouyer et al., 2011). **Omics datasets:** GPL10918 (ChIP-chip H3K27me3 and H3K4me3 in 20-day-old *fie* and wild-type seedlings) (Bouyer et al., 2011). GSE95562 (ChIP-seq of FIE in 30-h-old *pRNAi-BPC*; *pRNAi-ZnF* double knockdown mutant and wild-type Ws accession); GSE84483 (ChIP-seq of pFIE: FIE-HA in 30-h-old *fie-11* in C24 accession) (Xiao et al., 2017b).

(Continued)

TABLE 1 Continued

OE/ mutant	Phenotypes	Genetic and genomic insights related to seeds and seedlings
clf	**Dormancy and germination:** Two mutant alleles of *clf* show reduced seed yield and increased cell size and seed size (Liu et al., 2016). **Storage compounds:** Increased level of FA and oil.	**Selected marker genes:** Upregulation of *LAFL*, several *OLE*, and *WRI* in *clf* siliques; upregulation of *FUS3* and *ABI3* and downregulation of *LEC1* in mat green *clf* embryo. *clf* seedlings: decreased H3K27me3 at *FUS3*, *ABI3* (Liu et al., 2016). **Omics datasets:** GSE7065 (ChIP-chip of 10- to 14-day-old 35S::GFP-CLF in *clf-50* in Ws accession). GSE67322 (see description in the *atbmi1a atbmi1b* row in this table). GSE103361 (RNA-seq and H3K27me3 ChIP-seq of 3-week-old shoots of *pkl*, *clf*, and wild-type plants) (Carter et al., 2018). GSE155502 (HiChIP of H3K27me3 and H3K9ac and Hi-C in 14-day-old wild-type and *clf* seedlings) showed altered H3K27me3 repressive loops in *clf* (Huang et al., 2021).
swn clf	**Dormancy and germination:** Delayed germination, stronger than *fie* (Bouyer et al., 2011). **Storage compounds:** Increase accumulation of storage reserves. **Callus and SE:** Development of somatic embryos and callus-like structures in seedlings. Swollen root that produced green shoot-like tissue, similar to *fie* (Chanvivattana et al., 2004; Makarevich et al., 2006).	**Selected marker genes:** *LEC1*, *LEC2*, and *FUS3* derepression in seedlings; strong *FUS3* (not *LEC1* or *LEC2*) upregulation in 3DAP siliques (Makarevich et al., 2006). **Omics datasets:** GSE67322 (see description in the *atbmi1a atbmi1b* row in this table). GSE89358 (see description in the *atbmi1a atbmi1b* row in this table). GSE98301 (RNA-seq of 10-day-old *clf-50 swn-1* and wild-type seedlings treated with ABA for 5 h or 4 days. RNA-seq of 10-day-old *clf-50*, *swn-1*, *clf-50 swn-1*, and wild-type seedlings mock-treated for 4 days) (Liu et al., 2019). GSE108960 (RNA-seq of 10-day-old *clf*, *swn*, *clf swn*, and wild-type seedlings, and ChIP-seq of 10-day-old CLF-GFP and SWN-GFP seedlings) showed CLF and SWN function redundantly to deposit H3K27me3 at *LAFL* loci (Shu et al., 2019). GSE155378 (see description in the *atring1a atring1b* row in this table).
swn clf pkl	**Storage compounds:** Accumulation of storage lipids detected by Fat Red B staining (Aichinger et al., 2009). **Callus and SE:** *pkl* enhances embryo and callus-like structure compared to *swn clf* (Aichinger et al., 2009).	**Selected marker genes:** *LEC1* and *FUS3* are synergistically upregulated in *pkl clf* double mutant (Aichinger et al., 2009).
emf2 vrn2	**Storage compounds:** Ectopic storage lipid and chlorophyll accumulation in *emf2 vrn2* seedling roots (Ikeuchi et al., 2015). **Callus and SE:** Somatic embryos in *emf2 vrn2* (Schubert et al., 2005). Ectopic shoot on *emf2 vrn2* root (Ikeuchi et al., 2015).	**Selected marker genes:** Ectopic expression of *LEC1*, *LEC2*, and *FUS3* in *emf vrn2* roots (Ikeuchi et al., 2015).
emf2 sdg8	**Storage compounds:** Accumulation of storage proteins and lipids (Tang et al., 2012). **Callus and SE:** No somatic embryos in *emf2*, *vrn2*, or *sdg8*. Somatic embryos in *emf2 sdg8* (Tang et al., 2012).	**Selected marker genes:** Derepression of *FUS3*, seed maturation genes (*2S*, *LEA*, and *LTP*), and GA deactivation genes (*GA2ox*) in *sdg8* (Tang et al., 2012). Depression of *LAFL*, seed maturation genes (*2S*, *LEA*, and *LTP*), and GA deactivation genes (*GA2ox*) in *sdg8 emf2* (Tang et al., 2012).
bpc1 bpc2	**Dormancy and germination:** No dormancy and germination phenotype. Instead, double mutant is pleiotropic, exhibiting ovule and seed abortion, dwarfism, and reduced lateral roots (Monfared et al., 2011).	**Selected marker genes:** BPCs repress *LEC2* (Xiao et al., 2017b), *FUS3* (Xiao et al., 2017b; Wu et al., 2020), and *ABI4* (Mu et al., 2017). BPCs activate *LEC2* (Berger et al., 2011). **Omics datasets:** GSE84483 (ChIP-seq of 30-h-old gBPC1-Myc in Col-0 accession) (Xiao et al., 2017b).
val1 (*hsi2*)	**Dormancy and germination:** Similar to WT (Chen et al., 2020). **Storage compounds:** Embryonic phenotypes such as cotyledon-like organs and ectopic embryos on leaves displayed by 23% of seedlings treated by the GA-biosynthesis inhibitor, paclobutrazol (Suzuki et al., 2007). **Callus and SE:** A small % of callus observed at cotyledon margin (only in *val1–1* in the WS background) (Suzuki et al., 2007).	**Selected marker genes:** Derepression of *CRC* and *2S2* in seedlings, and derepression of *LAFL*, *L1L*, *CRC*, and *2S1* in seedlings rescued from embryos 9 days after pollination (Suzuki et al., 2007; Jia et al., 2013). Depression of *LEC1*, *FUS3*, *ABI3*, *AGL15*, and *DOG1*, and reduced H3K27me3 level at *LEC1*, *ABI3*, *AGL15*, and *DOG1* loci in *hsi2–2*. ChIP-PCR showed that *AGL15* and *DOG1* are direct targets of HSI2 (Veerappan et al., 2012, 2014; Chen et al., 2018, 2020). **Omics datasets:** ATH1 microarray from 5-day-old seedlings of Col-0, *val1–2* and *val2–1*, and 7.5-day-old seedlings of *val1–2 val2–1* double mutants) (Suzuki et al., 2007).

(Continued)

TABLE 1 Continued

OE/ mutant	Phenotypes	Genetic and genomic insights related to seeds and seedlings
		GSE74692 (RNA-seq time course of *val1* and wild-type developing seeds) (Schneider et al., 2016).
val1 val2 (*hsi2 hsl1*)	**Dormancy and germination:** Reduced and delayed germination (Suzuki et al., 2007; Chen et al., 2020). **Storage compounds:** Over-accumulation of storage reserves in seedlings (Suzuki et al., 2007; Tsukagoshi et al., 2007). **Callus and SE:** Embryonic callus found both in shoot and root (Suzuki et al., 2007). Prevalent formation of callus if developing seeds of *val1 val2* are rescued on MS plates (Jia et al., 2013).	**Selected marker genes:** Stronger derepression of *LAFL*, *L1L*, and seed storage genes in *val1 val2* seedlings grown from mature seeds or rescued embryos than in *val1* seedlings (Suzuki et al., 2007; Jia et al., 2013). Derepression of *LEC1*, *LEC2*, *FUS3*, and seed storage genes in seedlings since 4 days after imbibition, the magnitude of which is elevated by sucrose (Tsukagoshi et al., 2007). VAL1 and VAL2 binding and correlation with K27me3 at selected loci (*DOG1*) by ChIP-PCR in seedlings (Chen et al., 2020). **Omics datasets:** ATH1 microarray (see description in the *val1* row in this table) GSE119715 (RNA-seq of 14-day-old *val1*, *val2*, *val1 val2*, and wild-type seedlings). GSE145387 (ChIP-seq of VAL1-GFP and VAL2-GFP in wild-type background, and ChIP-seq of H3K27me3 of *val1 val2* and wild-type samples using 14-day-old seedlings). GSE159499 (ChIP-seq of SWN-GFP in *val1 val2* or wild-type background using 14-day-old seedlings) (Yuan et al., 2021).
vil1 (*vrn5*)	**Dormancy and germination:** Hypersensitive to ABA, delayed germination and cotyledon greening, and reduced root length on ABA plates (Zong et al., 2022).	**Selected marker genes:** VIL1 binds to *ABI3*, *ABI4*, and *RD29*. In the *vil1* mutant, these genes show upregulated transcript abundance and reduced H3K27me3 (Zong et al., 2022). **Omics datasets:** GSE180587 (RNA-seq of 1-day-old *vil1* and Col-0 seedlings) shows that DE genes are enriched for genes related to ABA responses and seed germination programs (Zong et al., 2022); PRJNA973989 (ChIP-seq of VELs before and after vernalization) (Franco-Echevarría et al., 2023).
vil1 swn		**Selected marker genes:** Further upregulation of *ABI3* and *ABI4* compared to vil1 single mutant (Zong et al., 2022).
pkl	**Dormancy and germination:** *pkl* mutant is hypersensitive to ABA, and shows delayed germination and cotyledon greening on ABA plates (Perruc et al., 2007). **Storage compounds:** Low penetrance of storage lipids and proteins at the root tip, and the trait is enhanced by GA biosynthesis inhibitors (Ogas et al., 1997). **Callus and SE:** Embryogenic callus formed from explants of pickle roots, cotyledons, and hypocotyls from *pkl* plants (Ogas et al., 1997; Henderson et al., 2004).	**Selected marker genes:** Increased and prolonged accumulation of *ABI3* transcript and proteins in imbibed *pkl* seeds and young seedlings (Perruc et al., 2007). Elevated master TF expression in seeds (*FUS3* and *ABI3*) and seedlings (*LAFL*) (Shen et al., 2015). **Omics datasets:** GSE103361 (see description in the *clf* row in this table) GSE186152 (RNA-seq of 14-day-old wild-type and *pkl* seedlings), GSE186156 (RNA-seq of 14-day-old wild-type and *val1 val2 pkl* seedlings), and GSE186157 (ChIP-seq of PKL in 14-day-old wild-type and *val1 val2* seedlings) (Liang et al., 2022).
pkl pkr2	**Storage compounds:** *pkr2*. but not *pkr1*, increases penetrance of *pkl* root phenotype (Aichinger et al., 2009).	**Selected marker genes:** Enhanced/synergistic de-repression of *LEC1, FUS3*, and *ABI3* in *pkl pkr2* (no *LAFL* derepression in *pkl2*), but no PKL association/binding in CHIP; decreased H3K27me3 at *LEC1* and *ABI3*, but not *FUS3* (Aichinger et al., 2009). **Omics datasets:** E-MEXP-2140 (microarray of root tips of 5-day-old *pkl*, *pkr2*, *pkl pkr2*, and wild-type seedlings) (Aichinger et al., 2009).
hda19		**Selected marker genes:** *LEC1*, *LEC2*, *ABI3*, and multiple seed storage genes, but not *FUS3* are derepressed in 14-day-old *hda19* seedlings; increased active histone marks and reduced repressive histone marks at the derepressed loci. Binding of HDA19 to *LEC1* and *LEC2* (Zhou et al., 2013). Increased levels of H3ac, H4ac, and H3K4me3 active marks, and decreased level of H3K9me2 and H3K27me3 repressive marks in *7S1*, *OLE1*, *ABI3*, *CRA1*, and *LEC2* in *hda19–1*. (Zhou et al., 2020). **Omics datasets:** Fourteen-day-old wild-type and hda19 seedlings were profiled by ChIP-seq for H3K14ac and H3K9me2 (Zhou et al., 2013).

(Continued)

TABLE 1 Continued

OE/ mutant	Phenotypes	Genetic and genomic insights related to seeds and seedlings
		GSE166090 (RNA-seq and H3ac ChIP-seq of 14-day-old *hda19, hda6*, and wild-type seedlings) (Feng et al., 2021).
hda6 hda19	**Dormancy and germination:** Post-germinative growth arrest (no cotyledon expansion or greening) in ~70% *hda6*/*hda19*:RNAi double repression seedlings (Tanaka et al., 2008). *hda6* RNAi and *axe5–1* (*hda6* splice mutant) are hypersensitive to ABA and salt stress (Chen et al., 2010b). Delayed germination and seedling growth arrest induced by HDAC inhibitor Trichostatin A (TSA) is not affected by GA (Tanaka et al., 2008). **Storage compounds:** Increased accumulation of storage compounds inferred from upregulation of *LEC1, ABI3*, and *FUS3* and upregulation of seed storage genes in TSA-treated seedlings that exhibit similar phenotypes to hda6/19:RNAi (Tanaka et al., 2008). **Callus and SE:** Embryo-like structures on *hda6/19* RNAi shoot (Tanaka et al., 2008).	**Selected marker genes:** *LEC1, FUS3*, and *ABI3* de-repression in *hda6/19* RNAI (Tanaka et al., 2008). *CRA/B/C* are derepressed in TSA-induced somatic embryos (Tanaka et al., 2008). *CRA1, OLE1, 2S2*, and *7S1* are derepressed in *hda19–1* seedlings (Zhou et al., 2020).
hda19 val2	**Dormancy and germination:** Embryo lethal (Zhou et al., 2013).	
sdr4l (sfl1/odr1)	**Dormancy and germination:** Delayed germination of mature seeds (Cao et al., 2020; Liu et al., 2020; Wu et al., 2022; Zheng et al., 2022). **Storage compounds:** Increased accumulation of storage lipids (Wu et al., 2022; Zheng et al., 2022).	**Selected marker genes:** *ABI3, FUS3*, and *DOG1* are upregulated in 15-DAP *Atsdr4l-1* seeds (Zheng et al., 2022). *LAFL* and *DOG1* are derepressed in 4-day-old *Atsdr4l-3* and *Atsdr4l-4* seedlings (Wu et al., 2022). *DOG1* is upregulated in maturing *Atsdr4l-1* and *Atsdr4l-2* seeds at 12–18 DAP (Cao et al., 2020). ABA biosynthesis genes *NCED6* and *NCED9* have elevated expression in *Atsdr4l-2* seeds harvested at fresh (Liu et al., 2020). **Omics datasets:** PRJNA663767 (RNA-seq of 4-day-old *Atsdr4l-3, Atsdr4l-4*, and Col-0 seedlings); GSE185388 (ChIP-seq of 4-day-old estradiol induced AtSDR4L seedlings) (Wu et al., 2022). GSE246997 (ChIP-seq of AtSDR4L expressed from its native promoter in ABA-treated 1-DAI seedlings, and H3K27me3 in 1-DAI and 3-DAI Col-0, *Atsdr4l-4, Atsdr4l-5* seedlings grown with 1% sucrose) (Lu et al., 2024).
sfl1 sfl4 (sdr4l dig2)	**Dormancy and germination:** Severely reduced germination (Zheng et al., 2022). **Storage compounds:** Increased accumulation of storage lipids (Zheng et al., 2022).	**Selected marker genes:** *LAFL* genes exhibit dynamic expression over the course of seed maturation in *Atsdr4l dig2* mutant: *LEC1* and *LEC2* have comparable expression levels between the double mutant and wild type from 9 to 18 DAP; *ABI3* expression is downregulated in 12-DAP, but upregulated in 15- and 18-DAP mutants seeds; *FUS3* exhibits lower expression than wildtype in 9-DAP, but higher in 12-, 15-, and 18-DAP mutant seeds. *DOG1* is downregulated in the mutant at 12 DAP, and upregulated at 15 and 18 DAP (Zheng et al., 2022).

terminating the embryonic program during seed-to-seedling transition. Table 1 and Figure 1 summarize the phenotypes of these mutants and dissect the regulation by the general epigenetic machinery (e.g. core proteins in the PRCs and TrxG) that affects all stages of plant development, as well as developmental stage-specific epigenetic regulators, including the facultative accessory protein of PRCs. Although there is extensive overlap in mutant phenotypes, some epigenetic mutants only show a subset of phenotypes and with variable penetrance. This is because some PRC2 complex subunits are present in all complexes, e.g., FIE and MSI, while others are specific for some developmental stages (e.g., EMF2, VRN2, FIS2, and MEA), and/or their mutant phenotypes are so severe (embryo lethality) that their roles in other developmental stages can be difficult to uncover (FIE, FIS2, and MSI1). For instance, double mutants lacking PRC2 paralogs that are involved in vegetative development and flowering (*swn clf* and *emf2 vrn2*), show delayed germination (*swn clf*) and develop embryonic callus (*swn clf* and *emf2 vrn2*). FIE is a single subunit required in all PRC2 complexes and *fie* mutants are not viable, due to endosperm over-proliferation and embryo arrest at the heart stage. Similar phenotypes are shown by *prc2* mutants, *fis2*, *mea*, and *msi1* (Meinke, 2020). However, evaluation of embryonically-rescued *fie* plants allowed for the discovery of its role in seed-to-seedling transition. Indeed, embryonically rescued *fie* lines displayed delayed germination, somatic embryos, and embryonic callus-like structures. Embryonic callus also develops in double mutants lacking PRC1 paralogs (*atbmi1a atbmi1b* and *atring1a atring1b*), as well as accessory proteins that promote histone deacetylation (*hda6 hda19 RNAi*) or recruit PRCs (*val1 val2*), albeit with different frequencies. The penetrance of the embryonic callus phenotype can be further increased in mutants simultaneously lacking core subunits of PRC1 and PRC2 (*atring1a atring1b clf*) or in the absence

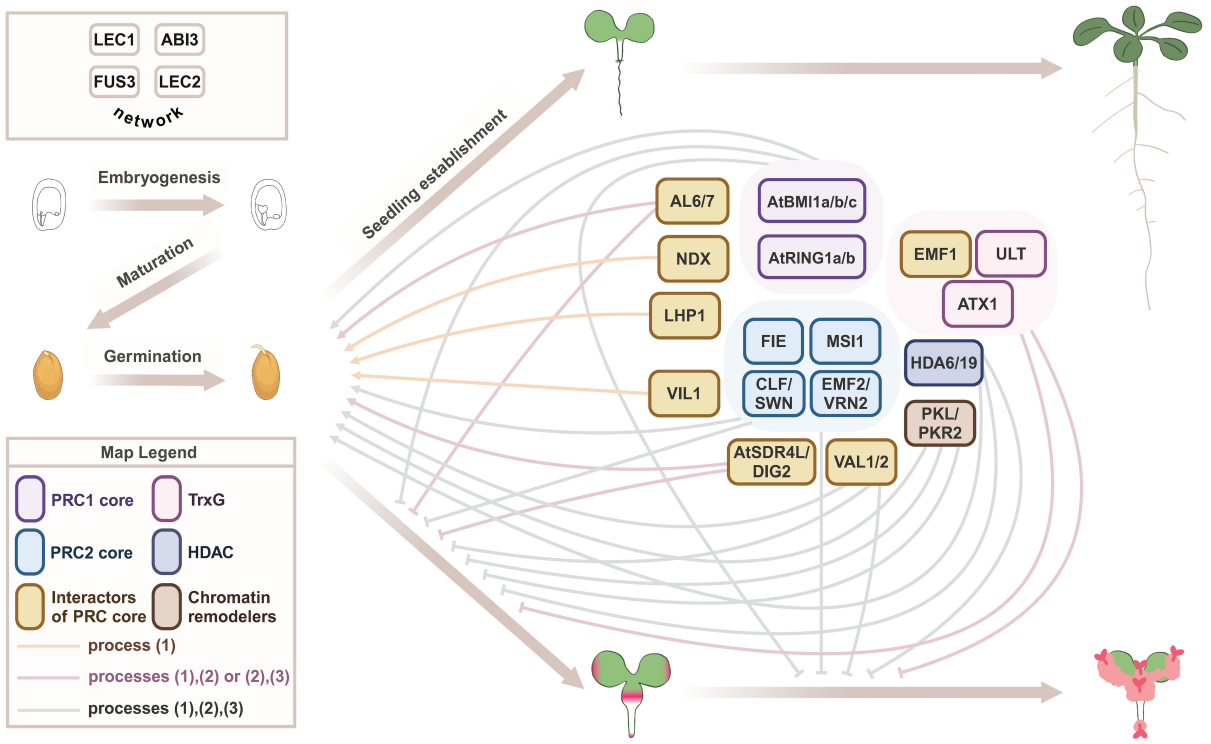


FIGURE 1
Epigenetic regulators that repress the seed maturation program during seed-to-seedling transition. A proposed model on the regulation of early developmental phase transitions by general and stage-specific epigenetic machinery in Arabidopsis (Table 1). Embryogenesis and maturation phases of seed development are regulated by the LAFL network (top left). The seed-to-seedling phase transition comprises the break of dormancy, germination, and seedling establishment. This normal developmental trajectory requires repression of the seed maturation program by a suite of epigenetic regulators. Loss-of-function mutations in these epigenetic regulators often lead to an alternative developmental trajectory, with mutant seeds exhibiting defects in one or more processes: (1) germination, (2) repression of post-embryonic accumulation of storage compounds, with pink color indicating the accumulation of storage lipids, and (3) repression of subsequent formation of callus and somatic embryos. Severe mutants of general epigenetic machinery usually show all three types of defects, whereas the role of stage-specific regulators may be limited to one or two processes.

of PRC2 and chromatin remodeling factors (*swn clf pkl*) (Table 1). Thus, a stable shutdown of *LAFL* and the embryonic program during vegetative development requires a suite of epigenetic regulators.

Interestingly, mutants in accessory proteins such as *al6 al7*, *lhp1*, and *vil1* affect germination, dormancy, and sensitivity to ABA and stress, and show repression of late-acting *LAFL* such as *ABI3*, suggesting more stage- and context-specific roles for these accessory proteins. While TrxG and PcG play opposite roles in regulating flowering time through FLC (Pien et al., 2008), TrxG's synergistic role with PRC in transcriptional repression during seed-to-seedling transition does not conform to the norm of TrxG in exerting transcriptional activation. The TrxG mutants *atx* and *ult* strongly enhance the phenotype of *emf1*, with *emf1 atx ult* showing swollen roots and embryo- and callus-like structures (Xu et al., 2018). Similarly, the TrxG homolog SDG8 acts synergistically with PRC2 EMF2 in repressing seed maturation genes, as shown by the development of somatic embryos in *emf2 sdg8* (Table 1). Altogether, *LAFL* and downstream seed maturation genes serve as excellent models for comprehending the repression of the seed maturation program during the transition from seed to seedling, similar to *FLC* for the transition from vegetative to reproductive phases (Whittaker and Dean, 2017).

Omics studies facilitate a comprehensive understanding of seed-to-seedling transition

Mutant phenotypes and marker genes are powerful tools to associate regulation with a specific developmental stage. These tools together with omics studies allow for the efficient and comprehensive characterization of biological processes. Transcriptomic and epigenomic datasets confirmed the observation that general epigenetic machinery participates in all major development transitions, and their regulatory specificity is often determined by accessory proteins and interacting transcription factors (Merini et al., 2017; Xiao et al., 2017b). Starting from profiling canonical histone marks such as H3K27me3 (Table 1), our understanding of the epigenetic regulation of plant phase transition has been substantially advanced through the integration of genetic, biochemical, cell

biology, and multi-omic results. Here, we discuss some representative examples of how transcriptomic and epigenomic data enable a quantitative assessment of regulatory specificity, suggest interactions between regulators, and uncover crosstalk between regulatory machinery.

Genome-wide comparisons reveal regulatory specificity of PRC components

Both PRC1 and PRC2 contribute to the repression of the embryonic and seed maturation programs (Table 1). Omics studies have strengthened the observation originally made based on marker genes that PRC1 and PRC2 have both shared and unique functions (Wang et al., 2016; Zhou et al., 2017). Furthermore, multi-omic profiling, including chromatin accessibility, H2AK121ub, H3K27me3, and transcriptome of mutants of core and accessory components of PRC1 and PRC2 and wild-type seedlings have shown that PRC1 regulates chromatin accessibility (Yin et al., 2021), and defined PRC1-dependent and -independent repression by H3K27me3 (Kralemann et al., 2020; Yin et al., 2021).

Genome-wide profiles also aid in the characterization of the accessory proteins, which are crucial for general machinery such as PRCs to exert specific roles in plant development and stress responses. The association of LHP1 with PRC2 was supported by extensive overlap between H3K27me3 and genome-wide binding of LHP1 (Turck et al., 2007; Zhang et al., 2007). Additionally, omics assays help to identify or rule out regulators at specific developmental stages. Cross-comparison of differentially expressed (DE) genes and H3K27me3 in the mutants of *lhp1* and PRC core components showed that LHP1 regulates vegetative-to-reproductive transition, but lacks a major role in seed-to-seedling transition (Wang et al., 2016). Furthermore, the accumulation of H2AK121ub is similar between *lhp*1 and wild-type seedlings (Zhou et al., 2017). Collectively, despite the physical association of LHP1 with both PRC1 (Xu and Shen, 2008) and PRC2 (Hecker et al., 2015), the binding of LHP1 to the dormancy promoting loci, ABI3 and DELAY OF GERMINATION 1 (DOG1) (Molitor et al., 2014), and the modest upregulation of DOG1 in lhp1 mutant (Chen et al., 2020) showed a connection to PRC1 and a minor role in germination, with genome-wide evidence indicating that LHP1 mainly functions after the seed-to-seedling transition as a PRC2 accessory protein.

Genome-wide comparisons support functional redundancy and protein–protein interactions

As discussed above, the epigenetic machinery functions through multi-protein complexes. ChIP-seq has been widely used to examine target sites of epigenetic regulators. The sheer number of binding sites across the genome provides numerous data points to assess binding similarities of epigenetic regulators and infer functional redundancy and/or protein–protein interactions (PPIs). In the case of PRC core proteins CLF and SWN, the nearly identical binding patterns (Shu et al., 2019), synergistic phenotype of the *clf swn* double mutant (Chanvivattana et al., 2004), and the absence of data showing their physical interaction indicate that these two methyltransferases function redundantly in different variants of PRC2 core complexes. By contrast, genomic data can also be used as supporting evidence in functional characterization of VAL proteins. Both VAL1/HSI2 and VAL2/HSL1 bind to the RY motif (CATGCA/TGCATG), and they homo- and heterodimerize via the PHD-L domain (Chen et al., 2020). The physical interaction and association with the same CRE are further supported by the extensive overlap of VAL1 and VAL2 across the genome (Yuan et al., 2021). Similarly, the physical interaction of the VAL proteins with the PRC2 core components SWN and CLF and with the chromatin remodeler PKL was demonstrated by both PPI assays and genome-wide binding similarities (Yuan et al., 2021; Liang et al., 2022).

Genome-wide comparisons suggest crosstalk of epigenetic machinery

The seedling is arguably one of the most vulnerable stages of a plant's life cycle. Germinating seeds must strike a balance between preserving limited resources to survive uncertain weather patterns in spring and fall and the rapid consumption of storage compounds to establish themselves and outcompete nearby seedlings. To cope with these two seemingly conflicting priorities, master TFs of seed maturation such as *ABI3* and *FUS3* remain inducible by abiotic stresses during the early stages of germination, while these TFs and the seed maturation program are robustly repressed within a few days after germination if environment is favorable. The robust repression requires coordination of various regulators. For instance, the repression involves a transient increase of histone deacetylase activity soon after germination (Tai et al., 2005), reduced accumulation of active histone marks and increased accumulation of repressive marks at seed maturation and dormancy loci within the first 3 days of germination (Yang et al., 2022a; Pan et al., 2023a), and participation of histone variants (Zhao et al., 2022b). Consistent with the multifaceted regulation, higher-order mutants defective in multiple epigenetic machinery often exhibit more severe phenotypes in germination and seedling establishment (Table 1). The coordination of gene repression is often facilitated by PPI. For instance, VAL1 serves an interaction hub to unite the HDAC and PRC activities (Baile et al., 2021; Mikulski et al., 2022), presumably to reduce the level of active marks such as H3ac, enhance the level of repressive histone marks, and limit chromatin accessibility through H2AK121ub and H3K27me3 (Mikulski et al., 2022). Besides VAL proteins, other TFs that possess an EAR motif can also recruit HDAC and enhance H3K27me3 marking through their physical interaction with TPL or SAP18 (Baile et al., 2021). Crosstalk between repressive machinery has been revealed by omics assays. For instance, LHP1 interacts with ATRX, a chromatin remodeler that deposits histone variant H3.3. The intersection of LHP1 target genes with DE genes in *atrx* mutant connects repressive histone marks with histone variants (Wang et al., 2018). Another example is the potential crosstalk between PRC and constitutive heterochromatin in the pericentromeric regions. NDX was discovered as a PRC1-associated protein that regulates ABA sensitivity (Zhu et al., 2020). Recently, genome-wide profiling revealed that NDX binds to heterochromatic small RNA loci and affects non-CG DNA methylation (Karányi et al., 2022), suggesting a potential connection between PRC1 with constitutive heterochromatin and chromatin topology.

A new puzzle piece: *Sdr4* family in model and crop species

AtSDR4L/ODR1/SFL1 is a nuclear-localized, plant-specific transcriptional corepressor that is devoid of known DNA-binding domains (Moon et al., 2016; Subburaj et al., 2016; Cao et al., 2020; Liu et al., 2020; Wu et al., 2022). Here, we review the functional characterization of AtSDR4L based on the features summarized in Table 1, Figure 1, and omics tools discussed in the previous section.

AtSDR4L and its paralogs are novel corepressors in Arabidopsis seed-to-seedling transition

The role of AtSDR4L is specific to the seed-to-seedling transition, and its expression increases during seed maturation, peaks in dry seeds, and decreases upon imbibition, subsequent to the expression patterns of *LAFL* that primarily span from embryogenesis to seed maturation (Stone et al., 2001; Baumbusch, 2003; Cao et al., 2020). Loss-of-function mutants of *Atsdr4l* share many phenotypic and molecular characteristics with mutants listed in Table 1. Mature seeds of *Atsdr4l* are more dormant (Cao et al., 2020; Liu et al., 2020; Wu et al., 2022; Zheng et al., 2022). *Atsdr4l* seedlings exhibit stunted growth, with seed storage compounds accumulating to various degrees depending on exogenous sucrose and the duration of after-ripening and cold stratification (Wu et al., 2022; Zheng et al., 2022). A large number of seed maturation genes are upregulated in *Atsdr4l* seedlings, and AtSDR4L binds to the upstream region of a subset of these, including *LEC1* and *ABI3* (Wu et al., 2022). Furthermore, AtSDR4L physically interacts with VAL2, and H3K27me3 at a distal regulatory region upstream of *ABI3* is decreased in 3-day-old *Atsdr4l* seedlings (Lu et al., 2024). The lack of strong dedifferentiation phenotypes, such as those shown in *LEC1* and *LEC2* OE and mutants of PRC core components, indicates that not all genes required for the formation of callus-like structures and somatic embryos are misregulated. *Atsdr4l* seedlings resemble *ML1:FUS3*, suggesting that *FUS3* could be an indirect target activated by elevated *ABI3*. Collectively, these data suggest that AtSDR4L functions together with VAL2 to recruit PRC2 to directly or indirectly repress *LAFL* and other seed maturation genes.

A recent study suggested that AtSDR4L shares partial functional redundancies with its paralogs to form a repressive module in Arabidopsis (Zheng et al., 2022). The paralogs are collectively known as Dynamic Influencer of Gene expression (DIGs) and DIG-like (DILs)/ABA-inducible transcriptional repressors (AITRs)/Seed dormancy Four-Like (SFLs) (Song et al., 2016; Tian et al., 2017; Zheng et al., 2022). Similar to AtSDR4L, DIGs and DILs are nuclear localized (Song et al., 2016; Tian et al., 2017) and physically interact with VAL2 (Lu et al., 2024). An *sfl1 sfl4* (*Atsdr4l dig2*) double mutant shows strongly enhanced seed dormancy and embryonic traits in seedlings compared to *Atsdr4l*, suggesting synergy between AtSDR4L and its paralogs (Zheng et al., 2022). On the other hand, *sfl2 sfl3 sfl4* (*aitr2 aitr6 aitr5*, *dig1 dil1 dig2*) triple mutant seeds exhibited reduced dormancy when they were freshly harvested from siliques at 24 days after flowering (DAF), suggesting antagonistic interaction between AtSDR4L and its paralogs at certain developmental stages (Zheng et al., 2022). Additionally, triple mutant seedlings are hyposensitive to ABA and resistant to drought (Tian et al., 2017), while seedlings overexpressing *DIG1* or *DIG2* are hypersensitive to ABA and salt (Song et al., 2016). These lines of evidence suggest that AtSDR4L and its paralogs may have context-specific functions that require further investigation.

Sdr4 prevents pre-harvest sprouting in rice

Sdr4 is a major quantitative trait locus and a positive regulator for seed dormancy in rice (Sugimoto et al., 2010; Zhao et al., 2022a). Rice *Sdr4*, herein referred to as *OsSdr4*, is expressed in the embryo and the protein is localized to the nucleus (Sugimoto et al., 2010). *OsSdr4* transcripts begin to accumulate after 7 DAF and increase as seed ripens (Sugimoto et al., 2010). The expression control of *OsSdr4* and its orthologs in wheat, a monocot, and Arabidopsis, a dicot, seems well conserved (Figure 2A). Upstream regulatory sequences of *OsSdr4* and orthologs contain multiple RY and G-box (CACGTG) motifs, and binding by ABI3/VP1 and bZIP TFs were shown by *in vitro* or *in vivo* assays in multiple species (O'Malley et al., 2016; Tian et al., 2020; Chen et al., 2021; Wu et al., 2022; Liu et al., 2024). Knocking down or knocking out *OsSdr4* leads to PHS (Sugimoto et al., 2010; Zhao et al., 2022a). In contrast, *Atsdr4l* mutant seeds harvested at maturity exhibit delayed germination (Cao et al., 2020; Liu et al., 2020; Wu et al., 2022). The seemingly opposite mutant phenotypes of seed germination between rice and Arabidopsis may be attributable to the downstream target genes of Sdr4 (Figure 2B). In accordance with elevated PHS of *sdr4* mutants in rice, the expression of a gibberellin biosynthesis gene, *OsGA20ox-1*, is significantly upregulated and an ABA-responsive gene, *OsLEA3*, is significantly downregulated compared to wild-type seeds (Sugimoto et al., 2010; Chen et al., 2021; Zhao et al., 2022a). A few other *OsLEA* genes and a dormancy regulator *OsMFT2* are downregulated in the rice *sdr4* mutant as opposed to the upregulation of their orthologs in the *Atsdr4l* mutants (Wu et al., 2022; Zhao et al., 2022a). Interestingly, a recent spatiotemporal single-cell transcriptomic profile of germinating rice embryo revealed that both *OsSdr4* and *OsMFT2* are expressed in the scutellum, and share a similar temporal expression pattern as the transcript abundance of both genes sharply decreases after imbibition (Yao et al., 2024). While mature *Atsdr4l* seeds are more dormant, mutant seeds prematurely harvested at 14 DAF germinate better than wild-type seeds (Zheng et al., 2022), suggesting that temporal-specific regulation may also contribute to the phenotypic difference between rice and Arabidopsis mutants. Storage reserve genes were upregulated in *Atsdr4l* seedlings in Arabidopsis whereas seed storage catabolism genes are upregulated in *Ossdr4* seeds in rice (Wu et al., 2022; Zhao et al., 2022a). The major forms of storage reserves in Arabidopsis are lipids and storage proteins that are

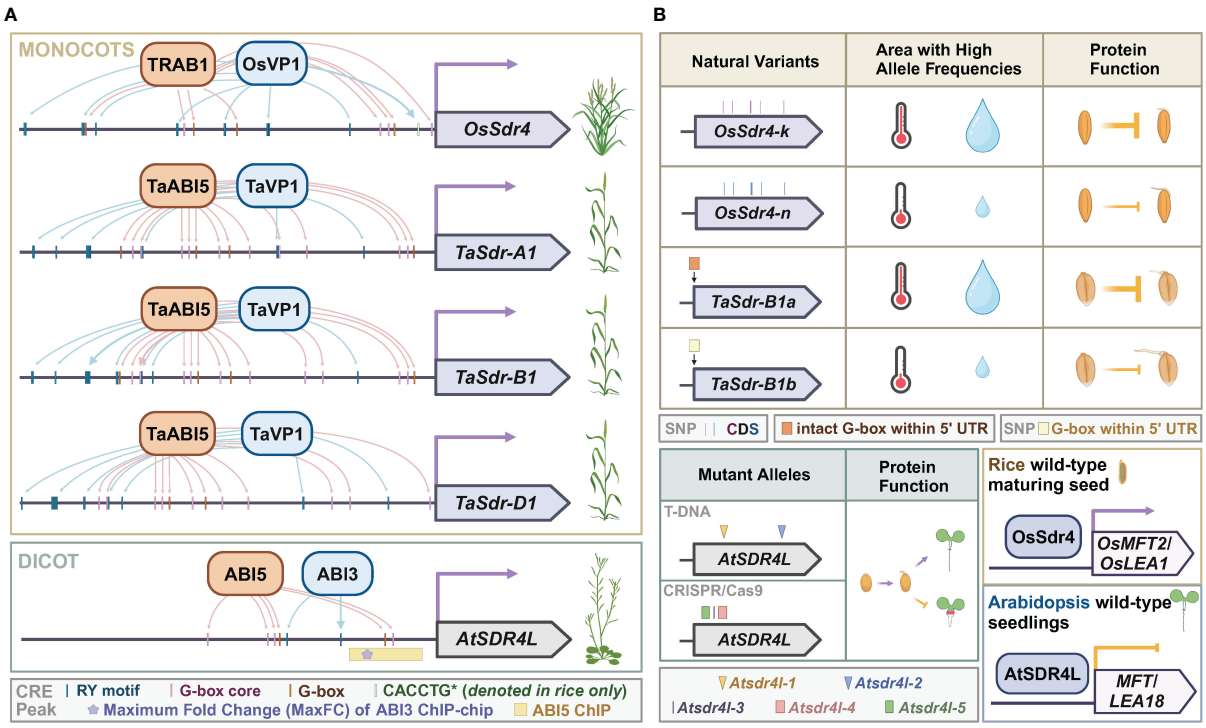


FIGURE 2
Sdr4 and its orthologs in monocots and dicots. **(A)** Putative *cis*-regulatory motifs of *OsSdr4* in rice, *TaSdrs* in wheat, and *AtSDR4L* in Arabidopsis from 1 bp to 1,500 bp upstream of the corresponding transcription start site. TF binding sites with experimental evidence are indicated by thick arrows protruding from the regulators of *SDR4*. Data were summarized from O'Malley et al. (2016); Tian et al. (2020); Chen et al. (2021); Wu et al. (2022), and Liu et al. (2024). **(B)** Top: occurrence of natural variants of *OsSdr4* and *TaSdr* in various climates, and their repression of preharvest sprouting. Bottom left: Mutant alleles of *Atsdr4l*, and function of wild-type AtSDR4L in promoting germination and preventing fatty acid over-accumulation. Bottom right: the upregulation of the dormancy-related gene, *MFT*, and the seed reserve-related gene, *OsLEA*, by Sdr4 and the downregulation of these orthologs by AtSDR4L. Data were summarized from Sugimoto et al. (2010); Zhang et al. (2014); Wu et al. (2022), and Zhao et al. (2022a).

deposited in cotyledons, and most of the endosperm except the peripheral endosperm layer in Arabidopsis is consumed by the embryo for nutrient uptake during seed maturation (Sreenivasulu and Wobus, 2013; Doll and Ingram, 2022). In contrast, most of the endosperm in Poaceae is retained and accumulates a substantial amount of starch and storage proteins, followed by programmed cell death without full degradation (Sreenivasulu and Wobus, 2013). Thus, the contrasting role of Sdr4 and its annotated orthologs in grasses and dicots may be associated with opposite regulation of key downstream genes, different types of the major forms of storage compounds, and a higher embryo-to-endosperm ratio in Arabidopsis than that in wheat and rice. In summary, Sdr4 homologs in rice and Arabidopsis share similar expression patterns in seeds and the nuclear localization. However, their function in regulating dormancy is species-, developmental stage-, and tissue-dependent.

Allelic variations of *Sdr4* correlate with regional weather patterns

Allelic variants of *Sdr4* and its orthologs in coding and regulatory sequences are associated with quantitative differences in seed dormancy, germination, and post-germinative growth. In Arabidopsis, developmental defects include inhibited root growth, swollen hypocotyl, and excess buildup of storage lipids. These defects are more severe in the CRISPR/Cas9 frameshift or segmental deletion mutants of *Atsdr4l-3*, *Atsdr4l-4*, and *Atsdr4l-5* than in the T-DNA insertion mutants of *Atsdr4l-1* and *Atsdr4l-2*, possibly because the mutations reside in the CRISPR/Cas9 lines are in the closer downstream of *AtSDR4L* start codon (Wu et al., 2022; Lu et al., 2024) (Figure 2B). In rice, a nearly isogenic line of *OsSdr4* [NIL(*Sdr4*)], in which the genomic segment containing *OsSdr4* from the Kasalath (*indica*) group was introgressed into a Nipponbare (*japonica*) background, had a substantially lower rate of seed germination than that of Nipponbare, demonstrating that the Kasalath allele of *Sdr4* (*OsSdr4-k*) confers deeper dormancy than the Nipponbare allele (*OsSdr4-n*) (Sugimoto et al., 2010) (Figure 2B). The amino acid sequences of *OsSdr4-k* and *OsSdr4-n* alleles differ by approximately 10 amino acid residues, which could potentially affect OsSdr4's characteristics as a cofactor, thus changing binding behaviors to downstream target loci. The *japonica* group has only the *OsSdr4-n* allele, whereas the *indica* group has both *OsSdr4-k* and *Sdr4-n*. Analysis of SNPs flanking the *OsSdr4-n* locus in the *indica* cultivars indicated their *OsSdr4-n* allele was introgressed from the *japonica* group. A subsequent larger-scale study revealed a correlation between allele frequency and

weather patterns (Zhao et al., 2022a). Allele frequency of *OsSdr4-k* and sequence-similar *OsSdr4-k′* is higher in regions with high annual temperatures and precipitation, whereas *OsSdr4-n* is more prevalent in areas with lower annual temperatures and precipitation. Interestingly, different geographic distribution for *Sdr4* alleles was also reported in wheat (Zhang et al., 2014). Among three homeologs of wheat *Sdr4*, namely, *TaSdr-A1*, *TaSdr-B1*, and *TaSdr-D1*, the *TaSdr-B1b* allele carries an SNP that abolishes a G-box in the 5' UTR immediately upstream of its start codon and is associated with increased germination compared to that of *TaSdr-B1a* allele with an intact G-box. The mutation in G-box could affect the binding by bZIP and bHLH TFs t *TaSdr-B*. Between the two alleles, *TaSdr-B1a* is likely positively selected for resistance to PHS, since the allele frequency of the *TaSdr-B1a* is high in areas that are more susceptible to severe PHS and low in areas with reduced rainfall and less damage by PHS. These observations in rice and wheat suggest that selection for *Sdr4* alleles best adapted to local climates is a shared feature for the adaptation of staple grains.

transcriptional activation, as well as VAL1/HSI2 and VAL2/HSL1 for transcriptional repression (Reidt et al., 2000; Nakabayashi et al., 2005; Suzuki et al., 2007; Tsukagoshi et al., 2007; Suzuki and McCarty, 2008; Jia et al., 2014; Yuan et al., 2021) (Figure 3). VAL1 and VAL2 can homo- or heterodimerize to target RY motifs in the *DOG1* promoter and repress its expression in seedlings (Chen et al., 2020). However, the B3 domains of these TFs exhibit high similarity but differential binding affinity to the target CREs, with the B3 domain of LEC2 (LEC2-B3) and FUS3-B3 having greater affinity than that of VAL1-B3 and ABI3-B3 (Jia et al., 2021), and with VAL1-B3 binding more effectively than VAL2-B3 (Chen et al., 2020). These varying RY-binding efficacies are likely owing to the slightly different structural bases of the B3 domains and the absence or presence of additional domains (Jia et al., 2021). This difference may also explain the more constrained and specific roles of LEC2 and FUS3 in the establishment and maintenance of the embryonic states, as well as lending greater flexibility for ABI3 to incorporate additional cues into more complex target recognition (Jia et al., 2013, 2021; Roscoe et al., 2015).

Regulatory symmetries of activators and repressors at the same CREs for developmental transitions

Regulatory symmetry via the RY motif

The regulatory symmetry is the activation and repression of genes through the same CRE. The regulatory summary of the seed maturation programs by B3 proteins is well established (Suzuki and McCarty, 2008). The RY motif is enriched in many seed maturation genes and is often bound by the B3 TFs FUS3, LEC2, and ABI3 for

RY and G-box motifs function coordinately in the activation of seed maturation genes

Transcriptional activation of maturation genes *via* the RY motif is often coupled with the G-box motif. G-box motifs are preferentially bound by the basic leucine zipper (bZIP) and basic helix-loop-helix (bHLH) TFs, and these CREs are an overrepresented CRE in seed maturation genes (Nakabayashi et al., 2005; Mönke et al., 2012; Mendes et al., 2013; Ezer et al., 2017; Jo et al., 2020). ABI5, a bZIP TF that interacts with ABI3, was found to transactivate the promoter of *AtEm6* (Nakamura

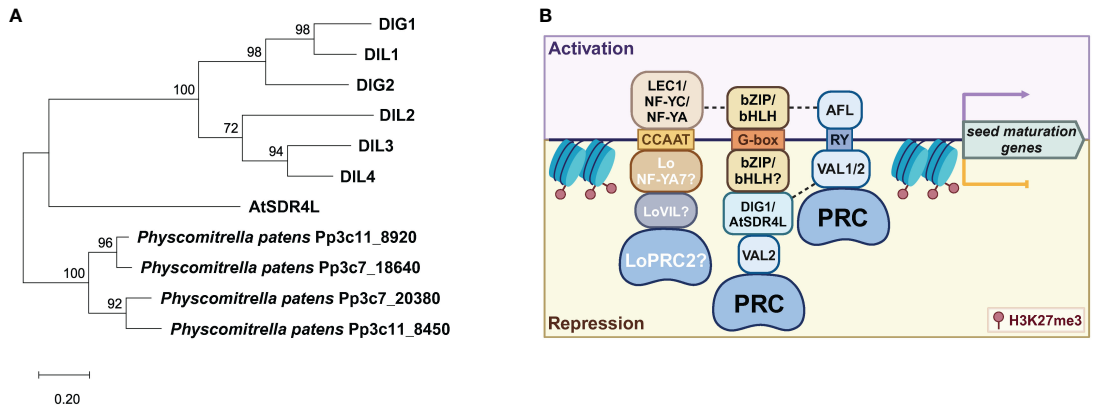


FIGURE 3
Regulatory symmetry in the activation and repression of seed maturation genes. **(A)** A phylogenetic tree of the AtSDR4L family in Arabidopsis by the maximum-likelihood method and the JTT-matrix-based model using MEGA11 (Tamura et al., 2021). AtSDR4L's orthologs in *Physcomitrella patens* were used as the outgroup (Lang et al., 2018). **(B)** A working model of the regulatory symmetry between activators and repressors at shared CREs upstream of seed maturation genes. CCAAT is bound by the NF-Y activator complex (top) in Arabidopsis or by the repressor complex of LoNF-YA, LoVIL, and PRC2 in lily (bottom). In Arabidopsis, bZIP/bHLH binding to G-box by itself upregulates the seed maturation genes whereas the G-box binding by bZIP/bHLH coupled with a repressor module of DIG/AtSDR4L-VAL-VIL-PRC2 downregulates these genes. The RY motif can be bound by ABI3/FUS3/LEC1 (AFL) and VAL1/2-PRC, for the activation and repression of the downstream genes, respectively. Dashed lines indicated potentially synergistic effects through protein–protein interaction. Data were summarized from Lumba et al. (2014); Song et al. (2016); Myers and Holt (2018); Bryant et al. (2019); Chen et al. (2020); Liu et al. (2020); Wu et al. (2022), and Pan et al. (2023b).

et al., 2001; Lopez-Molina et al., 2002). G-box-mediated transactivation of gene expression by ABI5 is indicated by the numerous downregulated genes in *abi5* dry seed and the enrichment of G-box from these repressed genes (Nakabayashi et al., 2005). Binding to the same RY-containing region upstream of *DOG1*, AFL TFs may also upregulate *DOG1* expression through collaboration with bZIP67 during seed maturation (Bryant et al., 2019). Consistently, G-box and RY motifs are highly enriched in the regulons of ABI3 and FUS3 (Mönke et al., 2012; Wang and Perry, 2013), and ABI3 is believed to induce seed maturation genes via G-box motifs that are in close proximity with RY elements (Suzuki et al., 2003; Jo et al., 2020).

G-box motif and the repression of the seed maturation program

Loci with high occurrence of RY and G-box elements are frequently associated with high occupancy of the repressive histone mark H3K27me3 in the seedlings (Wang et al., 2016; Liu et al., 2019; Baile et al., 2021). Because both AFL and VAL possess the RY-binding B3 domain, AFL may compete with VAL-PRC for the same RY sites in the regulatory regions of seed maturation genes. Emerging evidence suggests that G-box might be recognized by gene repression machinery (Figure 3). Recent studies suggest that AtSDR4L and its paralogs are important for the G-box-dependent transcriptional repression of the seed maturation genes (Song et al., 2016; Liu et al., 2020; Wu et al., 2022). G-box is enriched in AtSDR4L binding regions (Wu et al., 2022). Since AtSDR4L is devoid of known DNA binding domains, it is most likely recruited by bZIP and bHLH family TFs with sequence-specific binding activities to the G-box (Lumba et al., 2014; Liu et al., 2020). It is proposed that AtSDR4L physically interacts with bHLH57 to indirectly inhibit the expression of ABA biosynthetic genes *9-CIS-EPOXYCAROTENOID DIOXYGENASE6* (*NCED6*) and *NCED9*, thereby downregulating ABA biosynthesis to counteract seed dormancy (Liu et al., 2020).

Interestingly, the RY motif (CATGCA) is also found in high frequency in AtSDR4L target sites (Lu et al., 2024). Additionally, *AtSDR4L* and its orthologs also harbor abundant G-box and RY CREs in their own upstream regulatory regions (Figure 2). In Arabidopsis, ABI3 binds to the *AtSDR4L* promoter and upregulates its expression in the developing seeds (Wu et al., 2022; Zheng et al., 2022). In turn, AtSDR4L represses *ABI3* to shut down the seed maturation program in the seedlings, promoting the shift to the vegetative phase (Wu et al., 2022). Intriguingly, AtSDR4L also targets itself, possibly through the G-box and RY motifs in its own promoter (Lu et al., 2024). Similarly, the *OsSdr4* promoter contains seven RY and six G-box elements, and one of the RY motifs is in close proximity to a G-box (Sugimoto et al., 2010). The rice ortholog of ABI3, OsVP1, perhaps in collaboration with a rice bZIP TF TRAB1, can induce the expression of *OsSdr4* (Sugimoto et al., 2010; Chen et al., 2021). *OsSdr4* expression substantially decreases in *Osvp1* mutant embryos at the maturation stage. A similar mechanism is conserved in wheat, as TaVP1 binds to the RY motifs, and TaABI5, an ortholog of *Arabidopsis* bZIP TF ABI5, binds to the G-box in *TaSdr* promoter to transcriptionally activate *TaSdr4* (Liu et al., 2024). These conserved regulatory mechanisms on a key locus of dormancy control in both the model plant and crop species provide further elucidation of the mirrored targeting behaviors by activators and repressors.

Regulatory symmetry in the activation and repression of seed maturation genes *via* CCAAT and GAGA motifs

Symmetry of activator-repressor binding is not limited to the RY and G-box pairing. The CCAAT motif is a CRE frequently found in the promoters of many genes and specifically targeted by the Nuclear Factor Y (NF-Y) factors for gene regulation (Calvenzani et al., 2012). In *Arabidopsis* and soybean, the pioneer TF LEC1 (NF-YB factor) can bind to CCAAT box elements as a trimeric complex with the NF-YA and NF-YC subunits to activate the embryonic programs (Yamamoto et al., 2009; Jo et al., 2020) (Figure 3). Conversely, LoNF-YA7 in lily bulbs has been reported to recruit LoVIL1-PRC2 machinery to *LoCALS3* locus at CCAAT motif for H3K27me3 deposition, thereby repressing *LoCALS3* expression to promote the release of bulb dormancy (Pan et al., 2023b). Similarly, GAGA box-binding BPCs can repress *LEC2* during germination (Xiao et al., 2017b) and *FUS3* (Wu et al., 2020) in the ovule integuments and endosperm, but activates *LEC2* in the embryo (Berger et al., 2011). This is in agreement with GAGA-binding proteins in animals, which function as both activators and repressors (Berger and Dubreucq, 2012). Collectively, these results show that transcriptional activation and repression through the same CREs may be a general mechanism in the activation and repression of the seed maturation program. Regulatory robustness and specificity may be determined by functional coordination of transcriptional regulators that binds to these CREs.

Conclusions and perspectives

Perspectives and challenges to advance basic knowledge

To date, many players involved in the repression of the seed maturation program during the seed-to-seedling transition have been identified. While existing omics data are tremendously useful to understand the general machinery, they often do not fully capture the dynamics of the regulation in seed germination and seedling establishment because many profiling assays were carried out using 10- to 14-day-old seedlings (Table 1). This mismatch with the developmental stages might also miss the identification of stage-specific regulators. For instance, trichostatin A (TSA) treatment to inhibit histone deacetylase activities in 3- and 16-day-old seedlings identified distinct sets of DE genes (Chang and Pikaard, 2005; Tai et al., 2005). Presumably, some regulators of the seed-to-seedling transition might yet to be discovered due to stage-specific regulation and functional redundancy of homologous genes. Additionally, cell

biology and biochemical approaches are not always readily applicable to seeds because the seed coat serves as a physical barrier that blocks light and many chemicals. Therefore, tissue- and cell type-specific datasets are often scarce for mature seeds. Previously, transcriptome profiling of dissected developing or germinating seeds (Belmonte et al., 2013; Dekkers et al., 2013) demonstrated tissue- and stage-specific gene expression in Arabidopsis. Single-cell and spatial transcriptomics (Yao et al., 2024) will further advance our understanding of the heterogeneity of gene expression in seeds, allowing a superior statistical power to classify genes based on their expression patterns and designate marker genes to existing and new cell types, thus providing a foundation for understanding cell type-specific GRN. Additional advances may come from integrative analysis of histone modification with omics datasets such as time-course profiles of transcriptome (Narsai et al., 2017), DNA methylation (Bouyer et al., 2017; Kawakatsu et al., 2017), chromatin accessibility and non-coding RNA (Tremblay et al., 2024) during seed-to-seedling transition, and changes in chromatin topology. For instance, histone modifications have been profiled in many epigenetic mutants (Table 1). Combined with other assays such as Hi-C (Huang et al., 2021; Yin et al., 2023) and Hi-ChIP (Huang et al., 2021), these datasets help to reveal how histone marks such as H3K27me3 and H2AK121ub impact spatial genome organization by regulating local and long-range chromatin interactions. Collectively, these data elucidate the unique and shared role of PRC1 and PRC2 in the co-regulation of gene expression, and may contribute to a knowledge framework of multi-loci expression optimization and trait stacking for crop improvement.

Perspectives and challenges for knowledge transfer from model species in the lab to crops in the field

Several PRC-controlled traits, such as dormancy, stress responses, and flowering time, are related to plants' adaptation to various environments. Regulators specifically targeting these traits are likely to have immediate application value in the field. However, several gaps need to be addressed for knowledge transfer from model species to crop and from controlled laboratory environment to the field. For instance, many important crops and oilseeds are polyploid, which requires additional considerations for homeolog redundancy and subgenome dominance besides optimizing species- and lineage-specific regulation (Ramírez-González et al., 2018; Xiang et al., 2019; Khan et al., 2022). Environmental factors and plant–biotic interactions are prevalent in the field, making trade-offs important considerations to enhance plant performance. For instance, overexpression and mutant phenotypes of master TFs and general epigenetic regulators often reduce fitness, thus requiring more sophisticated engineering of these factors if increased yield under less water and fertilizer usage is the ultimate goal for crop improvement. In summary, research on transcriptional and epigenetic regulation has provided valuable insights into the phase transition from seed to seedlings, and multi-omic studies have revealed many target-specific regulations and crosstalk between regulatory machinery. Further research to identify developmental stage-specific regulators and CREs with minimized fitness trade-off holds strong potential to engineer crops that can adapt to the increasingly stressful environments associated with the increasingly volatile weather patterns from a warming climate.

Author contributions

DG: Conceptualization, Visualization, Writing – original draft. BL: Writing – original draft. MA: Writing – original draft. SG: Conceptualization, Funding acquisition, Writing – review & editing, Writing – original draft. LS: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors thank funding support from the UBC Four Year Doctoral Fellowships to BL and MA, the Natural Sciences and Engineering Research Council of Canada discovery grants (NSERC DG) to SG (NSERC-DG 480529) and LS (RGPIN-2019-05039), and the Canadian Foundation for Innovation (CFI JELF/BCKDF 38187) to LS.

Acknowledgments

The authors thank Dr. Randy Allen for comments and editing, Dr. Shao-shan Carol Huang for discussions, and Jiayi Shan for summarizing some literature. The figures were prepared using BioRender.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1416216/full#supplementary-material

References

Aichinger, E., Villar, C. B. R., Farrona, S., Reyes, J. C., Hennig, L., and Köhler, C. (2009). CHD3 proteins and polycomb group proteins antagonistically determine cell identity in arabidopsis. *PloS Genet.* 5, e1000605. doi: 10.1371/journal.pgen.1000605

Alizadeh, M., Hoy, R., Lu, B., and Song, L. (2021). Team effort: Combinatorial control of seed maturation by transcription factors. *Curr. Opin. Plant Biol.* 63, 102091. doi: 10.1016/j.pbi.2021.102091

Badia-i-Mompel, P., Wessels, L., Müller-Dott, S., Trimbour, R., Ramirez Flores, R. O., Argelaguet, R., et al. (2023). Gene regulatory network inference in the era of single-cell multi-omics. *Nat. Rev. Genet.* 24, 739–754. doi: 10.1038/s41576-023-00618-5

Baile, F., Gómez-Zambrano, Á., and Calonje, M. (2022). Roles of Polycomb complexes in regulating gene expression and chromatin structure in plants. *Plant Commun.* 3, 100267. doi: 10.1016/j.xplc.2021.100267

Baile, F., Merini, W., Hidalgo, I., and Calonje, M. (2021). EAR domain-containing transcription factors trigger PRC2-mediated chromatin marking in Arabidopsis. *Plant Cell* 33, 2701–2715. doi: 10.1093/plcell/koab139

Barabási, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* 37, 382–390. doi: 10.1038/ng1532

Baumbusch, L. O. (2003). LEC1, FUS3, ABI3 and Em expression reveals no correlation with dormancy in Arabidopsis. *J. Exp. Bot.* 55, 77–87. doi: 10.1093/jxb/erh014

Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., et al. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl. Acad. Sci.* 110, E435–E444. doi: 10.1073/pnas.1222061110

Berger, N., and Dubreucq, B. (2012). Evolution goes GAGA: GAGA binding proteins across kingdoms. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* 1819, 863–868. doi: 10.1016/j.bbagrm.2012.02.022

Berger, N., Dubreucq, B., Roudier, F., Dubos, C., and Lepiniec, L. (2011). Transcriptional Regulation of *Arabidopsis LEAFY COTYLEDON2* Involves *RLE*, a *cis*-Element That Regulates Trimethylation of Histone H3 at Lysine-27. *Plant Cell* 23, 4065–4078. doi: 10.1105/tpc.111.087866

Bieluszewski, T., Xiao, J., Yang, Y., and Wagner, D. (2021). PRC2 activity, recruitment, and silencing: a comparative perspective. *Trends Plant Sci.* 26, 1186–1198. doi: 10.1016/j.tplants.2021.06.006

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., et al. (2017). DNA methylation dynamics during early plant life. *Genome Biol.* 18, 179. doi: 10.1186/s13059-017-1313-0

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., et al. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PloS Genet.* 7, e1002014. doi: 10.1371/journal.pgen.1002014

Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J. C., and Calonje, M. (2010). Keeping cell identity in arabidopsis requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr. Biol.* 20, 1853–1859. doi: 10.1016/j.cub.2010.09.046

Braybrook, S. A., Stone, S. L., Park, S., Bui, A. Q., Le, B. H., Fischer, R. L., et al. (2006). Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc. Natl. Acad. Sci.* 103, 3468–3473. doi: 10.1073/pnas.0511331103

Brooks, M. D., Juang, C.-L., Katari, M. S., Alvarez, J. M., Pasquino, A., Shih, H.-J., et al. (2021). ConnecTF: A platform to integrate transcription factor–gene interactions and validate regulatory networks. *Plant Physiol.* 185, 49–66. doi: 10.1093/plphys/kiaa012

Bryant, F. M., Hughes, D., Hassani-Pak, K., and Eastmond, P. J. (2019). Basic LEUCINE ZIPPER TRANSCRIPTION FACTOR67 transactivates DELAY OF GERMINATION1 to establish primary seed dormancy in arabidopsis. *Plant Cell* 31, 1276–1288. doi: 10.1105/tpc.18.00892

Calonje, M., Sanchez, R., Chen, L., and Sung, Z. R. (2008). EMBRYONIC FLOWER1 participates in polycomb group–mediated *AG* gene silencing in *arabidopsis*. *Plant Cell* 20, 277–291. doi: 10.1105/tpc.106.049957

Calvenzani, V., Testoni, B., Gusmaroli, G., Lorenzo, M., Gnesutta, N., Petroni, K., et al. (2012). Interactions and CCAAT-binding of arabidopsis thaliana NF-Y subunits. *PloS One* 7, e42902. doi: 10.1371/journal.pone.0042902

Cao, H., Han, Y., Li, J., Ding, M., Li, Y., Li, X., et al. (2020). Arabidopsis thaliana SEED DORMANCY 4-LIKE regulates dormancy and germination by mediating the gibberellin pathway. *J. Exp. Bot.* 71, 919–933. doi: 10.1093/jxb/erz471

Carter, B., Bishop, B., Ho, K. K., Huang, R., Jia, W., Zhang, H., et al. (2018). The chromatin remodelers PKL and PIE1 act in an epigenetic pathway that determines H3K27me3 homeostasis in arabidopsis. *Plant Cell* 30, 1337–1352. doi: 10.1105/tpc.17.00867

Casson, S. A., and Lindsey, K. (2006). The *turnip* mutant of arabidopsis reveals that *LEAFY COTYLEDON1* expression mediates the effects of auxin and sugars to promote embryonic cell identity. *Plant Physiol.* 142, 526–541. doi: 10.1104/pp.106.080895

Chang, S., and Pikaard, C. S. (2005). Transcript profiling in arabidopsis reveals complex responses to global inhibition of DNA methylation and histone deacetylation* [boxs]. *J. Biol. Chem.* 280, 796–804. doi: 10.1074/jbc.M409053200

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z. R., et al. (2004). Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* 131, 5263–5276. doi: 10.1242/dev.01400

Chen, L., Cheng, J. C., Castle, L., and Sung, Z. R. (1997). EMF genes regulate Arabidopsis inflorescence development. *Plant Cell* 9, 2011–2024. doi: 10.1105/tpc.9.11.2011

Chen, D., Molitor, A., Liu, C., and Shen, W.-H. (2010a). The Arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. *Cell Res.* 20, 1332–1344. doi: 10.1038/cr.2010.151

Chen, L.-T., Luo, M., Wang, Y.-Y., and Wu, K. (2010b). Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response. *J. Exp. Bot.* 61, 3345–3353. doi: 10.1093/jxb/erq154

Chen, N., Veerappan, V., Abdelmageed, H., Kang, M., and Allen, R. D. (2018). HSI2/VAL1 silences *AGL15* to regulate the developmental transition from seed maturation to vegetative growth in arabidopsis. *Plant Cell* 30, 600–619. doi: 10.1105/tpc.17.00655

Chen, N., Wang, H., Abdelmageed, H., Veerappan, V., Tadege, M., and Allen, R. D. (2020). HSI2/VAL1 and HSL1/VAL2 function redundantly to repress *DOG1* expression in Arabidopsis seeds and seedlings. *New Phytol.* 227, 840–856. doi: 10.1111/nph.16559

Chen, W., Wang, W., Lyu, Y., Wu, Y., Huang, P., Hu, S., et al. (2021). OsVP1 activates Sdr4 expression to control rice seed dormancy via the ABA signaling pathway. *Crop J.* 9, 68–78. doi: 10.1016/j.cj.2020.06.005

Chiu, R. S., Saleh, Y., and Gazzarrini, S. (2016). Inhibition of FUSCA3 degradation at high temperature is dependent on ABA signaling and is regulated by the ABA/GA ratio. *Plant Signal. Behav.* 11, e1247137. doi: 10.1080/15592324.2016.1247137

Curaba, J., Moritz, T., Blervaque, R., Parcy, F., Raz, V., Herzog, M., et al. (2004). *AtGA3ox2*, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by *LEAFY COTYLEDON2* and *FUSCA3* in arabidopsis. *Plant Physiol.* 136, 3660–3669. doi: 10.1104/pp.104.047266

Dekkers, B. J. W., Pearce, S., Van Bolderen-Veldkamp, R. P., Marshall, A., Widera, P., Gilbert, J., et al. (2013). Transcriptional dynamics of two seed compartments with opposing roles in arabidopsis seed germination. *Plant Physiol.* 163, 205–215. doi: 10.1104/pp.113.223511

Doll, N. M., and Ingram, G. C. (2022). Embryo–endosperm interactions. *Annu. Rev. Plant Biol.* 73, 293–321. doi: 10.1146/annurev-arplant-102820-091838

Ezer, D., Shepherd, S. J. K., Brestovitsky, A., Dickinson, P., Cortijo, S., Charoensawan, V., et al. (2017). The G-box transcriptional regulatory code in arabidopsis. *Plant Physiol.* 175, 628–640. doi: 10.1104/pp.17.01086

Feng, C., Cai, X.-W., Su, Y.-N., Li, L., Chen, S., and He, X.-J. (2021). Arabidopsis RPD3-like histone deacetylases form multiple complexes involved in stress response. *J. Genet. Genomics* 48, 369–383. doi: 10.1016/j.jgg.2021.04.004

Ferrari, C., Manosalva Pérez, N., and Vandepoele, K. (2022). MINI-EX: Integrative inference of single-cell gene regulatory networks in plants. *Mol. Plant* 15, 1807–1824. doi: 10.1016/j.molp.2022.10.016

Finch-Savage, W. E., and Bassel, G. W. (2016). Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591. doi: 10.1093/jxb/erv490

Franco-Echevarría, E., Nielsen, M., Schulten, A., Cheema, J., Morgan, T. E., Bienz, M., et al. (2023). Distinct accessory roles of *Arabidopsis* VEL proteins in Polycomb silencing. *Genes Dev.* 37, 801–817. doi: 10.1101/gad.350814.123

Gazzarrini, S., and Song, L. (2024). LAFL factors in seed development and phase transitions. *Annu. Rev. Plant Biol.* 75, 28.1–28.30. doi: 10.1146/annurev-arplant-070623-111458

Gazzarrini, S., Tsuchiya, Y., Lumba, S., Okamoto, M., and McCourt, P. (2004). The transcription factor FUSCA3 controls developmental timing in arabidopsis through the hormones gibberellin and abscisic acid. *Dev. Cell* 7, 373–385. doi: 10.1016/j.devcel.2004.06.017

Giraudat, J., Hauge, B. M., Valon, C., Smalle, J., Parcy, F., and Goodman, H. M. (1992). Isolation of the Arabidopsis ABI3 gene by positional cloning. *Plant Cell* 4, 1251–1261. doi: 10.1105/tpc.4.10.1251

Guillotin, B., Rahni, R., Passalacqua, M., Mohammed, M. A., Xu, X., Raju, S. K., et al. (2023). A pan-grass transcriptome reveals patterns of cellular divergence in crops. *Nature* 617, 785–791. doi: 10.1038/s41586-023-06053-0

Hecker, A., Brand, L. H., Peter, S., Simoncello, N., Kilian, J., Harter, K., et al. (2015). The arabidopsis GAGA-binding factor BASIC PENTACYSTEINE6 recruits the POLYCOMB-REPRESSIVE COMPLEX1 component LIKE HETEROCHROMATIN PROTEIN1 to GAGA DNA motifs. *Plant Physiol.* 168, 1013–1024. doi: 10.1104/pp.15.00409

Henderson, J. T., Li, H.-C., Rider, S. D., Mordhorst, A. P., Romero-Severson, J., Cheng, J.-C., et al. (2004). *PICKLE* acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. *Plant Physiol.* 134, 995–1005. doi: 10.1104/pp.103.030148

Huang, Y., Sicar, S., Ramirez-Prado, J. S., Manza-Mianza, D., Antunez-Sanchez, J., Brik-Chaouche, R., et al. (2021). Polycomb-dependent differential chromatin compartmentalization determines gene coregulation in *Arabidopsis*. *Genome Res.* 31, 1230–1244. doi: 10.1101/gr.273771.120

Ikeuchi, M., Iwase, A., Rymen, B., Harashima, H., Shibata, M., Ohnuma, M., et al. (2015). PRC2 represses dedifferentiation of mature somatic cells in Arabidopsis. *Nat. Plants* 1, 15089. doi: 10.1038/nplants.2015.89

Jia, H., McCarty, D. R., and Suzuki, M. (2013). Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression. *Plant Physiol.* 163, 1293–1305. doi: 10.1104/pp.113.220988

Jia, H., Suzuki, M., and McCarty, D. R. (2014). Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks: Regulation of the seed to seedling developmental phase transition. *Wiley Interdiscip. Rev. Dev. Biol.* 3, 135–145. doi: 10.1002/wdev.126

Jia, H., Suzuki, M., and McCarty, D. R. (2021). Structural variation affecting DNA backbone interactions underlies adaptation of B3 DNA binding domains to constraints imposed by protein architecture. *Nucleic Acids Res.* 49, 4989–5002. doi: 10.1093/nar/gkab257

Jo, L., Pelletier, J. M., Hsu, S.-W., Baden, R., Goldberg, R. B., and Harada, J. J. (2020). Combinatorial interactions of the LEC1 transcription factor specify diverse developmental programs during soybean seed development. *Proc. Natl. Acad. Sci.* 117, 1223–1232. doi: 10.1073/pnas.1918441117

Junker, A., Mönke, G., Rutten, T., Keilwagen, J., Seifert, M., Thi, T. M. N., et al. (2012). Elongation-related functions of LEAFY COTYLEDON1 during the development of Arabidopsis thaliana: Characterization of the LEC1 regulon. *Plant J.* 71, 427–442. doi: 10.1111/j.1365-313X.2012.04999.x

Kagaya, Y., Okuda, R., Ban, A., Toyoshima, R., Tsutsumida, K., Usui, H., et al. (2005a). Indirect ABA-dependent regulation of seed storage protein genes by FUSCA3 transcription factor in arabidopsis. *Plant Cell Physiol.* 46, 300–311. doi: 10.1093/pcp/pci031

Kagaya, Y., Toyoshima, R., Okuda, R., Usui, H., Yamamoto, A., and Hattori, T. (2005b). LEAFY COTYLEDON1 controls seed storage protein genes through its regulation of FUSCA3 and ABSCISIC ACID INSENSITIVE3. *Plant Cell Physiol.* 46, 399–406. doi: 10.1093/pcp/pci048

Karányi, Z., Mosolygó-L, Á., Feró, O., Horváth, A., Boros-Oláh, B., Nagy, É., et al. (2022). NODULIN HOMEOBOX is required for heterochromatin homeostasis in Arabidopsis. *Nat. Commun.* 13, 5058. doi: 10.1038/s41467-022-32709-y

Kawakatsu, T., Nery, J. R., Castanon, R., and Ecker, J. R. (2017). Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biol.* 18, 171. doi: 10.1186/s13059-017-1251-x

Keith, K., Kraml, M., Dengler, N. G., and McCourt, P. (1994). fusca3: A heterochronic mutation affecting late embryo development in arabidopsis. *Plant Cell* 6, 589–600. doi: 10.1105/tpc.6.5.589

Khan, D., Ziegler, D. J., Kalichuk, J. L., Hoi, V., Huynh, N., Hajihassani, A., et al. (2022). Gene expression profiling reveals transcription factor networks and subgenome bias during *Brassica napus* seed development. *Plant J.* 109, 477–489. doi: 10.1111/tpj.15587

Kralemann, L. E. M., Liu, S., Trejo-Arellano, M. S., Muñoz-Viana, R., Köhler, C., and Hennig, L. (2020). Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable Polycomb-mediated gene repression in Arabidopsis. *Genome Biol.* 21, 144. doi: 10.1186/s13059-020-02062-8

Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801

Lepiniec, L., Devic, M., Roscoe, T. J., Bouyer, D., Zhou, D.-X., Boulard, C., et al. (2018). Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. *Plant Reprod.* 31, 291–307. doi: 10.1007/s00497-018-0337-2

Liang, Z., Yuan, L., Xiong, X., Hao, Y., Song, X., Zhu, T., et al. (2022). The transcriptional repressors VAL1 and VAL2 mediate genome-wide recruitment of the CHD3 chromatin remodeler PICKLE in Arabidopsis. *Plant Cell* 34, 3915–3935. doi: 10.1093/plcell/koac217

Liu, C., Cheng, J., Zhuang, Y., Ye, L., Li, Z., Wang, Y., et al. (2019). Polycomb repressive complex 2 attenuates ABA -induced senescence in Arabidopsis. *Plant J.* 97, 368–377. doi: 10.1111/tpj.14125

Liu, J., Deng, S., Wang, H., Ye, J., Wu, H.-W., Sun, H.-X., et al. (2016). *CURLY LEAF* regulates gene sets coordinating seed size and lipid biosynthesis. *Plant Physiol.* 171, 424–436. doi: 10.1104/pp.15.01335

Liu, S., Li, L., Wang, W., Xia, G., and Liu, S. (2024). TaSRO1 interacts with TaVP1 to modulate seed dormancy and pre-harvest sprouting resistance in wheat. *J. Integr. Plant Biol.* 66, 36–53. doi: 10.1111/jipb.13600

Liu, F., Zhang, H., Ding, L., Soppe, W. J. J., and Xiang, Y. (2020). REVERSAL OF RDO5 1, a homolog of rice seed dormancy4, interacts with bHLH57 and controls ABA biosynthesis and seed dormancy in arabidopsis. *Plant Cell* 32, 1933–1948. doi: 10.1105/tpc.20.00026

Lopez-Molina, L., Mongrand, S., McLachlin, D. T., Chait, B. T., and Chua, N.-H. (2002). ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. *Plant J.* 32, 317–328. doi: 10.1046/j.1365-313X.2002.01430.x

Lotan, T., Ohto, M., Yee, K. M., West, M. A. L., Lo, R., Kwong, R. W., et al. (1998). Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93, 1195–1205. doi: 10.1016/S0092-8674(00)81463-4

Lu, B., Alizadeh, M., Hoy, R., Zheng, R., Go, D., and Song, L. (2024). Co-repressors AtSDR4L and DIG1 interact with transcription factor VAL2 and promote Arabidopsis seed-to-seedling transition. *Plant Physiol.*, kiae225. doi: 10.1093/plphys/kiae225

Lumba, S., Toh, S., Handfield, L.-F., Swan, M., Liu, R., Youn, J.-Y., et al. (2014). A mesoscale abscisic acid hormone interactome reveals a dynamic signaling landscape in arabidopsis. *Dev. Cell* 29, 360–372. doi: 10.1016/j.devcel.2014.04.004

Lumba, S., Tsuchiya, Y., Delmas, F., Hezky, J., Provart, N. J., Shi Lu, Q., et al. (2012). The embryonic leaf identity gene FUSCA3 regulates vegetative phase transitions by negatively modulating ethylene-regulated gene expression in Arabidopsis. *BMC Biol.* 10, 8. doi: 10.1186/1741-7007-10-8

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., et al. (2006). Different *Polycomb* group complexes regulate common target genes in *Arabidopsis*. *EMBO Rep.* 7, 947–952. doi: 10.1038/sj.embor.7400760

Manosalva Pérez, N., Ferrari, C., Engelhorn, J., Depuydt, T., Nelissen, H., Hartwig, T., et al. (2024). MINI-AC : inference of plant gene regulatory networks using bulk or single-cell accessible chromatin profiles. *Plant J.* 117, 280–301. doi: 10.1111/tpj.16483

Meinke, D. W. (1992). A homoeotic mutant of arabidopsis thaliana with leafy cotyledons. *Science* 258, 1647–1650. doi: 10.1126/science.258.5088.1647

Meinke, D. W. (2020). Genome-wide identification of *EMBRYO - DEFECTIVE* (*EMB*) genes required for growth and development in Arabidopsis. *New Phytol.* 226, 306–325. doi: 10.1111/nph.16071

Meinke, D. W., Franzmann, L. H., Nickle, T., and Yeung, E. (1994). Leafy cotyledon mutants of arabidopsis. *Plant Cell* 6, 1049–1064. doi: 10.1105/tpc.6.8.1049

Mendes, A., Kelly, A. A., van Erp, H., Shaw, E., Powers, S. J., Kurup, S., et al. (2013). bZIP67 regulates the omega-3 fatty acid content of *arabidopsis* seed oil by activating *FATTY ACID DESATURASE3*. *Plant Cell* 25, 3104–3116. doi: 10.1105/tpc.113.116343

Merini, W., Romero-Campero, F. J., Gomez-Zambrano, A., Zhou, Y., Turck, F., and Calonje, M. (2017). The arabidopsis polycomb repressive complex 1 (PRC1) components atBMI1A, B, and C impact gene networks throughout all stages of plant development. *Plant Physiol.* 173, 627–641. doi: 10.1104/pp.16.01259

Mikulski, P., Wolff, P., Lu, T., Nielsen, M., Echevarria, E. F., Zhu, D., et al. (2022). VAL1 acts as an assembly platform co-ordinating co-transcriptional repression and chromatin regulation at Arabidopsis FLC. *Nat. Commun.* 13, 5542. doi: 10.1038/s41467-022-32897-7

Molitor, A. M., Bu, Z., Yu, Y., and Shen, W.-H. (2014). Arabidopsis AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes. *PloS Genet.* 10, e1004091. doi: 10.1371/journal.pgen.1004091

Monfared, M. M., Simon, M. K., Meister, R. J., Roig-Villanova, I., Kooiker, M., Colombo, L., et al. (2011). Overlapping and antagonistic activities of BASIC PENTACYSTEINE genes affect a range of developmental processes in Arabidopsis: Roles of BASIC PENTACYSTEINE genes. *Plant J.* 66, 1020–1031. doi: 10.1111/j.1365-313X.2011.04562.x

Mönke, G., Seifert, M., Keilwagen, J., Mohr, M., Grosse, I., Hähnel, U., et al. (2012). Toward the identification and regulation of the Arabidopsis thaliana ABI3 regulon. *Nucleic Acids Res.* 40, 8240–8254. doi: 10.1093/nar/gks594

Moon, H.-D., Lee, M.-S., Kim, S.-H., Jeong, W.-J., and Choi, D.-W. (2016). Identification of a drought responsive gene encoding a nuclear protein involved in drought and freezing stress tolerance in Arabidopsis. *Biol. Plant* 60, 105–112. doi: 10.1007/s10535-015-0567-1

Mozgova, I., and Hennig, L. (2015). The Polycomb group protein regulatory network. *Annu. Rev. Plant Biol.* 66, 269–296. doi: 10.1146/annurev-arplant-043014-115627

Mu, Y., Zou, M., Sun, X., He, B., Xu, X., Liu, Y., et al. (2017). Basic pentacysteine proteins repress *abscisic acid insensitive4* expression *via* direct recruitment of the polycomb-repressive complex 2 in *arabidopsis* root development. *Plant Cell Physiol.* 58, 607–621. doi: 10.1093/pcp/pcx006

Myers, Z. A., and Holt, B. F. (2018). NUCLEAR FACTOR-Y: still complex after all these years? *Curr. Opin. Plant Biol.* 45, 96–102. doi: 10.1016/j.pbi.2018.05.015

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed: Molecular profiling in Arabidopsis seed. *Plant J.* 41, 697–709. doi: 10.1111/j.1365-313X.2005.02337.x

Nakamura, S., Lynch, T. J., and Finkelstein, R. R. (2001). Physical interactions between ABA response loci of Arabidopsis. *Plant J. Cell Mol. Biol.* 26, 627–635. doi: 10.1046/j.1365-313x.2001.01069.x

Nambara, E., Keith, K., McCourt, P., and Naito, S. (1995). A regulatory role for the *ABI3* gene in the establishment of embryo maturation in *Arabidopsis thaliana*. *Development* 121, 629–636. doi: 10.1242/dev.121.3.629

Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y. V., Liew, L. C., et al. (2017). Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination. *Genome Biol.* 18, 172. doi: 10.1186/s13059-017-1302-3

O'Malley, R. C., Huang, S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., et al. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* 165, 1280–1292. doi: 10.1016/j.cell.2016.04.038

Ogas, J., Cheng, J.-C., Sung, Z. R., and Somerville, C. (1997). Cellular Differentiation Regulated by Gibberellin in the *Arabidopsis thaliana pickle* Mutant. *Science* 277, 91–94. doi: 10.1126/science.277.5322.91

Pan, J., Hu, Y., Wang, H., Guo, Q., Chen, Y., Howe, G. A., et al. (2020). Molecular mechanism underlying the synergetic effect of jasmonate on abscisic acid signaling during seed germination in arabidopsis. *Plant Cell* 32, 3846–3865. doi: 10.1105/tpc.19.00838

Pan, W., Li, J., Du, Y., Zhao, Y., Xin, Y., Wang, S., et al. (2023b). Epigenetic silencing of callose synthase by VIL1 promotes bud-growth transition in lily bulbs. *Nat. Plants* 9, 1451–1467. doi: 10.1038/s41477-023-01492-z

Pan, J., Zhang, H., Zhan, Z., Zhao, T., and Jiang, D. (2023a). A REF6-dependent H3K27me3-depleted state facilitates gene activation during germination in Arabidopsis. *J. Genet. Genomics* 50, 178–191. doi: 10.1016/j.jgg.2022.09.001

Parcy, F., Valon, C., Raynal, M., Gaubier-Comella, P., Delseny, M., and Giraudat, J. (1994). Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. *Plant Cell* 6, 1567–1582. doi: 10.1105/tpc.6.11.1567

Pelletier, J. M., Kwong, R. W., Park, S., Le, B. H., Baden, R., Cagliari, A., et al. (2017). LEC1 sequentially regulates the transcription of genes involved in diverse developmental processes during seed development. *Proc. Natl. Acad. Sci.* 114, E6710–E6719. doi: 10.1073/pnas.1707957114

Perruc, E., Kinoshita, N., and Lopez-Molina, L. (2007). The role of chromatin-remodeling factor PKL in balancing osmotic stress responses during Arabidopsis seed germination. *Plant J.* 52, 927–936. doi: 10.1111/j.1365-313X.2007.03288.x

Pien, S., Fleury, D., Mylne, J. S., Crevillen, P., Inzé, D., Avramova, Z., et al. (2008). ARABIDOPSIS TRITHORAX1 dynamically regulates *FLOWERING LOCUS C* activation via histone 3 lysine 4 trimethylation. *Plant Cell* 20, 580–588. doi: 10.1105/tpc.108.058172

Ramírez-González, 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

Ramirez-Prado, J. S., Latrasse, D., Rodriguez-Granados, N. Y., Huang, Y., Manza-Mianza, D., Brik-Chaouche, R., et al. (2019). The Polycomb protein LHP 1 regulates *Arabidopsis thaliana* stress responses through the repression of the MYC 2-dependent branch of immunity. *Plant J.* 100, 1118–1131. doi: 10.1111/tpj.14502

Raz, V., Bergervoet, J. H. W., and Koornneef, M. (2001). Sequential steps for developmental arrest in *Arabidopsis* seeds. *Development* 128, 243–252. doi: 10.1242/dev.128.2.243

Reidt, W., Wohlfarth, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., et al. (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. *Plant J.* 21, 401–408. doi: 10.1046/j.1365-313x.2000.00686.x

Roscoe, T. T., Guilleminot, J., Bessoule, J.-J., Berger, F., and Devic, M. (2015). Complementation of seed maturation phenotypes by ectopic expression of ABSCISIC ACID INSENSITIVE3, FUSCA3 and LEAFY COTYLEDON2 in arabidopsis. *Plant Cell Physiol.* 56, 1215–1228. doi: 10.1093/pcp/pcv049

Ryu, H., Cho, H., Bae, W., and Hwang, I. (2014). Control of early seedling development by BES1/TPL/HDA19-mediated epigenetic regulation of ABI3. *Nat. Commun.* 5, 4138. doi: 10.1038/ncomms5138

Santos Mendoza, M., Dubreucq, B., Miquel, M., Caboche, M., and Lepiniec, L. (2005). LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in *Arabidopsis* leaves. *FEBS Letters* 579, 4666–4670. doi: 10.1016/j.febslet.2005.07.037

Schneider, A., Aghamirzaie, D., Elmarakeby, H., Poudel, A. N., Koo, A. J., Heath, L. S., et al. (2016). Potential targets of VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing *Arabidopsis thaliana* embryos. *Plant J.* 85, 305–319. doi: 10.1111/tpj.13106

Schubert, D., Clarenz, O., and Goodrich, J. (2005). Epigenetic control of plant development by Polycomb-group proteins. *Curr. Opin. Plant Biol.* 8, 553–561. doi: 10.1016/j.pbi.2005.07.005

Shahan, R., Hsu, C.-W., Nolan, T. M., Cole, B. J., Taylor, I. W., Greenstreet, L., et al. (2022). A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev. Cell* 57, 543–560.e9. doi: 10.1016/j.devcel.2022.01.008

Shen, Y., Devic, M., Lepiniec, L., and Zhou, D.-X. (2015). Chromodomain, Helicase and DNA-binding CHD1 protein, CHR5, are involved in establishing active chromatin state of seed maturation genes. *Plant Biotechnol. J.* 13, 811–820. doi: 10.1111/pbi.12315

Shu, J., Chen, C., Thapa, R. K., Bian, S., Nguyen, V., Yu, K., et al. (2019). Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis* seedlings. *Plant Direct* 3, e00100. doi: 10.1002/pld3.100

Song, L., Huang, S.-S. C., Wise, A., Castanon, R., Nery, J. R., Chen, H., et al. (2016). A transcription factor hierarchy defines an environmental stress response network. *Science* 354, aag1550–aag1550. doi: 10.1126/science.aag1550

Sreenivasulu, N., and Wobus, U. (2013). Seed-development programs: A systems biology–based comparison between dicots and monocots. *Annu. Rev. Plant Biol.* 64, 189–217. doi: 10.1146/annurev-arplant-050312-120215

Stone, S. L., Braybrook, S. A., Paula, S. L., Kwong, L. W., Meuser, J., Pelletier, J., et al. (2008). Arabidopsis LEAFY COTYLEDON2 induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3151–3156. doi: 10.1073/pnas.0712364105

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., et al. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci.* 98, 11806–11811. doi: 10.1073/pnas.201413498

Subburaj, S., Cao, S., Xia, X., and He, Z. (2016). Phylogenetic Analysis, Lineage-Specific Expansion and Functional Divergence of seed dormancy 4-Like Genes in Plants. *PloS One* 11, e0153717. doi: 10.1371/journal.pone.0153717

Sugimoto, K., Takeuchi, Y., Ebana, K., Miyao, A., Hirochika, H., Hara, N., et al. (2010). Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. *Proc. Natl. Acad. Sci.* 107, 5792–5797. doi: 10.1073/pnas.0911965107

Sung, Z. R., Belachew, A., Shunong, B., and Bertrand-Garcia, R. (1992). *EMF*, an *arabidopsis* gene required for vegetative shoot development. *Science* 258, 1645–1647. doi: 10.1126/science.258.5088.1645

Suzuki, M., Ketterling, M. G., Li, Q.-B., and McCarty, D. R. (2003). Viviparous1 alters global gene expression patterns through regulation of abscisic acid signaling. *Plant Physiol.* 132, 1664–1677. doi: 10.1104/pp.103.022475

Suzuki, M., and McCarty, D. R. (2008). Functional symmetry of the B3 network controlling seed development. *Curr. Opin. Plant Biol.* 11, 548–553. doi: 10.1016/j.pbi.2008.06.015

Suzuki, M., Wang, H. H.-Y., and McCarty, D. R. (2007). Repression of the *LEAFY COTYLEDON 1/B3* regulatory network in plant embryo development by *VP1/ABSCISIC ACID INSENSITIVE 3 - LIKE* B3 genes. *Plant Physiol.* 143, 902–911. doi: 10.1104/pp.106.092320

Tai, H. H., Tai, G. C. C., and Beardmore, T. (2005). Dynamic histone acetylation of late embryonic genes during seed germination. *Plant Mol. Biol.* 59, 909–925. doi: 10.1007/s11103-005-2081-x

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tanaka, M., Kikuchi, A., and Kamada, H. (2008). The arabidopsis histone deacetylases HDA6 and HDA19 contribute to the repression of embryonic properties after germination. *Plant Physiol.* 146, 149–161. doi: 10.1104/pp.107.111674

Tang, X., Lim, M.-H., Pelletier, J., Tang, M., Nguyen, V., Keller, W. A., et al. (2012). Synergistic repression of the embryonic programme by SET DOMAIN GROUP 8 and EMBRYONIC FLOWER 2 in Arabidopsis seedlings. *J. Exp. Bot.* 63, 1391–1404. doi: 10.1093/jxb/err383

Tian, H., Chen, S., Yang, W., Wang, T., Zheng, K., Wang, Y., et al. (2017). A novel family of transcription factors conserved in angiosperms is required for ABA signalling. *Plant Cell Environ.* 40, 2958–2971. doi: 10.1111/pce.13058

Tian, R., Wang, F., Zheng, Q., Niza, V. M. A. G. E., Downie, A. B., and Perry, S. E. (2020). Direct and indirect targets of the arabidopsis seed transcription factor ABSCISIC ACID INSENSITIVE3. *Plant J. Cell Mol. Biol.* 103, 1679–1694. doi: 10.1111/tpj.14854

Tremblay, B. J. M., Santini, C. P., Cheng, Y., Zhang, X., Rosa, S., and Qüesta, J. I. (2024). Interplay between coding and non-coding regulation drives the Arabidopsis seed-to-seedling transition. *Nat. Commun.* 15, 1724. doi: 10.1038/s41467-024-46082-5

Tsai, A. Y.-L., and Gazzarrini, S. (2012). AKIN10 and FUSCA3 interact to control lateral organ development and phase transitions in Arabidopsis: Interaction between FUSCA3 and AKIN10. *Plant J.* 69, 809–821. doi: 10.1111/j.1365-313X.2011.04832.x

Tsukagoshi, H., Morikami, A., and Nakamura, K. (2007). Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. *Proc. Natl. Acad. Sci.* 104, 2543–2547. doi: 10.1073/pnas.0607940104

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.-L., Guillaume, E., Buisine, N., et al. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PloS Genet.* 3, e86. doi: 10.1371/journal.pgen.0030086

Veerappan, V., Chen, N., Reichert, A. I., and Allen, R. D. (2014). HSI2/VAL1 PHD-like domain promotes H3K27 trimethylation to repress the expression of seed maturation genes and complex transgenes in Arabidopsis seedlings. *BMC Plant Biol.* 14, 293. doi: 10.1186/s12870-014-0293-4

Veerappan, V., Wang, J., Kang, M., Lee, J., Tang, Y., Jha, A. K., et al. (2012). A novel HSI2 mutation in Arabidopsis affects the PHD-like domain and leads to derepression of seed-specific gene expression. *Planta* 236, 1–17. doi: 10.1007/s00425-012-1630-1

Veluchamy, A., Jégu, T., Ariel, F., Latrasse, D., Mariappan, K. G., Kim, S.-K., et al. (2016). LHP1 regulates H3K27me3 spreading and shapes the three-dimensional conformation of the arabidopsis genome. *PloS One* 11, e0158936. doi: 10.1371/journal.pone.0158936

Wang, H., Jiang, D., Axelsson, E., Lorković, Z. J., Montgomery, S., Holec, S., et al. (2018). LHP1 Interacts with ATRX through Plant-Specific Domains at Specific Loci Targeted by PRC2. *Mol. Plant* 11, 1038–1052. doi: 10.1016/j.molp.2018.05.004

Wang, H., Liu, C., Cheng, J., Liu, J., Zhang, L., He, C., et al. (2016). Arabidopsis flower and embryo developmental genes are repressed in seedlings by different combinations

of polycomb group proteins in association with distinct sets of cis-regulatory elements. *PloS Genet.* 12, e1005771. doi: 10.1371/journal.pgen.1005771

Wang, F., and Perry, S. E. (2013). Identification of direct targets of FUSCA3, a key regulator of arabidopsis seed development. *Plant Physiol.* 161, 1251–1264. doi: 10.1104/pp.112.212282

Wendrich, J. R., Yang, B., Vandamme, N., Verstaen, K., Smet, W., Van De Velde, C., et al. (2020). Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* 370, eaay4970. doi: 10.1126/science.aay4970

West, M., Yee, K. M., Danao, J., Zimmerman, J. L., Fischer, R. L., Goldberg, R. B., et al. (1994). LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in arabidopsis. *Plant Cell* 6, 1731–1745. doi: 10.1105/tpc.6.12.1731

Whittaker, C., and Dean, C. (2017). The *FLC* locus: A platform for discoveries in epigenetics and adaptation. *Annu. Rev. Cell Dev. Biol.* 33, 555–575. doi: 10.1146/annurev-cellbio-100616-060546

Wu, T., Alizadeh, M., Lu, B., Cheng, J., Hoy, R., Bu, M., et al. (2022). The transcriptional co-repressor SEED DORMANCY 4-LIKE (AtSDR4L) promotes the embryonic-to-vegetative transition in *Arabidopsis thaliana*. *J. Integr. Plant Biol.* 64, 2075–2096. doi: 10.1111/jipb.13360

Wu, J., Mohamed, D., Dowhanik, S., Petrella, R., Gregis, V., Li, J., et al. (2020). Spatiotemporal restriction of *FUSCA3* expression by class I BPCs promotes ovule development and coordinates embryo and endosperm growth. *Plant Cell* 32, 1886–1904. doi: 10.1105/tpc.19.00764

Xiang, D., Quilichini, T. D., Liu, Z., Gao, P., Pan, Y., Li, Q., et al. (2019). The transcriptional landscape of polyploid wheats and their diploid ancestors during embryogenesis and grain development. *Plant Cell* 31, 2888–2911. doi: 10.1105/tpc.19.00397

Xiao, J., Jin, R., and Wagner, D. (2017a). Developmental transitions: integrating environmental cues with hormonal signaling in the chromatin landscape in plants. *Genome Biol.* 18, 88. doi: 10.1186/s13059-017-1228-9

Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J. D., Pai, A., et al. (2017b). Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. *Nat. Genet.* 49, 1546–1552. doi: 10.1038/ng.3937

Xu, F., Kuo, T., Rosli, Y., Liu, M.-S., Wu, L., Chen, L.-F. O., et al. (2018). Trithorax group proteins act together with a polycomb group protein to maintain chromatin integrity for epigenetic silencing during seed germination in arabidopsis. *Mol. Plant* 11, 659–677. doi: 10.1016/j.molp.2018.01.010

Xu, L., and Shen, W.-H. (2008). Polycomb silencing of KNOX genes confines shoot stem cell niches in arabidopsis. *Curr. Biol.* 18, 1966–1971. doi: 10.1016/j.cub.2008.11.019

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., and Hattori, T. (2009). Arabidopsis NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J.* 58, 843–856. doi: 10.1111/j.1365-313X.2009.03817.x

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL- and atBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in arabidopsis. *Curr. Biol.* 23, 1324–1329. doi: 10.1016/j.cub.2013.05.050

Yang, Z., Liu, X., Wang, K., Li, Z., Jia, Q., Zhao, C., et al. (2022b). ABA-INSENSITIVE 3 with or without FUSCA3 highly up-regulates lipid droplet proteins and activates oil accumulation. *J. Exp. Bot.* 73, 2077–2092. doi: 10.1093/jxb/erab524

Yang, D., Zhao, F., Zhu, D., Chen, X., Kong, X., Wu, Y., et al. (2022a). Progressive chromatin silencing of ABA biosynthesis genes permits seed germination in Arabidopsis. *Plant Cell* 34, 2871–2891. doi: 10.1093/plcell/koac134

Yao, J., Chu, Q., Guo, X., Shao, W., Shang, N., Luo, K., et al. (2024). Spatiotemporal transcriptomic landscape of rice embryonic cells during seed germination. *Dev. Cell.* doi: 10.1016/j.devcel.2024.05.016

Yin, X., Romero-Campero, F. J., de Los Reyes, P., Yan, P., Yang, J., Tian, G., et al. (2021). H2AK121ub in Arabidopsis associates with a less accessible chromatin state at transcriptional regulation hotspots. *Nat. Commun.* 12, 315. doi: 10.1038/s41467-020-20614-1

Yin, X., Romero-Campero, F. J., Yang, M., Baile, F., Cao, Y., Shu, J., et al. (2023). Binding by the Polycomb complex component BMI1 and H2A monoubiquitination shape local and long-range interactions in the Arabidopsis genome. *Plant Cell* 35, 2484–2503. doi: 10.1093/plcell/koad112

Yuan, L., Song, X., Zhang, L., Yu, Y., Liang, Z., Lei, Y., et al. (2021). The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in *Arabidopsis*. *Nucleic Acids Res.* 49, 98–113. doi: 10.1093/nar/gkaa1129

Zhang, M., Cao, X., Jia, Q., and Ohlrogge, J. (2016). *FUSCA3* activates triacylglycerol accumulation in Arabidopsis seedlings and tobacco BY2 cells. *Plant J.* 88, 95–107. doi: 10.1111/tpj.13233

Zhang, X., Garreton, V., and Chua, N.-H. (2005). The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation. *Genes Dev.* 19, 1532–1543. doi: 10.1101/gad.1318705

Zhang, X., Germann, S., Blus, B. J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S. E. (2007). The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.* 14, 869–871. doi: 10.1038/nsmb1283

Zhang, Y., Miao, X., Xia, X., and He, Z. (2014). Cloning of seed dormancy genes (TaSdr) associated with tolerance to pre-harvest sprouting in common wheat and development of a functional marker. *Theor. Appl. Genet.* 127, 855–866. doi: 10.1007/s00122-014-2262-6

Zhao, M., He, W., Tang, J., Zou, Q., and Guo, F. (2021). A comprehensive overview and critical evaluation of gene regulatory network inference technologies. *Brief. Bioinform.* 22, bbab009. doi: 10.1093/bib/bbab009

Zhao, T., Lu, J., Zhang, H., Xue, M., Pan, J., Ma, L., et al. (2022b). Histone H3.3 deposition in seed is essential for the post-embryonic developmental competence in Arabidopsis. *Nat. Commun.* 13, 7728. doi: 10.1038/s41467-022-35509-6

Zhao, B., Zhang, H., Chen, T., Ding, L., Zhang, L., Ding, X., et al. (2022a). *Sdr4* dominates pre-harvest sprouting and facilitates adaptation to local climatic condition in Asian cultivated rice. *J. Integr. Plant Biol.* 64, 1246–1263. doi: 10.1111/jipb.13266

Zheng, L., Otani, M., Kanno, Y., Seo, M., Yoshitake, Y., Yoshimoto, K., et al. (2022). Seed dormancy 4 like1 of Arabidopsis is a key regulator of phase transition from embryo to vegetative development. *Plant J.* 112, 460–475. doi: 10.1111/tpj.15959

Zhou, Y., Romero-Campero, F. J., Gómez-Zambrano, Á., Turck, F., and Calonje, M. (2017). H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. *Genome Biol.* 18, 69. doi: 10.1186/s13059-017-1197-z

Zhou, Y., Tan, B., Luo, M., Li, Y., Liu, C., Chen, C., et al. (2013). HISTONE DEACETYLASE19 interacts with HSL1 and participates in the repression of seed maturation genes in *arabidopsis* seedlings. *Plant Cell* 25, 134–148. doi: 10.1105/tpc.112.096313

Zhou, Y., Yang, P., Zhang, F., Luo, X., and Xie, J. (2020). Histone deacetylase HDA19 interacts with histone methyltransferase SUVH5 to regulate seed dormancy in *Arabidopsis*. *Plant Biol.* 22, 1062–1071. doi: 10.1111/plb.13158

Zhu, Y., Hu, X., Duan, Y., Li, S., Wang, Y., Rehman, A. U., et al. (2020). The arabidopsis nodulin homeobox factor atNDX interacts with atRING1A/B and negatively regulates abscisic acid signaling. *Plant Cell* 32, 703–721. doi: 10.1105/tpc.19.00604

Zhu, T., Wei, C., Yu, Y., Zhang, Z., Zhu, J., Liang, Z., et al. (2024). The BAS chromatin remodeler determines brassinosteroid-induced transcriptional activation and plant growth in Arabidopsis. *Dev. Cell* 59, 924–939.e6. doi: 10.1016/j.devcel.2024.01.021

Zong, W., Kim, J., Bordiya, Y., Qiao, H., and Sung, S. (2022). Abscisic acid negatively regulates the Polycomb-mediated H3K27me3 through the PHD-finger protein, VIL1. *New Phytol.* 235, 1057–1069. doi: 10.1111/nph.18156

Glossary

LEC1, FUS3, LEC2, ABI3 (LAFL)	
LEC1	LEAFY COTYLEDON 1
FUS3	FUSCA 3
LEC2	LEAFY COTYLEDON 2
ABI3	ABSCISIC ACID INSENSITIVE 3
Polycomb Repressive Complex 1 (PRC1) Core	
AtRING1A, 1B	*Arabidopsis thaliana* RING 1A, 1B
AtBMI1A, 1B, 1C	*Arabidopsis thaliana* BMI1A, 1B, 1C
Polycomb Repressive Complex 2 (PRC2) Core	
CLF	CURLY LEAF
SWN	SWINGER
MEA	MEDEA
EMF2	EMBRYONIC FLOWER 2
VRN2	VERNALIZATION 2
FIS2	FERTILIZATION INDEPENDENT SEED 2
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
MSI1	MULTICOPY SUPPRESSOR OF IRA 1
Proteins that Interact with PRC Core	
AL6, 7	ALFIN-LIKE 6, 7
VAL1/HSI2	VIVIPAROUS1/ABI3-LIKE 1/HIGH-LEVEL EXPRESSION OF SUGAR INDUCIBLE GENE 2
VAL2/HSL1	VIVIPAROUS1/ABI3-LIKE 2/HSI2-LIKE 1
VIL1/VRN5	VERNALIZATION INSENSITIVE 3-LIKE 1/ VERNALIZATION 5
LHP1/TFL2	LIKE HETEROCHROMATIN PROTEIN 1/TERMINAL FLOWER 2
EMF1	EMBRYONIC FLOWER 1
BPC1, 2	BASIC PENTACYSTEINE 1, 2
AtSDR4L/ SFL1/ODR1	*Arabidopsis thaliana* SEED DORMANCY FOUR-LIKE 1/SEED DORMANCY FOUR-LIKE 1/REVERSAL OF RDO5 1
DIG1/ SFL2/AITR2	DYNAMIC INFLUENCER OF GENE 1/SEED DORMANCY FOUR-LIKE 2/ABA-INDUCED TRANSCRIPTION REPRESSOR 2
DIL1/ SFL3/AITR6	DIG-LIKE 1/SEED DORMANCY FOUR-LIKE 3/ABA-INDUCED TRANSCRIPTION REPRESSOR 6
DIG2/ SFL4/AITR5	DYNAMIC INFLUENCER OF GENE 2/SEED DORMANCY FOUR-LIKE 4/ABA-INDUCED TRANSCRIPTION REPRESSOR 5
NDX	NODULIN HOMEOBOX
HDA6, 19	HISTONE DEACETYLASE 6, 19
SAP18	SIN3 ASSOCIATED POLYPEPTIDE 18

(Continued)

Continued

Other Epigenetic Regulators of Seed-to-Seedling Transition	
ATX1	ARABIDOPSIS TRITHORAX 1
ULT1	ULTRAPETALA 1
PKL	PICKLE
PKR2	PICKLE RELATED 2
SDG8	SET DOMAIN GROUP 8
TPL	TOPLESS
TPR1, 2, 3, 4	TOPLESS-RELATED 1, 2, 3, 4

frontiers | Frontiers in Plant Science

TYPE Mini Review
PUBLISHED 11 February 2025
DOI 10.3389/fpls.2025.1429225

Check for updates

OPEN ACCESS

EDITED BY
Raju Datla,
Global Institute for Food Security (GIFS), Canada

REVIEWED BY
José Ángel Huerta Ocampo,
National Council of Science and Technology (CONACYT), Mexico
Talia Hernandez,
Unidad Irapuato (CINVESTAV), Mexico

*CORRESPONDENCE
Nii Patterson
nii.patterson@nrc-cnrc.gc.ca

RECEIVED 07 May 2024
ACCEPTED 07 January 2025
PUBLISHED 11 February 2025

CITATION
Nosworthy MG, Yu B, Zaharia LI, Medina G and Patterson N (2025) Pulse protein quality and derived bioactive peptides.
Front. Plant Sci. 16:1429225.
doi: 10.3389/fpls.2025.1429225

COPYRIGHT
© 2025 His Majesty the King in Right of Canada, as represented by the National Research Council of Canada. 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.

Pulse protein quality and derived bioactive peptides

Matthew G. Nosworthy[1,2], Bianyun Yu[3], L. Irina Zaharia[3], Gerardo Medina[4] and Nii Patterson[3]*

[1]Guelph Research and Development Center, Agriculture and Agri-Food Canada, Guelph, ON, Canada, [2]College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada, [3]Aquatic and Crop Resource Development, National Research Council of Canada, Saskatoon, SK, Canada, [4]Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada

There is a growing consumer interest in sources of dietary protein that are plant-based. Pulse crops, such as lentils, beans, chickpeas, and peas, are gaining popularity due to their environmental sustainability, nutrient density, and functional attributes. The protein content and quality of pulses vary across different pulse classes and processing methods. The biological properties of the protein and the physiologically active peptides make pulse crops attractive as potentially functional or health-promoting foods. This review highlights the nutritional quality of pulse proteins as determined by the Protein Efficiency Ratio and Protein Digestibility Corrected Amino Acid Score as well as bioactive properties of specific bioactive peptides related to amelioration of hypertension and diabetes. Additionally, the use of proteomics platforms, such as mass spectrometry, in combination with bioinformatics tools, enables the identification and characterization of bioactive peptides in pulse crops. These technologies facilitate the development of pulse-derived products with enhanced nutritional values. Overall, the high nutritional quality of pulse-based proteins supports the benefits of pulse inclusion in the diet, which can also exert beneficial bioactivities resulting in improving outcomes in non-communicable diseases.

KEYWORDS

pulse, protein quality, bioactive peptide, protein efficiency ratio, protein digestibility corrected amino acid score, proteomics

Introduction

The dietary consumption patterns of humans are recognized as an important issue in overall health status, and also as a strategy to prevent and/or attenuate the development of several diseases (Muzquiz et al., 2012). The global management of agricultural production, as well as the adoption of alternative and sustainable food sources for the increasing world population, will require high-yielding crops with improved nutritional quality that can be productive under increasingly variable climate conditions. Recently, the nutritional content of foods has re-emerged as a key element of improving the quality of food production and diets due to increased consumer knowledge and nutrient desire. Worldwide, there is a growing shift in diet to nutritionally dense and functional foods to address health concerns

(Singh and Basu, 2012). In 2018, the global market size of functional foods was about 150 billion US dollars, with this number expected to rapidly climb up to about 250 billion US dollars by 2024 (Rattanachaikunsopon and Phumkhachorn, 2018). The anticipated expansion of the market will be driven by growing effectiveness of functional foods and their bioactive components that guarantee the dietary nutritional needs.

In addition to the nutritional value of a food, or a crop, it is important to consider the bioactive components that are inherently present. Dietary proteins are known to carry a wide range of nutritional, functional, and biological properties. Nutritionally, proteins are a source of energy and amino acids essential for maintenance of growth, homeostasis, and combatting disease, among other metabolic interactions. Functionally, proteins contribute to the physicochemical and sensory properties of foods. Many dietary proteins also possess specific biological properties which make these components potential ingredients for the development of functional foods. These properties are attributed to physiologically active peptides encrypted in dietary protein sequences. These peptides are inactive, or have reduced activity, while within the native protein sequence and can be released during gastrointestinal digestion or via enzymatic exposure during food processing and germination (Korhonen and Pihlanto, 2006; Daliri et al., 2017). The size of these biologically functional peptides generally ranges between 2 and 20 amino acid residues and numerous important regulatory functions have been associated with these peptide fragments, including antihypertensive, antimicrobial, immunomodulatory, opioid, antioxidant, and mineral binding activities (Rutherfurd-Markwick, 2012; Soory, 2012; Sánchez and Vázquez, 2017). Currently, bioactive peptides (BPs) or protein hydrolysates can be commercialized as nutraceutical products or functional ingredients (López-Barrios et al., 2014).

Although many foods exhibit therapeutic potential, pulse-based foods are currently being valued worldwide due to their environmental sustainability, nutrient density, and functional attributes. Pulses such as lentils, common beans, chickpeas, and dry peas, are widely recognized for their potential to simultaneously promote human and environmental health (Foyer et al., 2016; Bessada et al., 2019). Recently there has been a growing interest in pulses and new pulse ingredients due to health promoting benefits (Aguilera et al., 2011), particularly the prevention and management of highly prevalent chronic inflammation and oxidative stress-related noncommunicable diseases (Guardado Félix and Gutierrez Uribe, 2019). In addition to the high protein content, a significant part of the potential health benefits and disease-preventing properties attributed to pulses are derived from the presence of 'non-nutrient' compounds that include dietary fiber, phytochemicals and antioxidants (Xu and Chang, 2007; Campos-Vega et al., 2010; Awika and Duodu, 2017; Giusti et al., 2017). Challenges therefore exist throughout the food development process, from the selection of best pulse type/cultivar (Carbas et al., 2020) to the best derived ingredients (Foschia et al., 2017). Several unique peptides generated from pulse protein isolates have been identified and associated with various functional attributes (Cosson et al., 2022; Asledottir et al., 2023). Given these points it is important to consider not only the nutritional value of a protein source such as pulses, but also the bioactivities of specific amino acids and peptides commonly found in pulses as well as their hydrolysates.

Pulse protein content and quality

Pulse crops have long been known to be an excellent source of dietary protein compared to most vegetable-based foods (Martín-Cabrejas, 2019). As protein is a vital nutrient in foods and ingredients, it is important to consider two aspects of its nature, total protein content and its nutritional quality. Plant-based protein sources typically have lower protein content when compared to animal sources; however there exists a considerable variation across different crops. The protein content of cereals falls in the range of 8-15% depending on the species, with wheat being reported as containing 13% protein (Mathai et al., 2017), while oilseeds such as soy can be as high as 46% (Nosworthy et al., 2023). Pulse protein content falls between cereals and soy, with a range of 20-30% depending on pulse class and processing method (Hall et al., 2017; Nosworthy et al., 2018a, 2020). Much like most plant-based foods, processing pulses is necessary due to the presence of antinutritional factors such as protease inhibitors and tannins which impact the digestibility/bioavailability of nutrients present in these foods. Processing also enhances the palatability of these foods leading to a more pleasant experience during consumption.

While pulses have a relatively high protein content compared to other plant sources, the nutritional quality must also be considered. There are multiple methods assessing protein quality, with many of them involved in regulating content claims in different global jurisdictions. These methods can be divided into two groups, those that require only chemical assessment of the food/ingredient and those that necessitate animal experimentation. For foods in Europe the guidelines stipulate that a "source" of protein must provide 12% of its energy from protein and 20% for a "high source" (available online: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32006R1924, European Commission, 2006). Similarly, Oceania bases their guidelines on the protein content per serving with 5 g allowing a content claim and 10 g providing a "good source" of protein available online: (https://www.legislation.gov.au/F2015L00394/2017-09-07/text, Food Standards Australia New Zealand 2015). Conversely, the Protein Efficiency Ratio (PER) and Protein Digestibility Corrected Amino Acid Score (PDCAAS), required in Canada and the United States respectively, necessitate animal experimentation.

The methods for determining PER and PDCAAS are well described elsewhere (Marinangeli and House, 2017). Briefly, PER is a growth measurement where the efficiency by which a rat can convert dietary protein into growth indicating the quality of dietary protein (Health Canda, 1981; available online: http://www.hc-sc.gc.ca/fn-an/alt_formats/hpfb-dgpsa/pdf/res-rech/fo-1-eng.pdf; accessed on 20 January 2024). This PER is subsequently compared to a casein control group to generate an adjusted PER score, which is then used to generate a Protein Rating where a value of ≥20 is a "good source" of protein and ≥40 is an "excellent source". PDCAAS

relies on the comparison between the amino acid composition of a food and human requirements which is then corrected by a measurement of protein digestibility determined in a rat (FAO/WHO, 1991). The PDCAAS score is then used to correct the protein content in a food serving where if the resulting corrected protein score is ≥ 5 g per serving the food is a "good source", ≥ 10 g is an "excellent source". Notably a recent change in Health Canada policy has allowed for the use of a PDCAAS score and a conversion factor of 2.5 to generate a PER value (Health Canada, 2023; available online: https://www.canada.ca/en/health-canada/services/food-nutrition/legislation-guidelines/policies/measuring-protein-quality-foods.html; accessed on 24 January 2024). There is a more recent method for determination of protein quality, the Digestible Indispensable Amino Acid Score (DIAAS), which is a modification of PDCAAS with updated amino acid reference patterns and consideration of individual amino acids as the nutrient rather than protein, it has yet to be adopted for regulatory purposes by any jurisdiction (FAO/WHO, 2013). Relative essential and non-essential amino acids presented in Figures 1A, B and the examples of pulse protein content and quality under different processing conditions are presented in Figures 1C-E. The essential amino acid content of the various pulse classes and processing methods remains consistent, with variation being more present among the non-essential amino acids. While still under review (Health Canada, 2023), use of alternative conversion factors between PER and PDCAAS may be advisable depending on the protein source and processing method.

Bioactivity of amino acids

In addition to the direct nutritional value of pulse protein, it has been well established that dietary amino acids can influence various aspects of metabolism including muscle growth, immune response, and prevention of non-communicable diseases such as hypertension and cardiovascular disease (Pedroche et al., 2002; Hong et al., 2008; Garcia-Mora et al., 2015). Many studies have investigated the biological activities of individual amino acids/peptides and the topic has been discussed in multiple reviews (Wu, 2009, 2010; Zaky et al., 2022). Individual amino acids have been demonstrated to have a wide range of functions including regulation of cell division and gene expression, nucleotide synthesis, regulation of protein turnover, and intracellular signaling among other activities (Wu, 2009). These activities are related to the

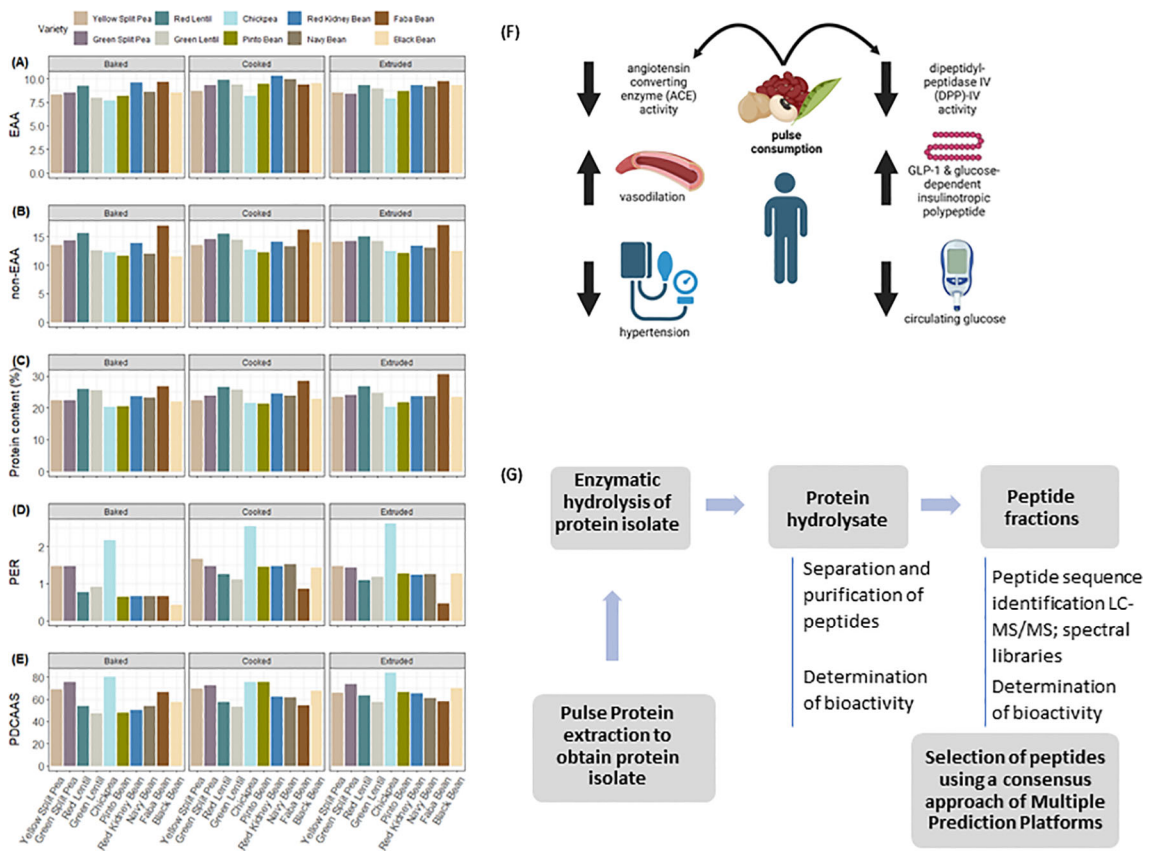


FIGURE 1
(A) Total essential and **(B)** non-essential amino acids in thermally processed pulses. EAA: Essential amino acids; NEAA: Non-essential amino acids. Data on processed peas were adapted from Nosworthy et al. (2017), lentils from Nosworthy et al. (2018b), chickpea from Nosworthy et al. (2020), and beans from Nosworthy et al. (2018a). **(C)** Protein content, **(D)** Protein Efficiency Ratio (PER), and **(E)** Protein Digestibility Corrected Amino Acid Score (PDCAAS) of thermally treated pulses. **(F)** Impact of pulse consumption angiotensin converting enzyme and Dipeptidyl-peptidase IV activity and downstream implications in human health. **(G)** Workflow for the identification and evaluation bioactive peptides derived from pulse protein.

individual amino acid, but during metabolic and enzymatic activities these amino acids can be converted into other compounds, such as creatine (tripeptide of arginine, glycine and methionine), glutathione (cystine, glutamine, and glycine), and taurine (formed from cystine), which are all antioxidants, as well as exhibiting other biological effects. This highlights that it is not simply the essential amino acids that need to be considered when investigating the bioactivity of different proteins and their constituent parts. Of particular interest in pulses are the amino acids arginine and leucine due to their high amounts in pulses and multifaceted bioactive effects.

Pulses are known to have a high arginine content, in particular chickpeas and fava beans (Boye et al., 2010; Bessada et al., 2019). Beyond protein synthesis, arginine is involved in many physiological processes including the synthesis of creatine, polyamines, and nitric oxide (NO) which is a known vasodilator (Tuteja et al., 2004; Maeda et al., 2006; Wu et al., 2021). Multiple studies have demonstrated that consumption of pulses lowers the incidence of cardiovascular disease and hypertension which may be tied to the higher arginine content and its function in vasodilation and cellular signaling, or the generation of bioactive peptides targeting angiotensin converting enzyme (Roy et al., 2010; Padhi and Ramdath, 2017). Leucine is a member of the branch-chain amino acids which also includes isoleucine and valine, and has been implicated in stimulating protein synthesis and muscle anabolism, as well as multiple immunomodulatory effects and attenuation of glucose response (Bonvini et al., 2018; Kamei et al., 2020). This is of particular importance in elderly individuals where a higher leucine content of a protein source has been demonstrated to enhance myofibrillar protein synthesis (Devries et al., 2018). Importantly, while content claims can be made for protein content as described previously, no regulatory body accepts health claims based on any specific amino acids (EFSA Panel on Dietetic Products, 2010; Krasniqi et al., 2016; Roberts, 2016). While this highlights the bioactivities of two amino acids with high concentration in pulses, it is the biological activity of peptides derived from pulse protein that has been the focus of significant amounts of research in recent years.

Pulse protein and derived bioactive peptides

Recent studies suggest that protein hydrolysates and bioactive peptides derived from pulses are among the most promising bioactive proteins (Guardado Félix and Gutierrez Uribe, 2019) thus positioning these legumes as excellent sources for the development of new protein-derived products. Bioactive peptides are small amino acid sequences derived from food proteins through *in vivo* or *in vitro* enzymatic proteolysis of the inactive precursor proteins (Udenigwe and Aluko, 2012) that can hold physiological properties (López-Barrios et al., 2014). Within the native protein, the amino acid sequence conforming the bioactive peptide is either inactive or exhibits diminished activity. However, once the peptide is released it can display diverse biological activities (Udenigwe and Aluko, 2012). Accumulating evidence from *in vitro* and *in vivo* studies indicate that pulse-derived bioactive peptides can be active toward several chronic diseases like autoimmune disorders, cardiovascular disease and diabetes, and have been used as opiates, immunomodulators, antimicrobials, antioxidants, antihypertensives and antithrombotic agents (Duranti, 2006). Many bioactive peptides have common structural properties, including a relatively short residue length, hydrophobic amino acid residues, and the presence of aromatic side chains, mainly with arginine, lysine or proline in the C-terminal position (Marambe and Wanasundara, 2012; López-Barrios et al., 2014). The biological and pharmacological properties of pulse protein-derived bioactive peptides have been recently reviewed (Tak et al., 2021; Fan et al., 2022). Of particular interest are the peptides demonstrating either ameliorative or protective effects on non-communicable diseases such as diabetes and cardiovascular disease (Figure 1F). Selected peptide sequences demonstrated to have inhibitory activities on angiotensin converting enzyme and dipeptidyl peptidase IV are presented in Table 1.

TABLE 1 Bioactive peptides from pule protein sources and their function.

Bioactive function	Protein source	Peptide sequence	Reference
ACE Inhibition	Common Beans	PVDDPQIH, KKSSG, LSFNT, KMARPV, KYMKS, SGSYS, CPGNK, INQGSLLPH, FVVAEQAGNEEQFE	Rui et al. (2013); Jakubczyk et al. (2017); Mojica et al. (2017);
	Peas	GGSGNY, DLKLP, GSSDNR, MRDLK, HNTPSR	Jakubczyk and Baraniak (2014);
	Lentils	KLRT, TLHGMV, VNRLM	Jakubczyk and Baraniak (2013);
	Chickpeas	MD, MFDL, MDLA, MDFLI, MDL	Pedroche et al. (2002); Yust et al. (2012);
	Faba bean	EEEDEDEPR, KEEEDEDEPR, VIPTEPPH, VIPTEPPHA, VVIPTEPPH, VVIPTEPPHA, NYDEGSEPR, PVNRPGEPQ, LDNINALEPDH	Martineau-Côté et al. (2022)
DPP-IV Inhibition	Common Beans	KKSSG, LSFNT, KMARPV, KYMKS, SGSYS, CPGNK	Mojica et al. (2017);
	Peas	EPF, SPGDVF, IPYWTY	Zhang et al. (2022, 2023);
	Chickpeas	AIPPGIPYW, PGIPYW	Zan et al. (2023);
	Faba bean	VIPTEPPH, VIPTEPPHA, VVIPTEPPH, VVIPTEPPHA, NYDEGSEPR, PVNRPGEPQ, LDNINALEPDH	Martineau-Côté et al. (2022)

Anti-hypertensive peptides (ACE inhibitors)

Hypertension, a known risk factor for cardiovascular disease, can be treated with the administration of drugs inhibiting an enzyme known as Angiotensin I-Converting Enzyme (ACE). This enzyme is responsible for the hydrolysis of angiotensin I to angiotensin II, which is a potent vasoconstrictor inducing elevated blood pressure. As hypertension can also be managed by alterations in diet, as well as exercise regimes, there is an interest in identifying compounds present in foods that can modify metabolic enzymes such as ACE. In the case of pulses there are multiple peptides that have been identified to inhibit ACE activity, thereby potentially leading to a reduction in blood pressure. A study investigating chickpea and pea hydrolysates demonstrated that desi and kabuli chickpeas, as well as yellow peas, are capable of inhibiting ACE activity (Barbana and Boye, 2010). Interestingly the enzymes used for hydrolysis impacted the IC_{50} for the protein hydrolysates with a range of 229-316 μg/mL for kabuli chickpea, 140-228 μg/mL for desi chickpea, and 128-412 μg/mL for yellow pea. For both types of chickpeas the use of a gastrointestinal digestion resulted in the lowest IC_{50} while papain digestion generated the lowest value for yellow peas, highlighting the importance of identifying the appropriate conditions depending on initial ingredient. A recent investigation into faba beans synthesized peptides based on those identified after *in vitro* gastrointestinal digestion and determined the differences in ACE inhibition based on peptide sequence (Martineau-Côté et al., 2024). While multiple peptides did not have any impact on ACE activity, four sequences were identified to have IC_{50} values between 43-118 μg/mL. Common among these peptides was the sequence VIPTEPPH, with the peptide having the lowest IC_{50} value of 43 μg/mL being VVIPTEPPH. In combination, this supports the idea that proper selection of initial ingredient combined with processing methods to generate a peptide with known activity can impact the development of hypertension.

Diabetic prevention peptides (DPP-IV)

Diabetes is a well-known non-communicable metabolic disorder characterized by inadequate production of insulin or development of insulin resistance thereby leading to reduced control over glucose homeostasis. While there are multiple methods of treatment available for individuals with diabetes, ranging from a modification of diet to oral and injectable drugs, one such treatment is inhibition of the enzyme dipeptidyl peptidase-IV, or DPP-IV (Silveira et al., 2013). This enzyme is a serine protease responsible for the degradation of glucagon-like peptide-1, as well as glucose-dependent insulinotropic polypeptide. As these two incretins are responsible for stimulating insulin release and decreased glucagon release, DDP-IV activity results in an increased blood glucose concentration, a concern for diabetics. The '-liptin' group of drugs such as sitagliptin are an approved drug for diabetic treatment that target this enzyme. Similar to that of ACE, there has been great interest in identifying peptides present in foods such as pulses that can also inhibit the activity of DPP-IV. One study investigated the generation of bioactive peptides targeting DPP-IV in post-simulated gastrointestinal digestion from pulses via milling, thermal treatments, fermentation, and germination (Di Stefano et al., 2019). The combination of heat treatment and fermentation increased DPP-IV inhibition in chickpeas, while fermenting ground green lentils and yellow peas induced increase of DPP-IV inhibition compared to other treatments investigated. Another study investigated the use of different digestive enzymes on the generation of DPP-IV inhibitory peptides from chickpea protein (Zan et al., 2023). In this case, application of Neutrase or Papain resulted in the greatest inhibition of DPP-IV at approximately 47% and 35% inhibition respectively. A time-course experiment determined that a digestion time of 60 minutes with Neutrase generated the highest inhibitory activity (52.50 ± 0.88%). During this experiment the IC_{50} values for the identified inhibitory peptides ranged from 12.43 μM, IAIPPGIPYW, to 110.50 μM, LAFP, with the positive control peptide exhibiting the greatest inhibition, diprotin, reaching 6.81 μM as a point of comparison. As there are a wide range of protein fragments and peptides generated by processing and digestion, it is crucial to employ methods for identification of these peptides and characterization of cultivars that can generate greater concentrations of bioactive peptides.

Proteomics platforms for analysis of pulse protein-derived bioactive peptides

The scientific advances in 'OMICS' technologies combined with molecular biology approaches in recent years have impacted crop breeding and food science significantly (Varshney et al., 2009; Kato et al., 2011). In particular, the application of these technologies to identify and select crop varieties that offer wider functional scope, enhance the nutritional profile of seed/grain and possess bioactive potential, is of significant interest to farmers, consumers, and product developers. OMICS technologies are driving a knowledge revolution toward a comprehensive understanding of functional attributes of foods, particularly the variability of bioactive molecules composition in food production and processing to improve nutritional value and health. Ahmed et al. (2022) emphasized three key advances in omics technology enabling unprecedented food composition analysis and application: (i) high-throughput platforms for analysis of a broad range of food molecules; (ii) high-resolution biochemical libraries; and (iii) data integration and machine learning.

Generally, analysis of plant protein-derived peptides is based on high-throughput proteomics platform to comprehensively profile untargeted and targeted peptides using mass spectrometry (MS) techniques, nuclear magnetic resonance, and bioinformatics tools. MS has arguably become the core technology for the qualitative and quantitative analysis of food proteins and peptides, and providing understanding of their nature, structure, functional properties and

impact on human health. Interfacing of high-resolution MS for food protein analysis with robust standard reference chemical libraries has enabled the identification of peptides based on their mass spectra with a high degree of accuracy and confidence. Some of the techniques used for bioactive peptides purification and identification have been recently reviewed by Du and Li (2022). The emergence of open-source protein databases (reviewed by Naeem et al., 2022) offers comprehensive spectral libraries that increase the efficiency of identifying bioactive molecules including molecular weight and structure.

To exploit the functional value of peptides, sequential purification and identification of various peptides are routinely used (Singh et al., 2023). This review is intended to highlight some of the MS-based strategy applications in pulse crops research defining structure/functional relation of protein and derived peptides. Commonly, ultrafiltration or in combination with chromatographic techniques allows protein hydrolysates to be separated into different fractions with specific molecular weights (MWs) for the determination of their biological activities (Figure 1G). In a study by Li and Aluko (2010), purified peptide fractions from pea protein isolate were analyzed by ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) to identify and obtain amino acid sequence of the most abundant peptide in each peak. The major peptides identified showed strong inhibitory properties toward ACE and renin that may be potentially useful as ingredients to formulate multifunctional food products and nutraceuticals. In a recent study to establish a peptide profile and identify specific pea protein-derived peptides that suppressed glucose production in mouse liver cell-line AML-12 cells, Liao et al. (2022) analyzed pea protein hydrolysate using a nano-UPLC coupled with a Q Exactive HFX Orbitrap instrument. The MS spectra of most of the peptides within the range of molecular weight of 1000 to 1499 kDa were processed using the Proteome Discoverer (PD) software (Version 2.4.0.305, Thermo Fisher Scientific) and the built-in Sequest HT search engine. The sequences were searched for in the UniProt FASTA databases (uniprot-Pisum sativum_3888.fasta). Consequently, this study revealed a new function of pea protein hydrolysate in its ability to exhibit anti-diabetic activity. A study by Karkouch et al. (2017) aimed to identify peptides with antioxidant, antityrosinase and antibiofilm activities released from *Vicia faba* (dry broad bean) seed proteins hydrolysate. The purified fraction was analyzed by liquid chromatography mass spectrometry (LC–MS/MS) coupled with an LTQ-Orbitrap hybrid mass spectrometer equipped with a nano-ESI source and annotation of peptide sequence based on MS/MS spectra was performed using the Peaks software (BSI, Canada). As such, seven peptides were identified and further demonstrated their potential as a natural source of bioactive peptides for applications in the cosmetic and pharmaceutical industries. A recent review by Hou et al. (2023) provided a comprehensive review of MS-based technologies for identification of peptides purified from mung bean, their biological activities, and their potential applications. Although there are some variations in MS equipment, applications of these instruments, the integration of large spectral datasets with bioinformatics has enabled and advanced quantitative determination and identification of pulse protein-derived peptides (Figure 1G). The review by Du and Li (2022) also highlights the available web-based tools to evaluate the relationship between the function and structure of bioactive peptides. Models based on quantitative structure/activity relationships (QSAR) and quantitative structure/property relationships (QSPR) are employed to evaluate specific bioactive peptide activities quantitatively (by QSAR) or qualitatively (by QSPR), using unknown activity data and structural information. In particular, PeptideRanker is widely used QSPR to predict the activity of different peptides by integrating peptides sequence datasets and structural parameters. The utilization of bioinformatics tools is essential when employing various MS-based technologies as it offers opportunities for the pulse industry to deploy products with enhanced nutritional values.

Conclusion

Pulses are a staple food in many regions of the world with a rising consumer interest in the Western world due to their nutritional density and environmental sustainability. Although the protein content and overall nutritional quality of pulses is high compared to other crops such as cereals, pulse class selection and method of processing can significantly impact the overall quality. This is particularly notable in the case of baked beans as they did not stimulate growth at a comparable rate to other thermal processing methods. In addition to their high nutritional quality, pulse protein also possesses numerous bioactive properties. Of particular interest is their inhibitory properties on ACE and DPPIV leading to a reduction in the severity of hypertension and diabetes. Identification and characterization of these bioactive peptides using modern technologies and tools such as proteomics, bioinformatics, peptide libraries and *in vitro* assays, will lead to the development of functional food ingredients and nutraceutical products with enhanced health benefits.

Author contributions

MN: Conceptualization, Visualization, Writing – original draft, Writing – review & editing. BY: Conceptualization, Writing – review & editing. LZ: Writing – review & editing. GM: Writing – review & editing, Visualization. NP: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We thank Dr. Alison Ferrie for critical review of the manuscript.

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.

The handling editor RD declared a past co-authorship with the authors BY, LZ, and NP.

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

Aguilera, Y., Estrella, I., Benitez, V., Esteban, R. M., and Martín-Cabrejas, M. A. (2011). Bioactive phenolic compounds and functional properties of dehydrated bean flours. *Food Res. Int.* 44, 774–780. doi: 10.1016/j.foodres.2011.01.004

Ahmed, S., de la Parra, J., Elouafi, I., German, B., Jarvis, A., Lal, V., et al. (2022). Foodomics: A data-driven approach to revolutionize nutrition and sustainable diets. *Front. Nutr.* 9. doi: 10.3389/fnut.2022.874312

Asledottir, T., Vegarud, G. E., Picariello, G., Mamone, G., Lea, T. E., Røseth, A., et al. (2023). Bioactive peptides identified in pea and faba bean after *in vitro* digestion with human gastrointestinal enzymes. *J. Funct. Foods* 102, 105445. doi: 10.1016/j.jff.2023.105445

Awika, J. M., and Duodu, K. G. (2017). Bioactive polyphenols and peptides in cowpea (Vigna unguiculata) and their health promoting properties: A review. *J. Funct. Foods* 38, 686–697. doi: 10.1016/j.jff.2016.12.002

Barbana, C., and Boye, J. I. (2010). Angiotensin I-converting enzyme inhibitory activity of chickpea and pea protein hydrolysates. *Food Res. Int.* 43, 1642–1649. doi: 10.1016/j.foodres.2010.05.003

Bessada, S. M. F., Barreira, J. C. M., and Oliveira, M. B. P. P. (2019). Pulses and food security: Dietary protein, digestibility, bioactive and functional properties. *Trends Food Sci. Technol.* 93, 53–68. doi: 10.1016/j.tifs.2019.08.022

Bonvini, A., Coqueiro, A. Y., Tirapegui, J., Calder, P. C., and Rogero, M. M. (2018). Immunomodulatory role of branched-chain amino acids. *Nutr. Rev.* 76, 840–856. doi: 10.1093/nutrit/nuy037

Boye, J., Zare, F., and Pletch, A. (2010). Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Res. Int.* 43, 414–431. doi: 10.1016/j.foodres.2009.09.003

Campos-Vega, R., Loarca-Piña, G., and Oomah, B. D. (2010). Minor components of pulses and their potential impact on human health. *Food Res. Int.* 43, 461–482. doi: 10.1016/j.foodres.2009.09.004

Carbas, B., MaChado, N., Oppolzer, D., Ferreira, L., Queiroz, M., Brites, C., et al. (2020). Nutrients, antinutrients, phenolic composition, and antioxidant activity of common bean cultivars and their potential for food applications. *Antioxidants* 9, 186. doi: 10.3390/antiox9020186

Cosson, A., Oliveira Correia, L., Descamps, N., Saint-Eve, A., and Souchon, I. (2022). Identification and characterization of the main peptides in pea protein isolates using ultra high-performance liquid chromatography coupled with mass spectrometry and bioinformatics tools. *Food Chem.* 367, 130747. doi: 10.1016/j.foodchem.2021.130747

Daliri, E. B., Oh, D. H., and Lee, B. H. (2017). Bioactive peptides. *Foods* 6, 32. doi: 10.3390/foods6050032

Devries, M. C., McGlory, C., Bolster, D. R., Kamil, A., Rahn, M., Harkness, L., et al. (2018). Protein leucine content is a determinant of shorter- and longer-term muscle protein synthetic responses at rest and following resistance exercise in healthy older women: a randomized, controlled trial. *Am. J. Clin. Nutr.* 107, 217–226. doi: 10.1093/ajcn/nqx028

Di Stefano, E., Tsopmo, A., Oliviero, T., Fogliano, V., and Udenigwe, C. C. (2019). Bioprocessing of common pulses changed seed microstructures, and improved dipeptidyl peptidase-IV and α-glucosidase inhibitory activities. *Sci. Rep.* 9, 15308. doi: 10.1038/s41598-019-51547-5

Du, Z., and Li, Y. (2022). Review and perspective on bioactive peptides: A roadmap for research, development, and future opportunities. *J. Agric. Food Res.* 9, 100353. doi: 10.1016/j.jafr.2022.100353

Duranti, M. (2006). Grain legume proteins and nutraceutical properties. *Fitoterapia* 77, 67–82. doi: 10.1016/j.fitote.2005.11.008

EFSA Panel on Dietetic Products, N. and A. (NDA) (2010). Scientific Opinion on the substantiation of health claims related to branched-chain amino acids (BCAA) and growth or maintenance of muscle mass (ID 442, 444, 445, 447, 448, 451, 1478), attenuation of the decline in muscle power following exercise at high. *EFSA J.* 8, 1790. doi: 10.2903/j.efsa.2010.1790

European Commission. (2006). *Regulation (EC) No 1924/2006 of the European Parliament and of the Council of 20 December 2006 on nutrition and health claims made on foods*. Available at: https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32006R1924 (Accessed January 20, 2024).

Fan, H., Liu, H., Zhang, Y., Zhang, S., Liu, T., and Wang, D. (2022). Review on plant-derived bioactive peptides: biological activities, mechanism of action and utilizations in food development. *J. Futur. Foods* 2, 143–159. doi: 10.1016/j.jfutfo.2022.03.003

FAO/WHO (1991). "Protein quality evaluation: report of the Joint FAO/WHO Expert Consultation," in *Food and nutrition paper*, vol. 51. (FAO, Rome).

FAO/WHO (2013). "Dietary protein quality evaluation in human nutrition. Report of an FAQ Expert Consultation," in *FAO food and nutrition paper*, vol. 92. (FAO, Rome), 1–66. Available at: https://www.fao.org/ag/humannutrition/35978-02317b979a686a57aa4593304ffc17f06.pdf (Accessed January 8, 2024).

Foschia, M., Horstmann, S. W., Arendt, E. K., and Zannini, E. (2017). Legumes as functional ingredients in gluten-free bakery and pasta products. *Annu. Rev. Food Sci. Technol.* 8, 75–96. doi: 10.1146/annurev-food-030216-030045

Foyer, C. H., Lam, H.-M., Nguyen, H. T., Siddique, K. H. M., Varshney, R. K., Colmer, T. D., et al. (2016). Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants* 2, 16112. doi: 10.1038/nplants.2016.112

Garcia-Mora, P., Peñas, E., Frias, J., Gomez, R., and Martinez-Villaluenga, C. (2015). High-pressure improves enzymatic proteolysis and the release of peptides with angiotensin I converting enzyme inhibitory and antioxidant activities from lentil proteins. *Food Chem.* 171, 224–232. doi: 10.1016/j.foodchem.2014.08.116

Giusti, F., Caprioli, G., Ricciutelli, M., Vittori, S., and Sagratini, G. (2017). Determination of fourteen polyphenols in pulses by high performance liquid chromatography-diode array detection (HPLC-DAD) and correlation study with antioxidant activity and colour. *Food Chem.* 221, 689–697. doi: 10.1016/j.foodchem.2016.11.118

Guardado Félix, D., and Gutierrez Uribe, J. A. (2019). "Legumes and Oxidative Stress," in *Legumes: Nutritional Quality, Processing and Potential Health Benefits*. Ed. M.Á. Martín-Cabrejas (London, UK: The Royal Society of Chemistry). doi: 10.1039/9781788015721-00261

Hall, C., Hillen, C., and Garden Robinson, J. (2017). Composition, nutritional value, and health benefits of pulses. *Cereal Chem.* 94, 11–31. doi: 10.1094/CCHEM-03-16-0069-FI

Health Canda. (1981). *Determination of Protein Rating*. Available at: http://www.hcsc.gc.ca/fn-an/alt_formats/hpfb-dgpsa/pdf/res-rech/fo-1-eng.pdf (Accessed January 20, 2024).

Health Canda. (2023). *Measuring the protein quality of foods*. Available at: https://www.canada.ca/en/health-canada/services/food-nutrition/legislation-guidelines/policies/measuring-protein-quality-foods.html (Accessed January 20, 2024).

Hong, F., Ming, L., Yi, S., Zhanxia, L., Yongquan, W., and Chi, L. (2008). The antihypertensive effect of peptides: A novel alternative to drugs? *Peptides* 29, 1062–1071. doi: 10.1016/j.peptides.2008.02.005

Hou, D., Feng, Q., Niu, Z., Wang, L., Yan, Z., and Zhou, S. (2023). Promising mung bean proteins and peptides: A comprehensive review of preparation technologies, biological activities, and their potential applications. *Food Biosci.* 55, 102972. doi: 10.1016/j.fbio.2023.102972

Jakubczyk, A., and Baraniak, B. (2013). Activities and sequences of the angiotensin I-converting enzyme (ACE) inhibitory peptides obtained from the digested lentil (ens culinaris) globulins. *Int. J. Food Sci. Technol.* 48, 2363–2369. doi: 10.1111/ijfs.12226

Jakubczyk, A., and Baraniak, B. (2014). Angiotensin I Converting Enzyme Inhibitory Peptides Obtained after *In Vitro* Hydrolysis of Pea (Pisum sativum var. Bajka) Globulins. *BioMed. Res. Int.* 2014, 438459. doi: 10.1155/2014/438459

Jakubczyk, A., Karaś, M., Złotek, U., and Szymanowska, U. (2017). Identification of potential inhibitory peptides of enzymes involved in the metabolic syndrome obtained by simulated gastrointestinal digestion of fermented bean (Phaseolus vulgaris L.) seeds. *Food Res. Int.* 100, 489–496. doi: 10.1016/j.foodres.2017.07.046

Kamei, Y., Hatazawa, Y., Uchitomi, R., Yoshimura, R., and Miura, S. (2020). Regulation of skeletal muscle function by amino acids. *Nutrients* 12, 261. doi: 10.3390/nu12010261

Karkouch, I., Tabbene, O., Gharbi, D., Ben Mlouka, M. A., Elkahoui, S., Rihouey, C., et al. (2017). Antioxidant, antityrosinase and antibiofilm activities of synthesized peptides derived from Vicia faba protein hydrolysate: A powerful agents in cosmetic application. *Ind. Crops Prod.* 109, 310–319. doi: 10.1016/j.indcrop.2017.08.025

Kato, H., Takahashi, S., and Saito, K. (2011). Omics and integrated omics for the promotion of food and nutrition science. *J. Tradit. Complement. Med.* 1, 25–30. doi: 10.1016/S2225-4110(16)30053-0

Korhonen, H., and Pihlanto, A. (2006). Bioactive peptides: Production and functionality. *Int. Dairy J.* 16, 945–960. doi: 10.1016/j.idairyj.2005.10.012

Krasniqi, E., Boshnjaku, A., and Petrusevska Tozi, L. (2016). Approved health claims for amino acids in/as food supplements. *Maced. Pharm. Bull.* 62, 35–46. doi: 10.33320/maced.pharm.bull.2016.62.01.003

Li, H., and Aluko, R. E. (2010). Identification and inhibitory properties of multifunctional peptides from pea protein hydrolysate. *J. Agric. Food Chem.* 58, 11471–11476. doi: 10.1021/jf102538g

Liao, W., Cao, X., Xia, H., Wang, S., and Sun, G. (2022). Pea protein-derived peptides inhibit hepatic glucose production via the gluconeogenic signaling in the AML-12 cells. *Int. J. Environ. Res. Public Health* 19, 10254. doi: 10.3390/ijerph191610254

López-Barrios, L., Gutiérrez-Uribe, J. A., and Serna-Saldívar, S. O. (2014). Bioactive peptides and hydrolysates from pulses and their potential use as functional ingredients. *J. Food Sci.* 79, R273–R283. doi: 10.1111/1750-3841.12365

Maeda, T., Wakasawa, T., Shima, Y., Tsuboi, I., Aizawa, S., and Tamai, I. (2006). Role of polyamines derived from arginine in differentiation and proliferation of human blood cells. *Biol. Pharm. Bull.* 29, 234–239. doi: 10.1248/bpb.29.234

Marambe, P. W. M. L. H. K., and Wanasundara, J. P. D. (2012). "Seed storage proteins as sources of bioactive peptides," in *Bioactive Molecules in Plant Foods*. Ed. F. O. Uruakpa (Hauppauge, USA: Nova Science Publishers, Inc), 49–80.

Marinangeli, C. P. F., and House, J. D. (2017). Potential impact of the digestible indispensable amino acid score as a measure of protein quality on dietary regulations and health. *Nutr. Rev.* 75, 658–667. doi: 10.1093/nutrit/nux025

Martín-Cabrejas, M. A. (2019). "Legumes: An Overview," in *Legumes: Nutritional Quality, Processing and Potential Health Benefits*. Ed. M.Á. Martín-Cabrejas (London, UK: The Royal Society of Chemistry). doi: 10.1039/9781788015721-00001

Martineau-Côté, D., Achouri, A., Karboune, S., and L'Hocine, L. (2024). Antioxidant and angiotensin-converting enzyme inhibitory activity of faba bean-derived peptides after *in vitro* gastrointestinal digestion: insight into their mechanism of action. *J. Agric. Food Chem.* 72, 6432–6443. doi: 10.1021/acs.jafc.4c00829

Martineau-Côté, D., Achouri, A., Wanasundara, J., Karboune, S., and L'Hocine, L. (2022). Health beneficial bioactivities of faba bean gastrointestinal (*In vitro*) digestate in comparison to soybean and pea. *Int. J. Mol. Sci.* 23. doi: 10.3390/ijms23169210

Mathai, J. K., Liu, Y., and Stein, H. H. (2017). Values for digestible indispensable amino acid scores (DIAAS) for some dairy and plant proteins may better describe protein quality than values calculated using the concept for protein digestibility-corrected amino acid scores (PDCAAS). *Br. J. Nutr.* 117, 490–499. doi: 10.1017/S0007114517000125

Mojica, L., Luna-Vital, D. A., and González de Mejía, E. (2017). Characterization of peptides from common bean protein isolates and their potential to inhibit markers of type-2 diabetes, hypertension and oxidative stress. *J. Sci. Food Agric.* 97, 2401–2410. doi: 10.1002/jsfa.8053

Muzquiz, M., Varela, A., Burbano, C., Cuadrado, C., Guillamón, E., and Pedrosa, M. M. (2012). Bioactive compounds in legumes: pronutritive and antinutritive actions. Implications for nutrition and health. *Phytochem. Rev.* 11, 227–244. doi: 10.1007/s11101-012-9233-9

Naeem, M., Malik, M. I., Umar, T., Ashraf, S., and Ahmad, A. (2022). A comprehensive review about bioactive peptides: sources to future perspective. *Int. J. Pept. Res. Ther.* 28, 155. doi: 10.1007/s10989-022-10465-3

Nosworthy, M. G., Franczyk, A. J., Medina, G., Neufeld, J., Appah, P., Utioh, A., et al. (2017). Effect of processing on the *in vitro* and *in vivo* protein quality of yellow and green split peas (Pisum sativum). *J. Agric. Food Chem.* 65, 7790–7796. doi: 10.1021/acs.jafc.7b03597

Nosworthy, M. G., Hernandez-Alvarez, A. J., Franczyk, A. J., Medina, G., Neufeld, J., Arcand, Y., et al. (2023). Effect of cooking on the *in vitro* and *in vivo* protein quality of soy, oat and wheat varieties. *Cereal Chem.* 100, 460–472. doi: 10.1002/cche.10623

Nosworthy, M. G., Medina, G., Franczyk, A. J., Neufeld, J., Appah, P., Utioh, A., et al. (2018a). Effect of processing on the *in vitro* and *in vivo* protein quality of beans (Phaseolus vulgaris and vicia faba). *Nutrients* 10, 671. doi: 10.3390/nu10060671

Nosworthy, M. G., Medina, G., Franczyk, A. J., Neufeld, J., Appah, P., Utioh, A., et al. (2018b). Effect of processing on the *in vitro* and *in vivo* protein quality of red and green lentils (Lens culinaris). *Food Chem.* 240, 588–593. doi: 10.1016/j.foodchem.2017.07.129

Nosworthy, M. G., Medina, G., Franczyk, A. J., Neufeld, J., Appah, P., Utioh, A., et al. (2020). Thermal processing methods differentially affect the protein quality of Chickpea (Cicer arietinum). *Food Sci. Nutr.* 8, 2950–2958. doi: 10.1002/fsn3.1597

Padhi, E. M. T., and Ramdath, D. D. (2017). A review of the relationship between pulse consumption and reduction of cardiovascular disease risk factors. *J. Funct. Foods* 38, 635–643. doi: 10.1016/j.jff.2017.03.043

Pedroche, J., Yust, M. M., Girón-Calle, J., Alaiz, M., Millán, F., and Vioque, J. (2002). Utilisation of chickpea protein isolates for production of peptides with angiotensin I-converting enzyme (ACE)-inhibitory activity. *J. Sci. Food Agric.* 82, 960–965. doi: 10.1002/jsfa.1126

Rattanachaikunsopon, P., and Phumkhachorn, P. (2018). Functional food: what are they? and why are they so popular? *Acta Sci. Nutr. Heal.* 2, 26–17.

Roberts, A. (2016). The safety and regulatory process for amino acids in Europe and the United States12. *J. Nutr.* 146, 2635S–2642S. doi: 10.3945/jn.116.234591

Roy, F., Boye, J. I., and Simpson, B. K. (2010). Bioactive proteins and peptides in pulse crops: Pea, chickpea and lentil. *Food Res. Int.* 43, 432–442. doi: 10.1016/j.foodres.2009.09.002

Rui, X., Boye, J. I., Simpson, B. K., and Prasher, S. O. (2013). Purification and characterization of angiotensin I-converting enzyme inhibitory peptides of small red bean (Phaseolus vulgaris) hydrolysates. *J. Funct. Foods* 5, 1116–1124. doi: 10.1016/j.jff.2013.03.008

Rutherfurd-Markwick, K. J. (2012). Food proteins as a source of bioactive peptides with diverse functions. *Br. J. Nutr.* 108, S149–S157. doi: 10.1017/S000711451200253X

Sánchez, A., and Vázquez, A. (2017). Bioactive peptides: A review. *Food Qual. Saf.* 1, 29–46. doi: 10.1093/fqsafe/fyx006

Silveira, S. T., Martínez-Maqueda, D., Recio, I., and Hernández-Ledesma, B. (2013). Dipeptidyl peptidase-IV inhibitory peptides generated by tryptic hydrolysis of a whey protein concentrate rich in β-lactoglobulin. *Food Chem.* 141, 1072–1077. doi: 10.1016/j.foodchem.2013.03.056

Singh, B. P., Bangar, S. P., Alblooshi, M., Ajayi, F. F., Mudgil, P., and Maqsood, S. (2023). Plant-derived proteins as a sustainable source of bioactive peptides: recent research updates on emerging production methods, bioactivities, and potential application. *Crit. Rev. Food Sci. Nutr.* 63, 9539–9560. doi: 10.1080/10408398.2022.2067120

Singh, J., and Basu, P. S. (2012). Non-nutritive bioactive compounds in pulses and their impact on human health: an overview. *Food Nutr. Sci.* 03, 1664–1672. doi: 10.4236/fns.2012.312218

Soory, M. (2012). Nutritional antioxidants and their applications in cardiometabolic diseases. *Infect. Disord. - Drug Targets* 12, 388–401. doi: 10.2174/187152612804142233

Tak, Y., Kaur, M., Amarowicz, R., Bhatia, S., and Gautam, C. (2021). Pulse derived bioactive peptides as novel nutraceuticals: A review. *Int. J. Pept. Res. Ther.* 27, 2057–2068. doi: 10.1007/s10989-021-10234-8

Tuteja, N., Chandra, M., Tuteja, R., and Misra, M. K. (2004). Nitric oxide as a unique bioactive signaling messenger in physiology and pathophysiology. *J. Biomed. Biotechnol.* 2004, 498591. doi: 10.1155/S1110724304402034

Udenigwe, C. C., and Aluko, R. E. (2012). Food protein-derived bioactive peptides: production, processing, and potential health benefits. *J. Food Sci.* 77, R11–R24. doi: 10.1111/j.1750-3841.2011.02455.x

Varshney, R. K., Close, T. J., Singh, N. K., Hoisington, D. A., and Cook, D. R. (2009). Orphan legume crops enter the genomics era! *Curr. Opin. Plant Biol.* 12, 202–210. doi: 10.1016/j.pbi.2008.12.004

Wu, G. (2009). Amino acids: metabolism, functions, and nutrition. *Amino Acids* 37, 1–17. doi: 10.1007/s00726-009-0269-0

Wu, G. (2010). Functional amino acids in growth, reproduction, and health. *Adv. Nutr.* 1, 31–37. doi: 10.3945/an.110.1008

Wu, G., Meininger, C. J., McNeal, C. J., Bazer, F. W., and Rhoads, J. M. (2021). *Role of L-Arginine in Nitric Oxide Synthesis and Health in Humans BT - Amino Acids in Nutrition and Health: Amino Acids in Gene Expression, Metabolic Regulation, and Exercising Performance*. Ed. G. Wu (Cham: Springer International Publishing), 167–187. doi: 10.1007/978-3-030-74180-8_10

Xu, B. J., and Chang, S. K. C. (2007). A comparative study on phenolic profiles and antioxidant activities of legumes as affected by extraction solvents. *J. Food Sci.* 72, S159–S166. doi: 10.1111/j.1750-3841.2006.00260.x

Yust, M.d. M., Millán-Linares, M.d. C., Alcaide-Hidalgo, J. M., Millán, F., and Pedroche, J. (2012). Hypocholesterolaemic and antioxidant activities of chickpea (Cicer arietinum L.) protein hydrolysates. *J. Sci. Food Agric.* 92, 1994–2001. doi: 10.1002/jsfa.5573

Zaky, A. A., Simal-Gandara, J., Eun, J.-B., Shim, J.-H., and Abd-El-Aty, A. M. (2022). Bioactivities, applications, safety, and health benefits of bioactive peptides from food and by-products: A review. *Front. Nutr.* 8. doi: 10.3389/fnut.2021.815640

Zan, R., Wu, Q., Chen, Y., Wu, G., Zhang, H., and Zhu, L. (2023). Identification of novel dipeptidyl peptidase-IV inhibitory peptides in chickpea protein hydrolysates. *J. Agric. Food Chem.* 71, 8211–8219. doi: 10.1021/acs.jafc.3c00603

Zhang, M., Zhu, L., Wu, G., Liu, T., Qi, X., and Zhang, H. (2022). Rapid screening of novel dipeptidyl peptidase-4 inhibitory peptides from pea (Pisum sativum L.) protein using peptidomics and molecular docking. *J. Agric. Food Chem.* 70, 10221–10228. doi: 10.1021/acs.jafc.2c03949

Zhang, M., Zhu, L., Zhang, H., Wang, X., Liu, T., Qi, X., et al. (2023). Identification of novel dipeptidyl peptidase-4 inhibitory peptides from pea proteins: A combined in silico and *in vitro* study. *Food Biosci.* 56, 103374. doi: 10.1016/j.fbio.2023.103374

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 17 February 2025
DOI 10.3389/fpls.2024.1395952

Check for updates

OPEN ACCESS

EDITED BY
George V. Popescu,
Mississippi State University, United States

REVIEWED BY
Tamas Varga,
Pacific Northwest National Laboratory (DOE), United States
João Paulo Rodrigues Marques,
University of São Paulo, Brazil

*CORRESPONDENCE
Teagen D. Quilichini
teagen.quilichini@usask.ca

RECEIVED 04 March 2024
ACCEPTED 26 December 2024
PUBLISHED 17 February 2025

CITATION
Ashe P, Tu K, Stobbs JA, Dynes JJ, Vu M, Shaterian H, Kagale S, Tanino KK, Wanasundara JPD, Vail S, Karunakaran C and Quilichini TD (2025) Applications of synchrotron light in seed research: an array of x-ray and infrared imaging methodologies. *Front. Plant Sci.* 15:1395952.
doi: 10.3389/fpls.2024.1395952

COPYRIGHT
© 2025 His Majesty the King in Right of Canada. 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.

Applications of synchrotron light in seed research: an array of x-ray and infrared imaging methodologies

Paula Ashe [1], Kaiyang Tu [2], Jarvis A. Stobbs [2], James J. Dynes [2], Miranda Vu [2], Hamid Shaterian [1], Sateesh Kagale [1], Karen K. Tanino [3], Janitha P. D. Wanasundara [4], Sally Vail [4], Chithra Karunakaran [2] and Teagen D. Quilichini [1,5]*

[1]Aquatic and Crop Resource Development, National Research Council Canada, Saskatoon, SK, Canada, [2]Canadian Light Source Inc., Saskatoon, SK, Canada, [3]Department of Plant Sciences, College of Agriculture and Bioresources, University of Saskatchewan, Saskatoon, SK, Canada, [4]Agriculture and Agri-Food Canada, Saskatoon Research Centre, Saskatoon, SK, Canada, [5]Department of Biology, College of Arts and Science, University of Saskatchewan, Saskatoon, SK, Canada

Synchrotron radiation (SR) provides a wide spectrum of bright light that can be tailored to test myriad research questions. SR provides avenues to illuminate structure and composition across scales, making it ideally suited to the study of plants and seeds. Here, we present an array of methodologies and the data outputs available at a light source facility. Datasets feature seed and grain from a range of crop species including *Citrullus* sp. (watermelon), *Brassica* sp. (canola), *Pisum sativum* (pea), and *Triticum durum* (wheat), to demonstrate the power of SR for advancing plant science. The application of SR micro-computed tomography (SR-μCT) imaging revealed internal seed microstructures and their three-dimensional morphologies in exquisite detail, without the need for destructive sectioning. Spectroscopy in the infrared spectrum probed sample biochemistry, detailing the spatial distribution of seed macronutrients such as lipid, protein and carbohydrate in the embryo, endosperm and seed coat. Methods using synchrotron X-rays, including X-ray absorption spectroscopy (XAS) and X-ray fluorescence (XRF) imaging revealed elemental distributions, to spatially map micronutrients in seed subcompartments and to determine their speciation. Synchrotron spectromicroscopy (SM) allowed chemical composition to be resolved at the nano-scale level. Diverse crop seed datasets showcase the range of structural and chemical insights provided by five beamlines at the Canadian Light Source, and the potential for synchrotron imaging for informing plant and agricultural research.

KEYWORDS

synchrotron, micro-computed tomography (μCT), mid-infrared spectroscopy, x-ray absorption spectroscopy (XAS), x-ray fluorescence spectroscopy (XRF), spectromicroscopy (SM), seed, embryo

1 Introduction

A synchrotron facility functions in the acceleration then redirection of electrons along a curved path; a process that emits strong radiation (Balerna and Mobilio, 2014). Synchrotron light or synchrotron radiation (SR) spans the electromagnetic spectrum, from low energy infrared, through soft X-rays to high energy hard X-rays, and is extremely bright, with high flux per unit area (Bharti and Goyal, 2019). The characteristic brightness and energy tunability of SR provides an adaptive tool with broad applications in science and medicine. Although applications of SR in medical research are prevalent, this tool has yet to become commonplace in agricultural research. The lag in synchrotron adoption by the life sciences, and particularly plant science research may result from perceived access restrictions to these facilities, and uncertainty among scientists surrounding the applications available and insights possible.

SR provides a unique tool for obtaining vast structural and chemical insights on the inner workings of organisms. In seed-bearing plants, offspring are efficiently packaged in a protected capsule to support independent survival. Seeds consist of three distinct structures, the embryo, endosperm and seed coat, which are intimately linked by physical proximity but differ genetically and biochemically (Xiong et al., 2021). Within a seed, the progeny progresses from a single-celled zygote to an embryo with shoot and root axes that is poised for germination (Ramalho et al., 2021). The developing embryo is surrounded by multinucleated cells of the triploid endosperm that primarily functions as a nutritive tissue (Sabelli and Larkins, 2009; Doll and Ingram, 2022). The abundance of endosperm varies dramatically among species, persisting as a prominent storage tissue in monocotyledonous (monocot) plants including cereals such as wheat, rice and maize, or absent in mature seeds of many dicotyledonous (dicot) plants, including legumes and brassicas (Sabelli and Larkins, 2009). A seed coat forms the outermost casing of a seed and envelops the endosperm and embryo within a hardened exterior. As a seed develops, the seed coat forms at the interface between offspring and mother plant, and later between offspring and the external environment (Radchuk and Borisjuk, 2014; Huss and Gierlinger, 2021). This crucial locale requires seed coats to be multifunctional, including as a conduit for nutrient transfers between generations during early seed development, then as a fortified, often dead, barrier that protects the filial generation and mediates seed dispersal, dormancy and germination. Beyond the crucial function seeds serve in the reproductive success of many plants, seeds and grains serve as a primary source of calories for human and animal diets, supplying diverse nutrient profiles rich in protein, fiber, fats, vitamins, and minerals (Preedy et al., 2010). Thus, the agricultural sector requires vigorous nutrient-dense seeds that meet consumer demands, maintain viability after sustained storage, and yield crops that are productive and stress resilient. For seed research, synchrotron imaging is particularly informative for resolving structural relationships amongst seed tissues to support quality, health and viability assessments, and for mapping compositional data to each seed subcompartment to assess nutrient landscapes and commercial value.

Adoption of synchrotron-based imaging in plant and agricultural research has grown in recent years, but is generally in its infancy (Vijayan et al., 2015; Indore et al., 2022). Research strategies that uncover the internal structural phenotypes and nutritional content of seeds offer opportunities to identify preferred traits and targets for improvement. Here, we showcase an assortment of seed datasets to demonstrate the potential for synchrotron imaging in plant science, using the Canadian Light Source (CLS) in Saskatoon, Canada. Methods presented cover seed preparation options to analyze mature, dry seeds as well as fresh, developing seeds undergoing embryogenesis at five CLS beamlines. Seeds analyzed come from a variety of valued dicot and monocot crops, including oil-rich canola, protein-rich pea, and carbohydrate-rich wheat. We feature CLS beamlines that deploy hard and soft X-rays, deemed widely relevant to life science research and available at synchrotron facilities around the globe. For each beamline, basic capabilities are detailed and examples are provided to demonstrate data output potential. Briefly, seed three-dimensional (3D) structure is analyzed in high resolution with synchrotron-based micro-computed tomography (SR-µCT) at the biomedical imaging and therapy low-energy beamline (BMIT-BM, 12.6 – 40.0 keV spectral range). Seed protein, carbohydrate and lipid macronutrients are spatially mapped across seeds with mid-infrared spectroscopy (MidIR spectroscopy 560-6000 cm^{-1}; with and without synchrotron light). Micronutrients, including calcium (Ca), potassium (K), manganese (Mn), magnesium (Mg), cadmium (Cd), zinc (Zn), iron (Fe), and phosphorus (P) are spatially mapped across seed tissues using a combination of hard X-ray fluorescence (XRF) imaging at the Biological X-ray absorption spectroscopy beamline (BioXAS-Imaging, at 5 – 21 keV), and XRF imaging at the soft X-ray microcharacterization beamline (SXRMB, 1.7 – 10 keV). The chemical speciation of elements is examined using X-ray Absorption Near Edge Structure (XANES) spectroscopy, to distinguish distributions of Zn (at BioXAS-Imaging). Fine elemental mapping with soft X-ray spectromicroscopy (SM, 130 – 3000 eV) is used to characterize subcellular distributions of P.

This article uses case studies and select examples to describe methods and to highlight the utility of synchrotron imaging technologies for seed research. While the article presents novel structural and compositional datasets, the article is not intended to describe complete research stories.

Case Study 1: SR-µCT of *Acidovorax citrulli* infected watermelon seeds. Bacterial fruit blotch caused by *Acidovorax citrulli* is a global threat to members of the Cucurbitaceae family such as watermelon and seeds are the primary source of inoculum. There are no known sources of resistance to this pathogen and antibacterial seed treatments have limited efficacy (Burdman and Walcott, 2012; Carvalho et al., 2012). Pathogen localization in the seed is related to the mode of infection and penetration into the embryo occurs with pistil infection (Dutta et al., 2012). SR-µCT provides the unique ability to visualize the extent of pathogen infection and resulting necrotic damage in 3D space. This application of SR-µCT imaging could be useful for testing outcomes of seed disinfestation treatments and for studies on the pathogenesis of plant disease.

Case Study 2: Multi-beamline analysis of the anatomic and compositional changes in pea seed tissues during development. Developing seeds require special consideration for comparison between genotypes and treatments, as developmental staging is critical to make informed conclusions. Destructive seed dissection to visualize the embryo can be used for developmental staging but results in the loss of 3D spatial information. Sectioning for microscopy and the resulting 2D images requires careful orientation of the plane of section to view developing embryos. SR-µCT can provide 3D visualization of the embryo for accurate developmental staging as well as provide cell layer specific information through the seed coat and embryo and volumetric analysis of seed components. This information across multiple pollination days, genotypes and treatments can be used to accurately match samples for subsequent 2D chemometric imaging. This case study provides an example of semi-correlative microscopy of developing seeds across multiple beam lines.

Case Study 3: Phosphorus distribution in high and low phytate canola genotypes. Phytate is the main storage form of P and is a known anti-nutritional factor due to chelation of cationic elements (Thompson, 1990; Madsen and Brinch-Pedersen, 2020). In canola, phytate is stored in globoids inside protein storage vacuoles primarily in the cotyledons. Multiple beamlines were used to interrogate the distribution of protein, lipid and P in canola genotypes differing in their phytate content. This case study is not only an example of cross beam correlative and semi-correlative microscopy, but also serves to demonstrate the use of beamlines with lower energies to better interrogate low atomic weight elements such as P. Limitations and considerations for examining P in biological samples are also discussed.

2 Materials and methods

2.1 Overview of the Methods

The methods in our study provide generalized procedures to prepare seeds for synchrotron imaging analysis (Figure 1). Given the heterogeneity amongst seeds, and the breadth of research questions that synchrotron light can address, the methodologies describe general workflow recommendations, as detailed protocols that will optimally prepare all specimens for synchrotron-based

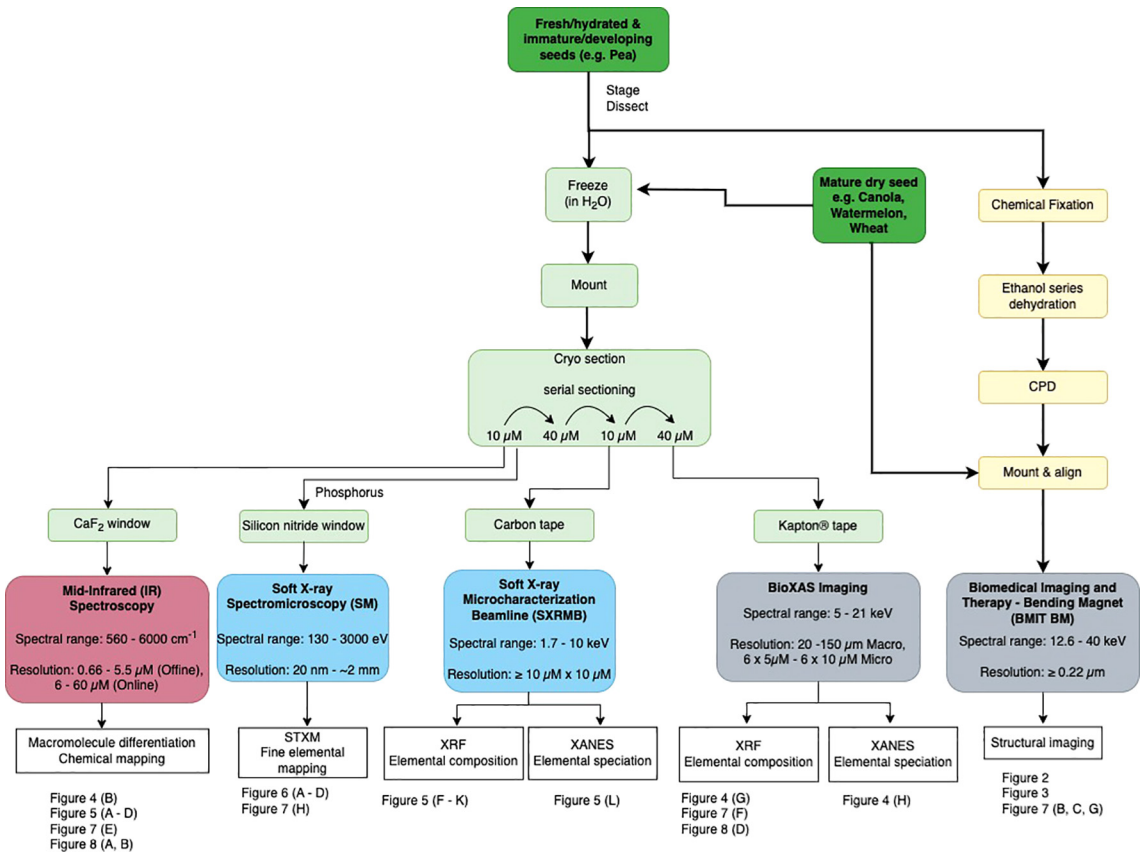


FIGURE 1
Workflow for imaging seeds with synchrotron light, for five agriculturally relevant beamlines at the Canadian Light Source. Dark green boxes indicate seed sample entry points. Procedural steps (detailed in Methods) to prepare seeds for each beamline are outlined, beginning with specimen input options (dark green), then methodologies to prepare specimens for the BMIT-BM beamline (yellow), and BioXAS-Imaging, SXRMB SM and MidIR spectroscopy beamlines (pale green). Beamlines are colour-coded according to the spectrum of electromagnetic radiation employed; red uses infrared light, blue use soft X-rays, and grey use hard X-rays. Beamline spectral range (in keV) and resolution capabilities are included with beamline titles. Resolutions reflect pixel size options for output data. Data outputs of each beamline, as featured in this study, are summarized in white boxes. Reference examples of the data outputs available are indicated by figure numbers below each beamline.

research are not feasible. The methods are intended to provide a framework for new users to develop, adapt and optimize approaches according to their research needs. We present methods and synchrotron imaging data for diverse crop seeds, known for their unique abundance of oil, proteins or carbohydrates, to demonstrate the potential for implementing these methods in many crop species and experimental pursuits. Methods include recommended approaches for preparing mature seed as well as immature, fresh seed in which embryos are developing. Each beamline used in this study has specific sample preparation requirements. Therefore, we outline key beamline-specific considerations for appropriately preparing materials. Readers are encouraged to consult additional published works employing synchrotron light to study seeds and seed germination, including studies employing XRF imaging (Kim et al., 2006; Iwai et al., 2012; Kopittke et al., 2018; Romeu et al., 2021; van der Ent et al., 2022; Montanha et al., 2022; Ren et al., 2022) and synchrotron X-ray CT approaches (Yamauchi et al., 2012; Rousseau et al., 2015; Le et al., 2019; Legland et al., 2022; Indore et al., 2024). Examples and reviews of seed MidIR spectroscopy can be found in Feng et al., 2020; Hossain et al., 2022; Guendel et al., 2018; Türker-Kaya and Huck, 2017; Yan et al., 2020 and references therein. It is advised that experimental design and methodologies are reviewed in consultation with synchrotron experts prior to initiating new studies, to appropriately tailor protocols according to beamline standards and recommendations.

2.2 Seeds

- Pea (*Pisum sativum*) seeds (USDA seed repository Pullman, WA).
- Canola (*Brassica napus*) genotype DH4079, a doubled haploid line obtained from the Swedish spring cultivar, Topas, (Dr. Alison Ferrie, NRC, Saskatoon). Two genotypes were selected from a *Brassica napus* diversity collection based on phytate levels (NAM13; high phytate and NAM45; low phytate) (Ebersbach et al., 2022).
- Wheat (*Triticum durum*) cultivar 'Pelissier' seeds (Dr. Raju Soolanayakanahally, Agriculture and Agri-Food Canada, Saskatoon).
- Watermelon (*Citrullus lanatus*) seeds infected with *Acidovorax citrulli*, a seedborne bacterium known to cause bacterial fruit blotch disease (Dr. Greg Welbaum, Virginia Tech) (Eckshtain-Levi et al., 2014).

2.3 Preparing seeds for synchrotron-based micro-computed tomography

Plant growth should involve controlled conditions, to remove unintended variations in seed structure and/or composition. For SR-based micro-computed tomography (SR-µCT) imaging (BMIT-BM) mature seeds do not require fixation or drying, and can be directly mounted to the stage and imaged. For SR-µCT imaging of fresh materials including immature seeds spanning embryogenesis, chemical fixation and dehydration are recommended (Section 2.3.1). For chemical imaging (MidIR spectroscopy, BioXAS-Imaging, SXRMB and SM), seeds must be cryo-sectioned and dried on beamline-appropriate substrates (Section 2.4.2).

- For preparing immature seeds, staging flowers as they form can assist in developmental timing decisions.

2.3.1 Chemical fixation and dehydration of immature pea seeds for SR-µCT imaging (BMIT-BM)

- An overview of the chemical fixation protocol recommended for fresh, immature seeds is provided here, with additional protocol details and suggestions provided (See Supplementary Methods).
- Use a dissecting stereo microscope and forceps (Dumont #5 forceps, Fine Science Tools; Cedarlane, Burlington, ON) to remove developing seeds from the parent plant without damaging the seed coat. Minimize time from dissection to chemical fixation (for SR-µCT).
- Recommended fixative solution: Glutaraldehyde (25% aqueous glutaraldehyde, Electron Microscopy Sciences; Cedarlane Burlington, ON) diluted to a final concentration of 2% (v/v) in 25 mM PIPES buffer, pH 7.0 (PIPES or 1,4-Piperazinediethanesulfonic acid, MilliporeSigma, Oakville, ON). Prepare fresh fixative solution on the day it will be used.
- To fix, submerge seeds in fixative solution in glass scintillation vials and leave at room temperature for 2 hours or overnight at 4°C. Plant material to fixative ratio should be 1:1000.
- Following fixation, replace with 25 mM PIPES buffer, pH 7.0 without glutaraldehyde, while ensuring samples remain submerged. Repeat 5 times, to ensure fixative has been removed.
- Dehydrate samples in ascending ethanol; 30%, 50%, 70%, 95%, 100% with a minimum 20 minute incubation at room temperature between each exchange. Ethanol (anhydrous absolute ethanol, Greenfield Global, Toronto, ON) solutions can be made with ultrapure water (Synergy UV Water Purification System, MilliporeSigma, Oakville, ON).
- Repeat a minimum of three 100% ethanol exchanges for 20 minutes each. For materials exceeding 1 cm in diameter, it is recommended to lengthen time between exchanges, and to allow seeds to remain in 100% ethanol for a minimum of 48 hours prior to critical point drying.
- Critical point dry the samples using stasis mode (Autosamdri®-931 Critical Point Dryer, Tousimis Research Corp, Rockville, MD); 3 cycles with solvent-substituted liquid CO_2 (as a drying agent). Thoroughly dried samples should have no detectable ethanol odour.
- Store samples in a sealed container away from light, with dust-free desiccant.

2.3.2 Mount seeds for SR-μCT imaging (BMIT-BM)

- Affix seed to a stub (SEM pin stubs, Ted Pella Inc, Redding, CA). Hot glue is well suited to mounting seeds, including critical pointed dried materials. Other mounting substrates such as dental wax or double sided carbon tape are also acceptable. It is critical that the sample is firmly mounted to eliminate motion during scanning.
- Sample can be stored in an SEM stub box (Ted Pella Inc, Redding, CA), or mounted directly on a goniometer head for SR-μCT imaging (see Section 2.5).

2.4 Preparation of seed sections for MidIR, BioXAS-imaging, SXRMB and SM

Accessing spatially-resolved internal chemistry typically requires thin cryo-sectioning of biological materials. For all chemical imaging methods described by this study, seed materials could not be chemically fixed or embedded in resin prior to sectioning, as this alters metabolic composition. This is critical for MidIR spectroscopy, however, if only elemental analysis is required (XRF or XANES), sample processing, such as fixation, with careful consideration is possible (for review see Pushie et al., 2022). Ultrapure water is predominantly used as an embedding medium to freeze seeds in preparation for cryosectioning. The cryo-sectioning methods outlined below are for samples embedded in water, where measurements using IR and XRF require clean analysis of the outermost seed layers (e.g. seed coat and aleurone). If analysis is restricted to inner tissues such as the embryo, and/or IR will not be used, direct embedding of seeds into cryo-gel (Pelco cryo-embedding compound; Ted Pella Inc, Redding, CA) or other non-infiltrating media is sufficient.

2.4.1 Cryo-preservation of seeds for cryo-sectioning

- Wear gloves throughout sample preparation. Do not touch samples or substrates as any contamination will impact subsequent analysis.
- Hydrated seed drastically improves the quality of cryo-sections. For mature seed, soaking pretreatment in ultrapure water for ~8 h to overnight is typically sufficient for re-hydration. For fresh seed including immature or green seed, no re-hydration step prior to embedding is required.
- Samples are frozen in water to ensure a hydration sphere around the tissue as this will minimize artifacts in subsequent analysis. Using water as the embedding medium ensures samples can be analyzed by IR. Ultrapure water (Synergy UV Water Purification System, MilliporeSigma, Oakville, Ontario) is recommended.
- Carefully place an isolated seed in a capsule (BEEM® embedding capsules, Electron Microscopy Sciences; Cedarlane Burlington, ON) or 2 mL microfuge tube (Avantor, Mississauga, ON) filled with water. The container should be selected to ensure sufficient space for the seed and at least a thin shell of surrounding water. Minimize time from dissection of seeds to cryo-preservation. Orient the sample to aid subsequent mounting for cryo-sectioning, if possible.
- Flash freeze the sample in liquid nitrogen (in a dewar) and store at -80˚C until cryo-sectioning. To minimize cryodamage to tissues due to the formation of ice crystals during flash freezing, samples may alternatively be frozen at higher temperatures in a slurry of dry ice/isopentane, liquid ethane or propane or under high-pressure. It should be noted that cryoprotectants such as sucrose or glycerol are not suitable as they will interfere with the generated MidIR spectra. Avoid storage exceeding 6 months.

2.4.2 Cryo-sectioning seeds

- Set the cryostat (Leica CM1950 cryostat, Leica Biosystems, Concord, ON) to the desired sectioning temperature and let it equilibrate. Temperature for sectioning will depend on the composition and nature of the sample as well as environmental conditions in the lab (e.g. humidity). A reasonable starting temperature is -20˚C. If the sections stick or roll, decrease the temperature. If the sections crack, increase the temperature. It is recommended to begin with a sacrificial sample to optimize sectioning parameters.
- In the cryostat, use cryo-mounting medium (Pelco cryo-embedding compound, Ted Pella Inc, Redding, CA) to secure the sample to a pre-cooled cryostat chuck.
- Remove the sample from the tube by warming the sides of the tubes until the ice block containing the sample slides out.
- Excess ice may be trimmed away, but a thin layer must be kept to ensure separation between the sample and the mounting medium.
- Equilibrate the mounted sample to the cryostat prior to initiating cryo-sectioning.
- The recommended substrates for collecting cryosections differ according to the intended method of analysis. Cryosection substrate options include:
    - For MidIR spectroscopy (online and offline), calcium fluoride (CaF_2) windows (Calcium Fluoride polished windows, 25 mm x 1 mm thick; Crystran, Dorset, UK). Thorough cleaning of windows ensures a clean background for IR measurements, and is crucial anytime a CaF_2 window is re-used. To clean, rinse with ultrapure water followed sequentially by ethanol, isopropanol, and acetone.
  - For BioXAS or SXRMB, Kapton® (Avantor, Mississauga, ON) or carbon tape (Ted Pella Inc, Redding, CA).

– For SM, silicon nitride windows (5 mm frame, 0.5 mm aperture, 200 nm membrane; Ted Pella Inc, Redding, CA).
– Substrates for cryo-sections should be at room temperature to aid section transfers from the cryostat.
- Section into frozen specimen block with 60 – 80 µm steps to approach region of interest, then reduce step size. Inspect sections with a compound microscope (Revolve Echo, San Diego, CA) to ensure region of interest has been reached.
- For MidIR, section at a thickness of 6 to 12 µm and collect sections on a clean CaF_2 window. Multiple sections can be collected on a single window.
– Section thickness should be decided for the sample and resolution required. 6 µm is ideal, but for some seeds this might not be achievable. 10-12 µm is a compromise to obtain high-quality sections that still allow for IR transmission through most IR regions of interest. Immediately prior to the transfer of a section to a CaF_2 window, bring the window into close proximity with a deionizer for electrostatic discharge (Haug Biel Deionizer, Mettler Toledo, Mississauga, ON). This prevents unintended movement of sections during collection, and may only be necessary in dry laboratory conditions.
– To transfer a section onto a CaF_2 window, bring the window into the cryostat (wearing gloves and holding the edges only), and position it directly above the section. The section should jump onto the window and dry rapidly. A pre-cooled paintbrush can be used to facilitate transfer. If transfer results in folding or wrinkles in the section, it can be removed with ultrapure water and isopropanol or acetone.
– Store window in a sterile mini petri dish, section side up, at room temperature. Avoid exposure to light to prevent degradation. Note, for IR, samples must be dry prior to analysis to avoid water interference in the spectra and to prevent movement during the measurement period.
- For BioXAS and SXRMB, section at a thickness of 20 to 80 µm and collect the section on Kapton® or Carbon tape, respectively.
– Section thickness should be decided based on the sample and resolution required. Thinner samples are better for resolution but increase dwell time (requiring longer measurements). A rough guide is to section 1 - 2 cell layers. For pea, sections collected were 40 µm.
– To collect a section on Kapton® or Carbon tape, ensure the block face is trimmed and sections of optimal thickness can be attained. Cut a piece of Kapton®/Carbon tape long enough to span the mini petri dish the section will be stored in. Place the adhesive side directly on the block face (with the long edge of the tape running parallel to the cutting plane, perpendicular to the knife edge) and use a pre-cooled cryostat roller to ensure thorough contact between the tape and sample block face. Proceed with section cutting, gently holding the tape along the bottom edge of the block face, to allow contact between the block and knife edge.
– For storage, suspend the section in the middle of a sterile petri dish by connecting the outer edges of the Kapton®/Carbon tape with the dish and store at room temperature. In some instances, such as with mobile elements, freeze drying of samples may be required. Additionally, some samples may be stored frozen as, at some beamlines, cryo-measurements with hydrated tissue are possible. Avoid exposure to light to prevent degradation due to oxidation.

When cross-beamline comparisons between regions of interest within a sample will be made, interpretation can be facilitated by selecting cryo-sections in close proximity. If collecting serial cryo-sections, sections should be interspersed rather than collecting for each beamline separately. For example, after collecting a 10 µm section on CaF_2 window, shift the cryostat to 40 µm step size, cut one section to shift mechanism to new thickness, then collect subsequent 40 µm section on Kapton® tape. Note: for small features such as a globular-stage embryo, the number of serial sections possible is finite.

2.5 Experimental parameters for synchrotron radiation-based micro-computed tomography (SR-µCT at BMIT-BM)

At the CLS, two beamlines for SR-µCT analyses are available at BMIT (https://bmit.lightsource.ca/) (Gasilov et al., 2024). The low energy bending magnet (BM) uses 12.6 – 40 keV hard X-rays and the higher energy insertion device (ID) beamline uses 28 – 140 keV. SR-µCT of seeds in this study used the BMIT-BM beamline.

- Detectors. There are multiple detector set-ups available which include both monochromic and white beam at the BMIT beamline at the CLS. These set-ups consist of a scintillator, objective lens and eye piece (0.9x) lens coupled with a camera. Using various magnifications and/or camera combinations, a variety of pixel sizes and fields of view (FOV) can be achieved (for more information see Gasilov et al. 2024). Available objectives are 2X, 5X, 7.5X, 10X and 20X. Objective changes require experienced beamline scientists and time for re-alignment. It is recommended to consult with experienced synchrotron scientists to tailor the imaging set-up to each experiment.
- Filters. When using a filtered white beam, a broad band polychromatic beam is produced. The correct filter combination for collecting SR-µCT scans of live plant materials will limit radiation damage. There are a number of filters and thicknesses available at the BMIT-BM beamline, including aluminum, glassy carbon, silver, molybdenum, copper, and tin. See Supplementary Table 1 for a list of filters.

- Stage holder. Place the sample on a goniometer head and secure to the beamline stage holder. A goniometer (Goniometer head 1005, Huber Diffraktionstechnik GmbH & Co. KG, Rimsting, Germany) allows for control of sample orientation in the x, x', y, y' and z axis to ensure optimal position adjustment for centering the sample within the field of view limitation of the beam. Note that a variety of custom holders are available.
- Data collection. Samples are rotated within the beam over 180° for conventional µCT for 500 - 3000 projections and over 360° for half-acquisition µCT with 1000 - 6000 projections. The range of the vertical stage is 30 - 90 mm depending on the setup. Multiple vertical views can be stitched together in post-processing. The time to scan each vertical view is dependent on the exposure time and the number of projections; as an example, each vertical view for pea took ~5 minutes. Details for all SR-µCT scans are shown in Table 1.

2.6 Experimental parameters for Mid-IR spectroscopy (Mid-IR beamline)

Multiple approaches, with or without synchrotron light, are available for the detection of biochemical compounds using Mid-IR spectroscopy depending on the resolution required. Access to both synchrotron IR and offline imaging instrumentation is provided at the Mid-IR beamline at the CLS. When larger samples are scanned, or high resolution with low noise spectra is not required, a globar source with an FPA detector is sufficient (offline). Synchrotron IR (SIR) mapping is used to optimize IR spectra quality, and may be used for acquisitions at the cellular level.

2.6.1 Chemical imaging analysis (developing pea seeds, canola and wheat)

IR imaging maps were obtained at the Mid-Infrared beamline (https://www.lightsource.ca/facilities/beamlines/cls/beamlines/mid-ir.php) at the Canadian Light Source. Offline IR images were acquired at the Agilent end station, equipped with an Agilent FTIR microscope and spectrometer (Agilent Technologies Cary 620 FTIR microscope with 128x128 pixel Focal Plane Array detector, 25x objective and 670 spectrometer, Santa Clara, CA). The spectra were collected in transmission mode in the Mid-IR range of 3900 - 900 cm^{-1} with 16 coadded scans and a spectral resolution of 4 cm^{-1}. For the parameters used to scan pea sections, each mm^2 of tissue area took approximately 13 minutes. The SIR images were acquired using the Bruker end station with a Bruker Vertex70v spectrometer (Bruker Optics Inc, Billerica, MA) with a Hyperion 3000 microscope (MCT-detector) equipped with a 36x objective and 0.3 circular objective in transmission mode in the Mid-IR range of 4000 - 900 cm^{-1} with 8 coadded scans and a spectral resolution of 4 cm^{-1} (Supplementary Table 2).

For offline and SIR measurements, section(s) on IR transparent windows were secured to a sample stage. A visible microscope image and IR background were acquired before sample measurement. For the offline Agilent end station, individual IR image tiles were acquired by the FPA detector and stitched together into a mosaic image by the instrument software (see Section 2.11.2 for data analysis). For the SIR Bruker end station, seed IR images were collected with point-by-point mapping the SIR beam spot over the area of interest and collecting an IR spectra at each pixel.

2.7 Experimental parameters for XRF-imaging (BioXAS imaging and SXRMB)

There are different beamlines available for XRF imaging, which employ hard X-rays (BioXAS-Imaging) or soft X-rays (SXRMB) to detect different elements. SGM (250 - 2000 eV) and VESPERS (6 - 30 keV) beamlines at the CLS provide XRF imaging with additional energy ranges and capabilities. At BioXAS Imaging, there are macro and micro imaging modes provide unique resolution capabilities. It is important to select dwell times that ensure adequate detection for the element(s) of interest. Prior to initiating an experiment, discussion with experienced beamline staff is recommended to address desired outcomes. See Supplementary Table 3 for specifications of data collection across all samples.

2.7.1 XRF elemental imaging at BioXAS-imaging

XRF-imaging was performed at the BioXAS-Imaging beamline (https://www.lightsource.ca/facilities/beamlines/cls/beamlines/bioxas-imaging.php#Techniques) at the Canadian Light Source. Here, XRF elemental imaging at BioXAS-Imaging was used to investigate elemental spatial distributions in cryo-sectioned immature pea seed, mature canola seed and mature wheat grain. BioXAS-Imaging is uniquely suited to accessing biologically important trace elements such as Cl, Ca, Zn, Fe, and K based on its X-ray energy range of 5 to 20 keV. At the beamline, high brilliance X-ray light is generated by an in-vacuum undulator insertion device. The primary optics of the beamline consist of a collimating mirror, a fixed-exit double crystal monochromator, and a post-monochromator vertically focusing mirror. For macro-mode imaging, the spot size of the incident X-ray beam onto the sample is shaped by a 20 µm tungsten aperture (40 µm for coarse scans). For micro-mode imaging, a set of Kirkpatrick-Baez micro-focusing mirrors (K-B-mirror) focus the beam to a spot size of 5 x 5 µm. Macro- and micro-mode measurements were collected with incident beam energy set at 13.45 keV (pea and wheat) or 15 keV (canola) and sample positioned at 45° to reduce scattering signal. The XRF signal for XRF-imaging and µ-XAS was collected by a Vortex-ME4 silicon drift X-ray detector for macro-mode and Vortex-ME3 silicon drift X-ray detector for micro-mode (Hitachi High Technologies Science American, Inc., Chatsworth, CA).

For sample measurements, seed sections on Kapton® tape were affixed to the sample plate on the imaging stage with sample side facing the incident beam. A coarse scan was collected for the sample plate to acquire an overview of all samples (macro-mode: 10 msec dwell time and 0.2 mm pixel size; micro-mode: 25 msec dwell time and pixel size to 50 microns). Single point XRF spectra at different

TABLE 1 SR-μCT imaging specifications.

	Watermelon	Canola	Pea	Root
Image capture				
Detector	PCO Edge 5.5	PCO Edge 5.5	PCO Edge 5.5	PCO 4000
Objective	Optic Peter 2x	Optic Peter 10x	Optic Peter 5x	Hamamastsu AA60x
Effective magnification	1.8x	9x	4.5x	1x
Scintillator	LuAg: Ce 200 μm	LSO: Tb 10 μm	LuAg: Ce 50 μm	LuAg: Ce 200 μm
Pixel size	3.61 μm	0.72 μm	1.44 μm	9 μm
Sample to detector	0.05 m	0.03 m	0.05 m	0.4 m
White beam	1.000 aluminum filter (mm)	1.000 aluminum filter	0.800 mm Al + 0.100 silver filter (mm)	–
Beam energy	20 keV	20 keV	25.5 keV	41 keV
Mode	conventional	conventional	conventional	half-acquisition[1]
Projections	3000 images	3000 images	2500 images	3000 images
Flats	10 images	20 images	50 images	20 images
Darks	10 images	20 images	20 images	20 images
Reconstruction				
Phase	yes	yes	yes	yes
Delta/Beta	300	150	150	400
Spot removal	Yes	Yes	Yes	No
Prominence of spot	10000 counts	1000 counts	5000 counts	–
Spot blur	7 pixels	3 pixels	2 pixels	–
Ring removal	Yes	Yes	Yes	Yes
Fourier-transform	2D	2D	2D	2D
Sigma horizontal	60	60	13	90
Sigma vertical	2	2	2	2
Clipped histogram	No	Yes (16-bit)	Yes (16-bit)	No
Min value	–	4.56E-15	-5.42E-04	NA
Max value	–	1.06E-14	1.46E-04	NA
Detector cropped	yes	yes	no	yes
x (width)	2560	2560	2560	4000
y (height)	1000	1000	2160	888
Field of View				
FOV	2560x2560x1000 pixels	2560x2560x1000 (pixels)	2560x2560x2160 pixels	4000x4000x888 pixels
FOV	9241 x 9241 x 3610 μm	1843 x 1843 x 1555 μm	3816 x 3816 x 3100 μm	36000 x 36000 x 8000 μm

[1]Half-acquisition mode is available for larger samples that do not fit within the horizontal field of view. This mode increases the field of view to approximately double that of conventional mode by offsetting the sample to the right or left of the detector and collecting projections over 360˚. Images 180˚ apart are then stitched together to increase the horizontal field of view.

dwell times on the sample were collected for optimization of dwell time required to obtain adequate counts for the elements of interest and reduce the total time required to scan the samples. Using the coarse scan, regions of interest were selected and cued for high-resolution measurement (pea: macro-mode 20 μm pixel with 75 msec dwell time; wheat: micro-mode 5 μm pixel with 100 msec dwell time; canola: micro-mode 5 μm with 200 msec dwell time). For the parameters used to scan pea sections, each mm^2 of tissue

area took approximately 3.5 minutes; scan time can be calculated based on dwell time and pixel size.

2.7.2 XRF elemental imaging at SXRMB

XRF-imaging at the Microprobe end station of the SXRMB beamline (1.7 – 10 keV) (https://www.lightsource.ca/facilities/beamlines/cls/beamlines/bioxas-imaging.php#Techniques) at the Canadian Light Source, was used to investigate phytate localization in mature canola seed. SXRMB is uniquely suited to access biologically important elements from Si - Ca due to its optimization to the tender/soft X-ray energy range of 1.7 to 5 keV (Xiao et al., 2017). The X-ray light source is a bending magnet with primary optics consisting of a collimating mirror, double crystal monochromator and toroidal mirror. The microprobe end station slits and K-B-mirror further shape the incident monochromatic beam at 4 keV into a 10 x 10 µm spot. XRF signal was measured by a seven-element silicon drift detector.

For sample measurements, the canola seed section on Kapton® tape was affixed to a Cu sample plate with double sided carbon tape and placed within the vacuum chamber and pumped down. Note, that in this study, the section was originally collected on Kapton® tape, requiring it to be affixed to carbon tape for imaging. The sample holder was placed within the vacuum chamber of the beamline end station and pumped down. A coarse scan of all of the samples on the plate (10 µm beam size, 0.5 s dwell time and pixel size to 50 µm) was collected to show general elemental distribution and highlight areas of interest for high resolution scans. These parameters result in a scan time per 1 mm^2 of section of approximately 3.5 minutes. High resolution scans at selected areas were acquired with 10 µm resolution. The resulting XRF signal of P was calculated over the P K $\alpha 1$ range from 1952 - 2075.4 eV.

2.8 Experimental parameters for µ-XAS collection of XANES spectra (BioXAS imaging and SXRMB)

2.8.1 Select the spot(s) of interest for µ-XAS

The XRF elemental map collected (Section 2.7.1 for BioXAS-Imaging and Section 2.7.2 for SXRMB imaging) can be used for µ-XAS measurement to identify area(s) of interest and the exact coordinate(s) with respect to the sample stage. For the pea zinc XANES spectra, a coarse scan macro-mode XRF map was collected at BioXAS at 10 keV with 40 µm pixel size. Once a region of interest was selected, a µ-XAS measurement at the coordinates was setup to collect XANES data. For the pea seed Zn K-edge (9659 eV), XANES spectra were collected from 9558.6 – 9795.3 eV. For µ-XAS measurement at SXRMB of P in canola, regions of interest can be set based on fine scans and P K-edge XANES collected from 2115 – 2210 eV. Replicate XANES measurements are strongly recommended at the same spot and/or nearby spot(s) within the same tissue. In the case of pea seed, a minimum of 4 XANES spectra at neighbouring spots were collected for each area of interest. Scan times are dependent on the required dwell time, the number of energies collected and the number of spots measured.

2.9 Experimental parameters for Scanning transmission x-ray microscopy

The X-ray imaging and spectroscopy were performed using the spectromicroscopy (SM, 10ID-1) beamline (https://www.lightsource.ca//facilities/beamlines/cls/beamlines/sm.php#top) at the Canadian Light Source (Kaznatcheev et al., 2007), to investigate subcellular P distributions in mature canola seed embryonic cotyledons. The raw transmitted X-ray flux signals were converted to optical densities (OD) (i.e., absorbance) using an incident flux signal measured at regions devoid of sample material. P was mapped in canola sections using the P K-edge. Specifically, off-resonance (2143 eV) and on-resonance (2155 eV) OD images were collected for the same area at a spatial resolution of 50 nm, and the difference of the images (i.e., subtraction) resulted in the P map (Dynes et al., 2015) (See Supplementary Table 3). The P map was then overlaid with the off-resonance map. The off-resonance map represents the non-P components, mainly the organic components. The P K-edge spectra were obtained by collecting a sequence of images (i.e., stack) at specific energies for $AlPO_4$ and selected canola seeds. The energy scale was calibrated using $AlPO_4$ (2154.5 eV) (Tofoni et al., 2023). Note that when the data was collected, a new monochromator was being commissioned on the SM beamline, and the energy scale was not reproducible, so the energy scale of the canola seeds is not expected to be accurate. Analyses were performed with aXis2000 (http://unicorn.mcmaster.ca/aXis2000.html; Hitchcock, 2023).

2.10 Complementary methods

ESEM images were collected using a Thermo Quattro S SEM at the Electron Imaging and Microanalysis Lab at the Canadian Light Source. Backscattered electron images were collected at ESEM pressure of 150 Pa with beam energy of 5 keV.

Oil and protein contents of NAM13 and NAM45 seeds were determined using NIR analysis (NIRS system model 6500, FOSS NIRSystems, Silver Springs, MD).

Phytate content of oil-free seed residue was determined by Phytate/Phosphorus assay (Megazyme K-PHYT, Megazyme, Wicklow, Ireland).

To remove oil from canola seeds, seeds were split in half using a razor blade. One half of the seed was soaked in 10 mL reagent grade mixed hexanes in a glass vial for 1 hour. The seed was removed, air dried, then placed in a clean vial.

2.11 Data analysis

2.11.1 SR-µCT image processing (BMIT-BM)

Image sequences were reconstructed and vertically stitched using ufo-kit tofu software (GitHub - ufo-kit/tofu: Helper scripts for tomographic reconstruction using the ufo-core framework (https://github.com/ufo-kit/tofu) (Faragó et al., 2022). During reconstruction of the projected images, flat field correction (FFC)

was performed using the acquired flat and dark signal images. Paganin TIE (transport-of-intensity equation) phase retrieval, a non-interferometric non-iterative method of deterministically solving the TIE using the intensity distribution at the in-focus plane and the axial intensity derivative (Zuo et al., 2020) was enabled. This increases the signal to noise ratio and the image contrast, especially with respect to water and plant tissue within the sample. Large spots were removed using predetermined thresholds. A Fourier-transform based filter was used to remove ring artifacts resulting from sample rotation. Vertical stitching was performed based on the image row overlap between adjacent vertical scans. The resulting 32-bit image sequence was clipped and converted to 16-bit where required. The images were then cropped as required to minimize background information using ImageJ software (https://imagej.net/ij/) (Schneider et al., 2012). The time required for image processing is dependent on the available computational resources but each sample can take a minimum of 1-2 hours to reconstruct and stitch with a high performance computer (current computer specifications comprised a Precision 7820 Tower with two Intel Xeon Silver 4215R processors, two RTX A8000, 48GB graphics cards and 512 GB RAM). Following the reconstruction, the target image sequence for segmentation with Avizo should be less than 20 GB. Image sequences were loaded into Avizo software (Thermo Fisher Scientific, Waltham, MA) for segmentation. Every ~40th slice was segmented and a smart interpolation of the image sequence segmentation was performed using Biomedisa (https://biomedisa.info/) (Lösel et al., 2020). Biomedisa results were returned to Avizo for the final segmentation, volume rendering and analysis. Reconstruction parameters for all samples are shown in Table 1. For additional methodology descriptions see Willick et al., 2020.

2.11.2 MidIR data processing (MidIR beamline)

The data output file was loaded into Quasar software (https://quasar.codes/) (Toplak et al., 2017, Toplak et al., 2021). The spectra were pre-processed to correct for atmospheric gas (CO_2/H_2O) using default parameters. A rubber band baseline correction was performed. For visualization of peak areas, baseline integrals from peaks of interest were calculated; specifically, 2819 - 2994 cm^{-1}, (lipid-rich, CH_2 and CH_3), 1700 - 1770 cm^{-1} (C=O stretch; lipid), 1600 - 1700 cm^{-1} (amide I), 1490 - 1580 cm^{-1} (amide II), 1200 - 1490 cm^{-1} (structural carbohydrates), and 900 - 1200 cm^{-1} (carbohydrate). Empty pixels were removed from the spectra based on thresholds in the fingerprint region (1800 cm^{-1} - 900 cm^{-1}). For the PCA and K-means cluster analysis shown, data was further processed by removing empty pixels through thresholding, limited to the fingerprint region (1800 cm^{-1} - 900 cm^{-1}), baseline corrected and vector normalized. All data was exported from Quasar for subsequent visualization in R (ggplot2) or Python (Matplotlib).

2.11.3 XRF data processing (BioXAS-imaging and SXRMB)

For XRF maps (generated at BioXAS-Imaging), the data file was loaded into PyMca software (PyMca 5.9.2 — https://www.silx.org/doc/PyMca/dev/index.html; Solé et al., 2007). An energy calibration was generated using the Fe K α_1 and Ca K α_1 peaks. A fit configuration file containing e.g. fit, background, beam and elements peak parameters was setup and used for batch fitting of the XRF map. The RGB visualization was generated within PyMca RBG correlator's matplotlib visualization tool. For the SXRMB XRF maps, the elemental maps were generated from the resulting XRF spectrum based on appropriate energy windows (P K α_1 1952 - 2075.4 eV, Ca K α_1 3608.14 - 3775.22 eV, Cd L α_1 3056.76 – 3210.7 eV) normalized with incident I0.

2.11.4 XANES data processing (BioXAS-imaging and SXRMB)

XANES data were processed using Athena software (https://bruceravel.github.io/demeter/aug/index.html ; Ravel and Newville, 2005). X-ray absorption spectra were calculated by dividing raw fluorescence with incident I0 signal. Pre-edge and post-edge parameters were adjusted to determine the pre-edge and normalization lines respectively. Normalized XANES spectra were calculated by subtraction of the pre-edge line, followed by division by the normalization constant determined by the pre-edge and normalization line. For visualization, data was exported and plotted using Origin Pro Software (OriginLab, Northampton, MA).

3 Results

To support the adoption of synchrotron-informed research in plant sciences, we present methods for "preparing an array of crop seeds for analysis by a wide range of electromagnetic radiation energies. An overview of method pipelines is illustrated in Figure 1, starting with input options for seeds to" accommodate mature, dry materials or developing, fresh seeds (dark green panels). Sample inputs for SR-µCT imaging (at the BMIT-BM beamline) can accommodate whole seeds without the need for sectioning, and benefits from the preparation of dry materials (Figure 1 yellow panels). Preparation of seed materials for all other beamlines described here involved spatially-resolving chemical compositional data and thus, used hydrated, frozen materials that were cryo-sectioned (and not chemically fixed; Figure 1 light green boxes). Note that XRF imaging on hydrated, non-sectioned, thin samples is a viable option depending on the element of interest if complementary multimodal imaging is not required. Utilizing the range of energies available at the CLS, from hard X-rays (5 - 40 keV; Figure 1 grey boxes), to soft X-rays (130 eV - 10 keV; Figure 1 blue boxes), and IR light (with or without SR) provides diverse opportunities for making biological discoveries. We focus on the structural information provided by SR-µCT (BMIT-BM) for revealing internal seed anatomy, and the chemical mapping tools for localizing chemistries across seed subcompartments, provided by hard X-ray synchrotron imaging technologies (BioXAS-Imaging beamline) and soft X-ray modalities (SXRMB and SM beamlines), as well as MidIR spectroscopy. Potential data outputs from each of these imaging techniques are summarized by beamline (Figure 1 white boxes), including the category of molecules that can be probed and the spatial resolutions that can be achieved. Data presented here features case study examples, using seeds from

Citrullus sp. (watermelon), *Brassica* sp. (canola), *Pisum sativum* (pea), and *Triticum durum* (wheat) seeds (see summary lists at the bottom of Figure 1, Figures 2-8, and Supplementary Table 4 for a list of Figures, samples and methods), and showcase the combined power of cross-beamline synchrotron imaging for deciphering complex living systems such as seeds.

Mature seeds are an ideal material for SR-µCT as they are easy to manipulate, require no fixation or pre-processing (Figure 1) and are dry at maturity, making them subject to minimal motion artifacts during scans. Further, SR-µCT imaging provides non-destructive interrogation of seed traits, after which seeds can remain viable. Case study 1 demonstrates the use of SR-µCT to image watermelon seeds (Figures 2A–D). Differential absorption of X-rays relates directly to density, and allowed details of seed anatomy and necrosis due to invading organisms to be distinguished. Pan-through videos of sequential SR-µCT slices through a healthy and a fruit blotch infected seed illustrate this point (Supplementary Video 1). Volume renderings following image segmentation isolated internal 3D features of interest including necrotic tissue (Figure 2A, blue) resulting from *Acidovorax citrulli*-induced bacterial fruit blotch, and seed vasculature (Figure 2B, red). Feature overlays (Figure 2C) allowed direct comparison of vasculature and necrotic networks within the seed. The volume rendering was critical to inform the degree of necrosis and the complexity of the vascular network. This was emphasized by a pan-through video of the segmented seed (Supplementary Video 2) differentiating the collective network of necrotic tissue from the vasculature and surrounding seed tissue in 3D space (Supplementary Video 3). An example of an SR-µCT slice used in image segmentation for 3D rendering is provided (Figure 2D).

Computational analysis of SR-µCT data facilitates accurate developmental staging and quantification of internal features following volume rendering as discussed in case study 2 and as shown in Figures 2E–I. Individual segmentation of immature pea seed subcompartments, revealed the embryo (Figure 2E, F) endosperm cavity (Figure 2G), and seed coat (Figure 2G) from SR-µCT reconstructed slices (Figure 2H). Volumetric analysis showed that the seed coat accounts for 64%, the endosperm cavity for 35% and the embryo for ~1% of the total volume of a seed (Figure 2I) at 7 days post flowering. In addition to this quantitative volumetric analysis, precise developmental staging information was readily accessible from 3D reconstructed data, which showed the presence of an early heart embryo. A single slice through a seed viewed by light or transmission electron microscopy (comparable to Figure 2H), in this case, could lead to inaccurate classification of the embryo as globular. Large and mature seeds with less permeable seed coats are susceptible to fixative penetration issues (Supplementary Methods). If volumetric quantification is not required, seeds can be cut in half and imaged (Supplementary Figure 1). This partial seed scan still gives important

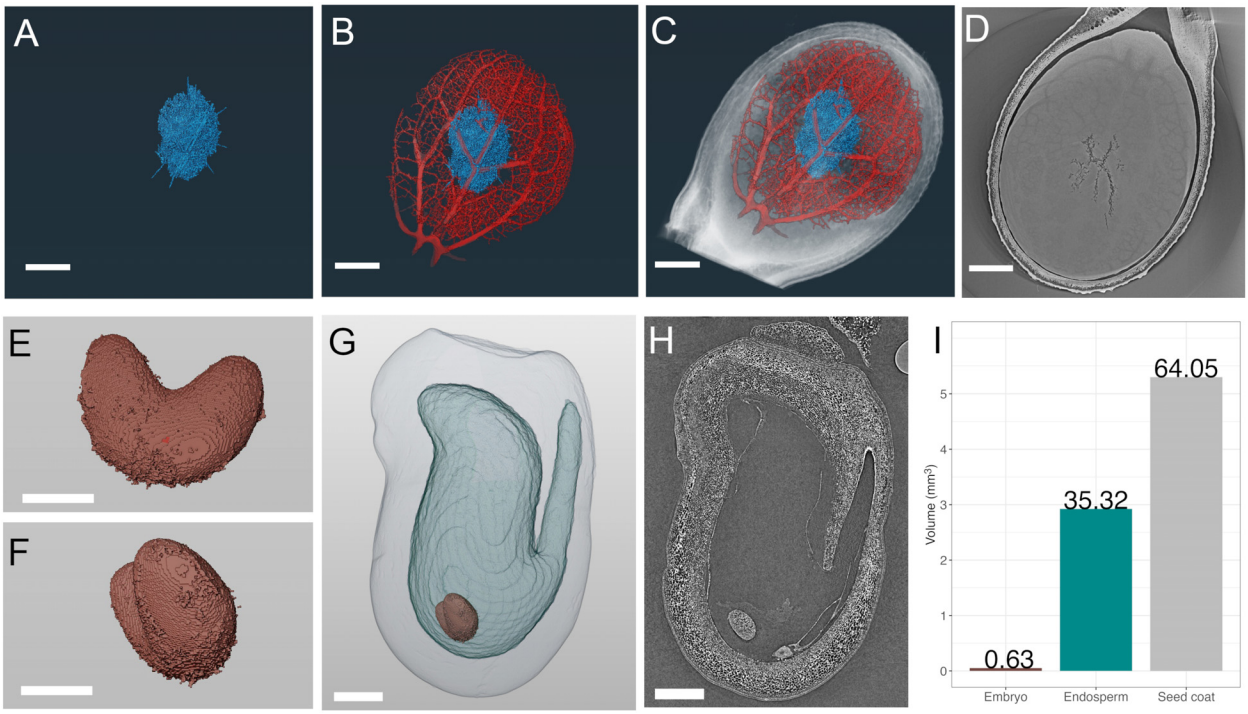


FIGURE 2
SR-µCT of mature watermelon (*Citrullus lanatus*) and immature pea (*Pisum sativum*) seeds produced at the Canadian Light Source BMIT-BM beamline. SR-µCT from a mature watermelon seed infected with *Acidovorax citrulli* **(A-D)** and a developing pea seed 7 days post flowering **(E-H)**. **(A)** Segmented and volume rendered necrotic path of bacterial fruit blotch infection alone **(A)**, in blue) and overlaid with volume rendered seed vasculature **(B)**, red), and whole watermelon seed **(C, D)** SR-µCT slice through infected watermelon seed, **(E, F)**. Segmented and volume rendered embryo from an early stage pea seed. **(G)** Segmented and volume rendered whole pea seed of seed coat (outermost light grey), endosperm cavity (light blue) and embryo (red) subcompartments. **(H)** SR-µCT slice through a pea seed. **(I)** Percent volume of each subcompartment of the pea seed in **(G)**. Scale bar **(A–D)**, 1000 µm; scale bar **(E, F)**, 250 µm; scale bar **(G, H)**, 500 µm.

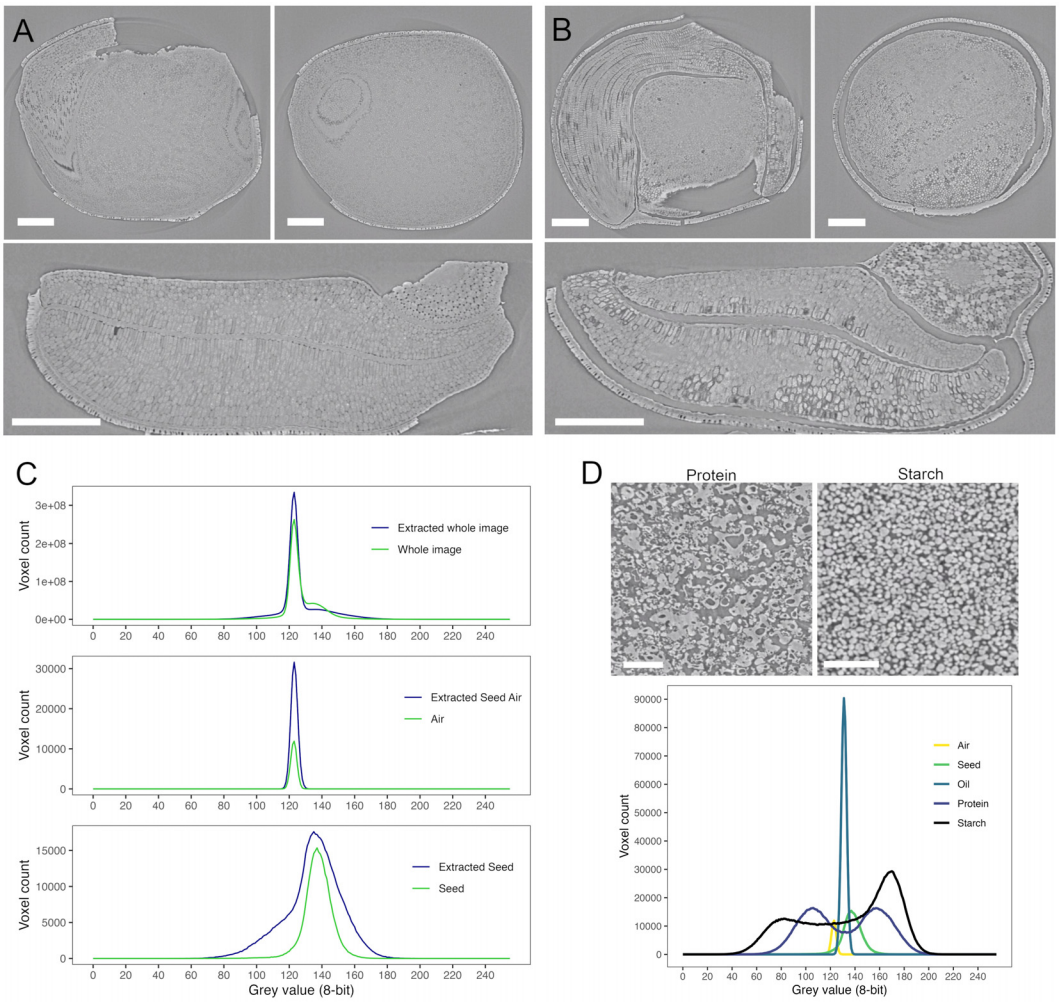


FIGURE 3
SR-μCT of a mature canola (*Brassica napus*) seed, produced at the Canadian Light Source BMIT-BM beamline. An SR-μCT slice is shown before **(A)** and after **(B)** extraction with hexane. Histograms of the 8-bit grey value for the seed, air and total image are shown in panel **(C)** SR-μCT of isolated protein and starch (panel **(D)** show distinct morphological differences as well as distinct grey values as seen in the histogram below. Scale bar **(A, B)** 200 μm; scale bar **(D)**, 100 μm.

3D architectural and developmental information and is useful for correlation with data from matched samples at different beamlines. When the requirement for imaging of plant materials exceeds the size capacity of the BMIT-BM beamline, the higher energies available at the BMIT-ID beamline can be applied. For reference, the complex subterranean root architecture of wheat is shown in Supplementary Figure 2. This highlights the capacity for BMIT-ID SR-μCT to isolate living plant features within a complex static matrix such as soil, and demonstrates the ability to investigate larger scale structures. The higher energy BMIT-ID beamline also has the ability to image significantly more dense materials than the BMIT-BM beamline.

SR-μCT is most often used for structural imaging. However, it is possible to glean some compositional information for biological materials using phase retrieval. These scans generate contrast based on differences in attenuation coefficients within the material which relate to the density of the material and the atomic weight of its elements. The application of Paganin propagation-based phase retrieval can increase the contrast in the generated images using differences in refractive indices. This is especially critical in biological samples where materials have similar absorption profiles and contain low atomic number elements. Grey value distribution can be used to differentiate between materials and serve to provide compositional insight. Histograms over single SR-μCT slices through a mature canola seed without extraction (Figure 3A) and after hexane extraction of lipids (Figure 3B) showed differences across the complete image (Figure 3C, top panel) where extraction with hexane sharpened the histogram peak. Hexane extraction increased the air component in the sample (Figure 3C, middle panel) and produced differences in the seed histogram (Figure 3C, bottom panel). Following extraction, the histogram of the seed component widened to include a larger number of grey values, indicating increased heterogeneity of particle size and/or pore distribution, and the peak of the seed was shifted to the right from that of the air (Figure 3C, middle

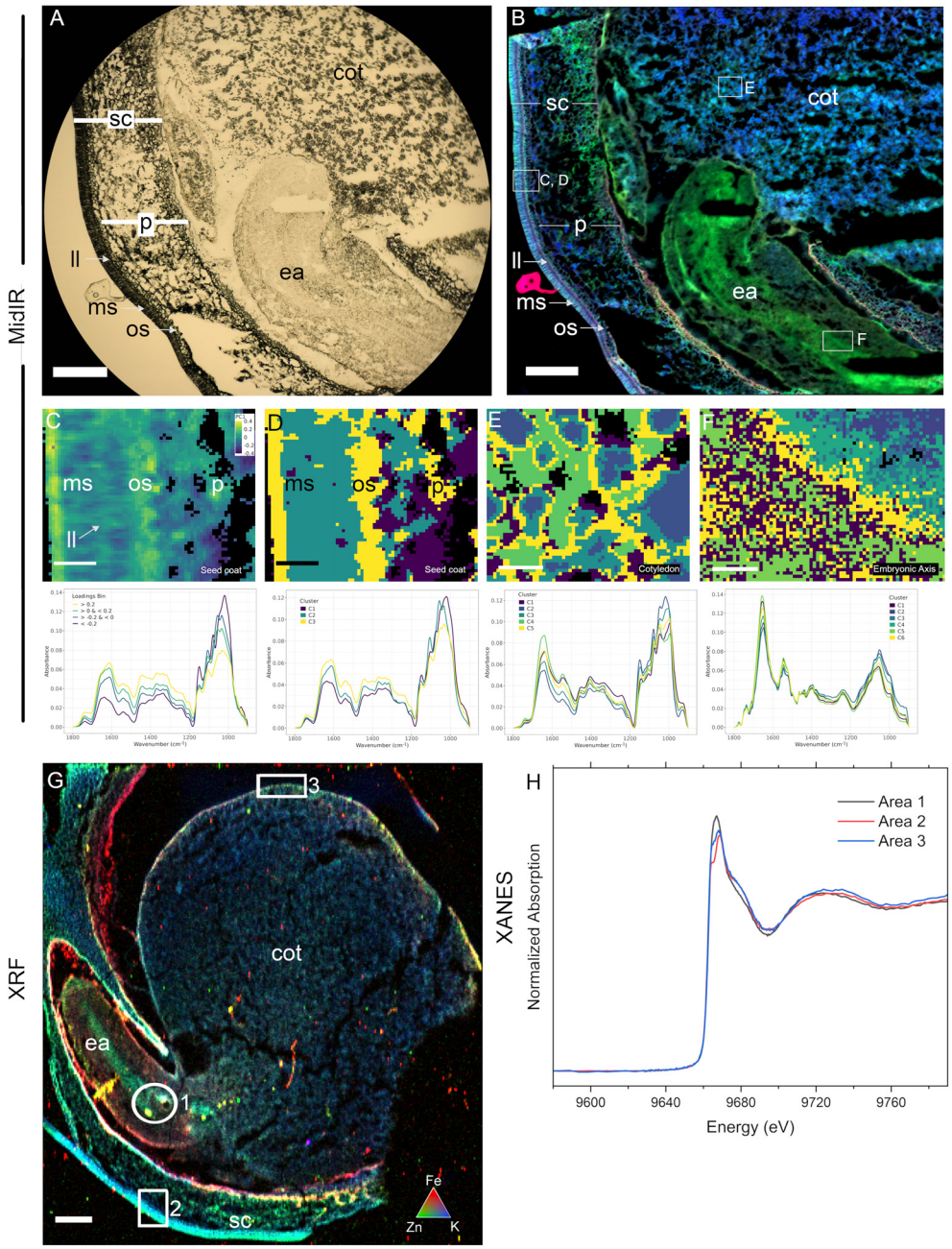


FIGURE 4
Cross-beamline semi-correlative analysis of nutrient distributions in a single immature pea (*Pisum sativum*) seed, produced at the Canadian Light Source MidIR spectroscopy and BioXAS-Imaging beamlines. Brightfield/transmitted light view of a cryosectioned pea seed showing the visual region scanned **(A)** and the resulting reconstructed MidIR image showing lipid (red), protein (green) and carbohydrate (blue) distributions **(B)**. Selected regions in **(B)** including in the seed coat **(C, D)**, cotyledon **(E)**, and embryonic axis **(F)** were assessed using principal component analysis **(C)** and K-means clustering **(D-F)**. An adjacent section was used for XRF imaging and an overlay of iron (red), zinc (green) and potassium (blue) distributions are shown **(G)**. Zinc speciation differences were investigated in thee selected regions in G using XANES **(H)**; areas 1, 2, and 3). Scale bar **(A, B, G)**, 500 μm; scale bar **(C–F)**, 50 μm. sc, seed coat; p, parenchyma; ll, light line; ms, macrosclereids; os, osteosclereids; ea, embryonic axis; cot, cotyledon.

panel) or complete seed (Figure 3C, top panel) (peak grey value of 135 versus 123). The increased greyscale range of the extracted seed relates primarily to the increased air volume within the sample. The sample to detector distance and the δ/β ratio (where δ is the phase shift and β is the absorption component of the refractive index) used for phase retrieval during image reconstruction impacts the image quality and the size of the fringe effect and can therefore affect the distribution of relative greyscale values. SR-μCT scans of isolated protein (Figure 3D, upper left), starch (Figure 3D, upper right) and oil showed distinct morphologies in packing density and particle size. For example, starch has smaller air spaces between granules than protein and this size difference resulted in a shift of approximately 20 grey values for the air peak between samples (Figure 3D, lower). These peaks were also shifted from that of bulk

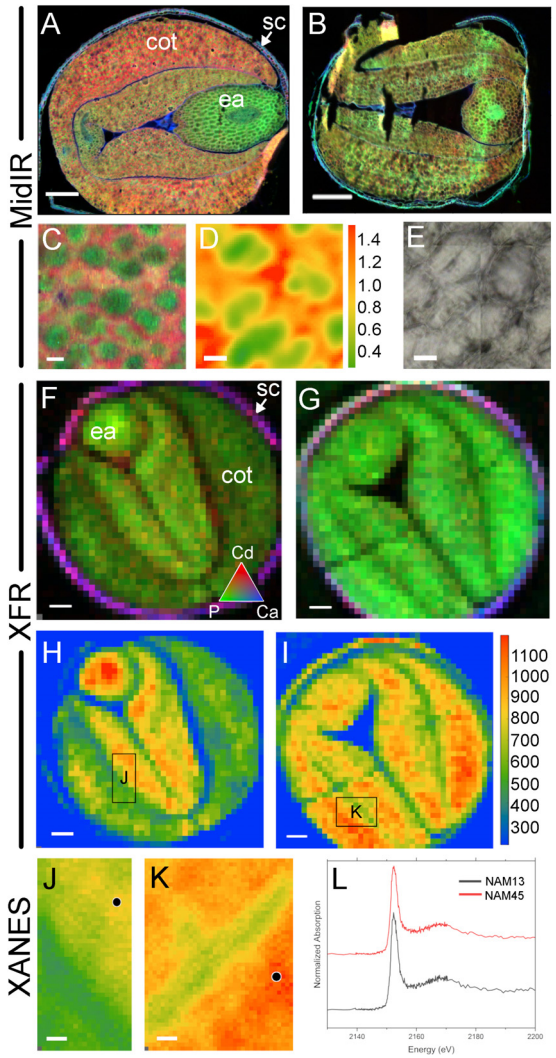


FIGURE 5
Cross-beamline analysis of nutrient distributions in high and low phytate mature canola (*Brassica napus*) seed, produced at the Canadian Light Source MidIR spectroscopy and SXRMB beamlines. Cryosectioned mature seed in a high level phytate genotype, NAM13 **(A, C-F, H, J)** and a low level phytate genotype, NAM45, **(B, G, I, K)**. MidIR reconstructed images of canola seed lipid (red), protein (green) and carbohydrate (blue) macronutrient distributions in NAM13 **(A)** and NAM45 **(B)**. High magnification images shows the tissue distribution of protein bodies (green) surrounded by lipid (red) with conventional IR source **(C)** and the lipid to protein rations with a synchrotron IR source **(D)** and corresponding visual image **(E)**. These same seeds were used to generate XRF images at the SXRMB to reveal cadmium (red), phosphate (green) and calcium (blue) distributions **(F, G)**, and the specific distribution of phosphate **(H, I)**. Regions of interest in the high phytate **(J)** and low phytate **(K)** genotype were scanned at higher resolution and black dots indicate specific regions for subsequent XANES **(L)**. Scale bar **(A, B, F-I)**, 200 µm; scale bar **(C-E)**, 10 µm.

air. This demonstrated that air can manifest greater than 20% of the dynamic grey value CT range and irrespective of material, the histogram can inform particle size and amount. The use of hexane extraction resolved the localization of oil versus protein given the histogram peaks overlap. Using these relative greyscale values it is possible, over selected regions within a seed, to gather information

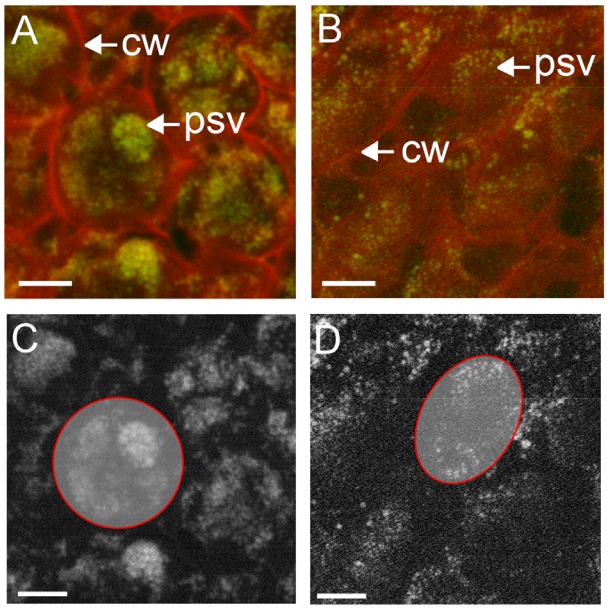


FIGURE 6
Subcellular distribution of phosphorus in cotyledon cells of high phytate **(A, C)** and low phytate **(B, D)** genotypes, measure with STXM at SM. Phosphorus (green) was mapped using the P K-edge, with on-resonance (2155 eV) and off-resonance (2143 eV) and shown against off-resonance optical density (red) **(A, B)**. Panels **(C, D)** are greyscale representations of phosphorus in **(A, B)**, respectively, with individual cells traced in red for semi-quantitative analysis of phosphorus content. Scale bar 10 µm. cw, cell wall; psv, protein storage vacuole.

regarding localized chemical composition based on phase shifts and comparison of the resulting grey values. An example of this microstructural characterization approach is described by Sivakumar et al., 2024. It should be noted that these changes in grey values are relative changes based on material density, not based on calibration to known standards. Transformation to standard values, as often seen in medical imaging (Hounsfield units), is not practical as air is not a single defined value.

While SR-µCT can support broad classification of chemical components in a sample and their localization patterns, MidIR spectroscopy and XRF imaging are more useful for interrogating macromolecule and element distributions in seeds (case study 2), including through the differentiation steps that form plant embryos. Processing of developing seeds for MidIR and XRF imaging (Figure 1) is more involved than for mature seeds. Figure 4A shows a visual of the section quality of a 10 µm cryosection for MidIR. The structure and general morphology are intact however, freezing and cryosectioning can create holes in the tissue. The resulting MidIR image (Figure 4B) showing lipid (red), protein (green) and carbohydrate (blue) demonstrates the minimal impacts of these sample preparation artifacts on data quality. In the case of pea, developing seeds (13 days post flowering) presented elevated levels of protein in the embryonic axis relative to the carbohydrate-rich cotyledons. The seed coat contained high carbohydrate levels, with protein enriched sub-domains (e.g. at the base of the sclereid layers) and lipid enrichment (particularly apparent in the light line region of the macrosclereid layer).

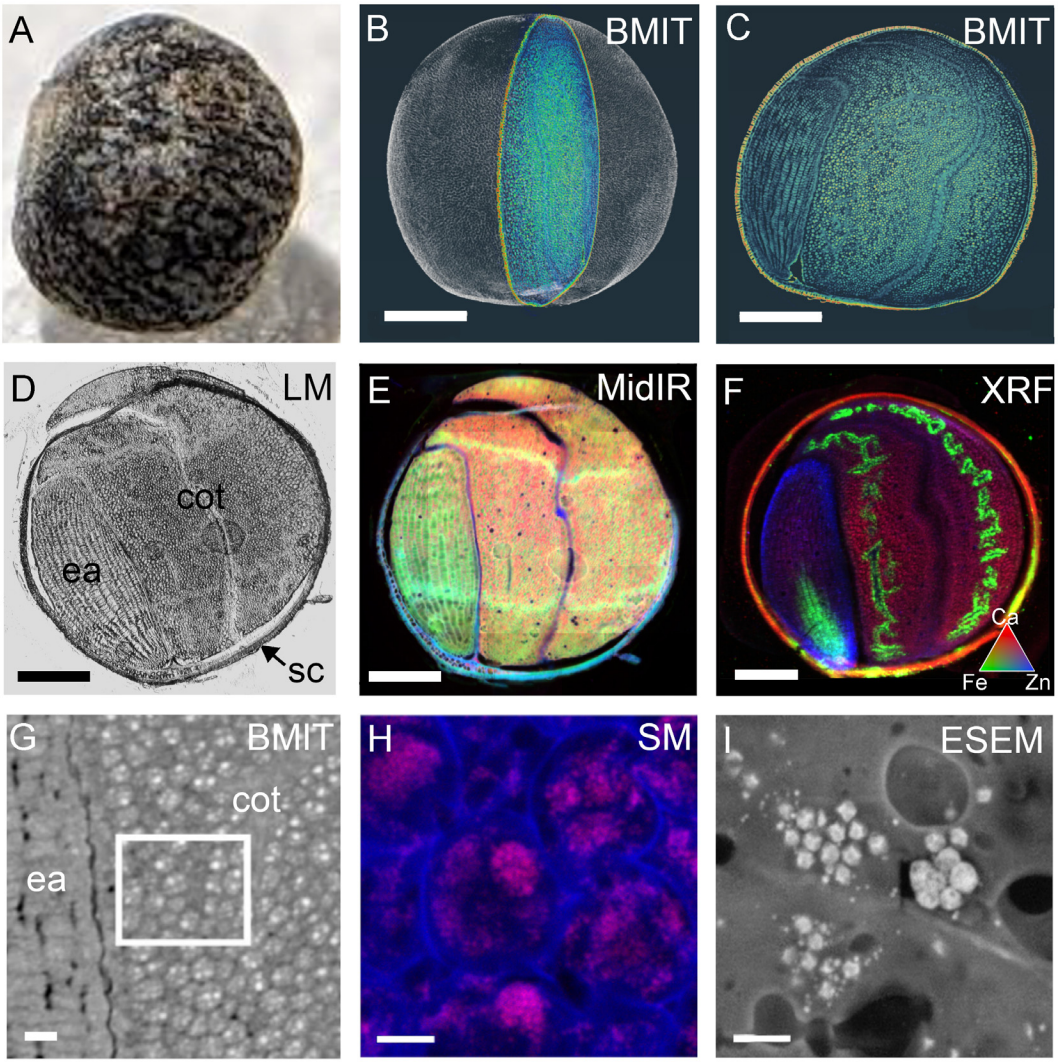


FIGURE 7
Cross-beamline correlative analysis of structure and nutrient distributions within a single mature canola (*Brassica napus*) seed, at the Canadian Light Source MidIR spectroscopy, BioXAS-Imaging, SM and Electron Imaging and Microanalysis Lab or EIML beamlines. A photograph of the mature, intact canola seed **(A)** and the 3D rendering following SR-µCT with a single section highlighted **(B)**. The SR-µCT single slice view **(C)** of the highlighted section in **(B)** corresponds to the same slice imaged by light microscopy **(D)**, MidIR spectroscopy **(E)**, and XRF mapping **(F)** following cryosectioning. MidIR spectroscopy **(E)** shows the distribution of lipid (red), protein (green) and carbohydrate (blue) and the XRF mapping **(F)** shows calcium (red), iron (green) and zinc (blue) distributions. A higher resolution slice showing cellular detail by SR-µCT **(G)**, white box) can be used to inform the locale for phosphorus mapping by SM **(H)**. Environmental scanning electron micrograph (ESEM) **(I)** shows protein storage vacuoles and globoids. Scale bar **(B, F)**, 500 µm; scale bar **(G)**, 50 µm; scale bar **(H)** 10 µm; scale bar **(I)**, 2 µm. sc, seed coat; ea, embryonic axis; cot, cotyledon.

Generalized overviews of macromolecule distribution provided by MidIR reconstructed images, can be analyzed in regions of interest by PCA (Figures 4C, B) and K-means clustering (Figures 4D, E) to identify compositional clustering features within the seed coat (Figures 4C, D), cotyledon (Figure 4E) and embryonic axis (Figure 4F). In the seed coat PCA (Figure 4C), principal component 1 (PC1), explained 47% of the variation, and showed distinct macromolecule compositions. Binning of the loadings according to the legend breaks followed by extraction and averaging of the spectra values allowed interrogation of the spectra to determine which macromolecules were the primary components at distinct cellular locations. The results of this binning method were in accordance with that of K-means clustering of the fingerprint regions from 900 - 1800 cm^{-1} (Figure 4D) where the cuticle and inner layer of the seed coat sclereids (yellow) grouped together in cluster 3. This cluster is consistent with the higher protein content regions shown in the false-colour RGB image (Figure 4B) as the average spectra in the cluster analysis (lower panel) showed well defined amide I (1600 - 1700 cm^{-1}) and amide II (1490 - 1580 cm^{-1}) peaks. Cluster 3 represented elevated structural carbohydrates (1200 - 1490 cm^{-1}) and the presence of a lipid component (peak at ~1730 - 1745 cm^{-1}) in the predicted macrosclereid light line. Cluster 2 represented a carbohydrate-rich area (see blue in panel B) and the average spectra showed a strong carbohydrate peak (900 - 1200 cm^{-1}) with specific peaks likely representing structural carbohydrates such as cellulose (~1110 cm^{-1}) and xyloglucans (~1070 cm^{-1}). K-means clustering in the cotyledon (Figure 4E) showed distinct regions of

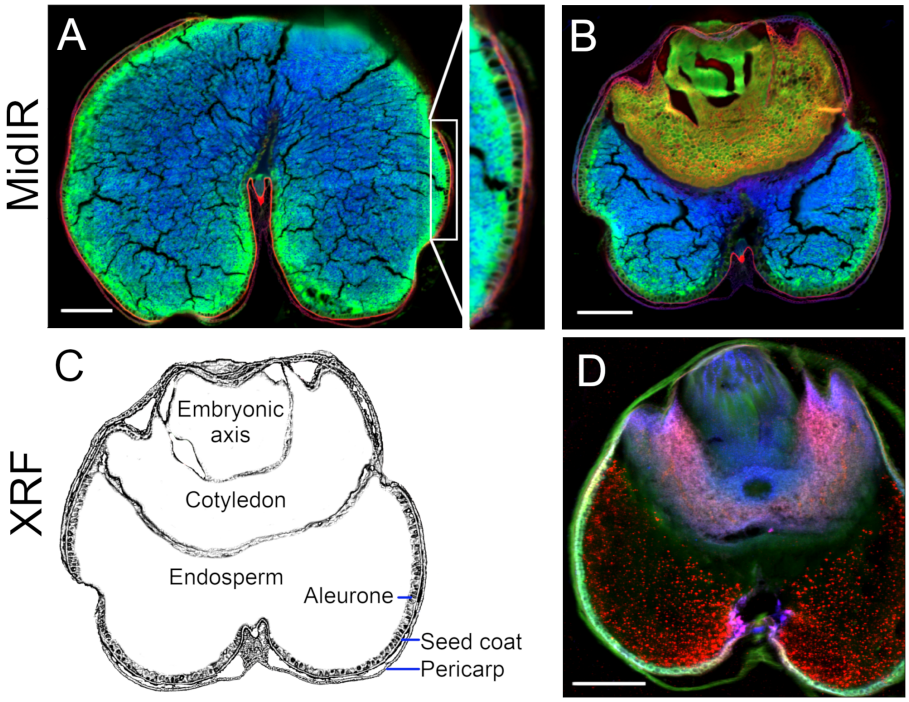


FIGURE 8
Cross-beamline analysis of nutrient distributions in mature wheat (*Triticum durum*) grain, produced at the Canadian Light Source MidIR spectroscopy and BioXAS-Imaging beamlines. MidIR spectroscopy **(A, B)** shows lipid (red), protein (green) and carbohydrate-enriched (blue) distribution patterns in mature wheat seeds in the endosperm **(A)** and embryo **(B)** regions of the seed. A simplified rendering of a wheat grain section with labels is provided for reference **(C)**. XRF imaging shows the distribution of calcium (red), potassium (green) and manganese (blue) through the embryo, endosperm and seed coat **(D)**. Scale bar 500 μm.

macromolecule distribution. Vacuole-like structures (blue, cluster 2 inner and cluster 3 outer) were tightly surrounded by a third cluster (yellow, cluster 5). Interrogation of the average spectra across these clusters (Figure 4E) supports carbohydrate-rich (900 - 1200 cm^{-1}) vacuole-like structures surrounded by high protein regions (cluster 4 and cluster 5) as seen by amide I (1600 - 1700 cm^{-1}) and amide II (1490 - 1580 cm^{-1}) peaks. The embryonic axis is a protein-rich region according to the false colour RGB image (Figure 4B) and K-means cluster analysis corroborates this observation. There are small differences across the region that can be differentiated (Figure 4F). However, average spectra (Figure 4F) of these clusters show minimal differences and mainly serve to highlight the pronounced amide I (1600 - 1700 cm^{-1}) and amide II (1490 - 1580 cm^{-1}) peaks in each of the clusters. There are subtle changes in the carbohydrate composition (900 - 1200 cm^{-1}) that contribute to cluster assignment.

In addition to macromolecule data for MidIR imaging, multi element data can be obtained from XRF spectroscopy of adjacent sections. A false-coloured reconstructed image depicts Fe (red), Zn (green) and K (blue) distributions in an immature pea seed (Figure 4G). Fe was evident in the remnants of endosperm adjacent to the seed coat, while K was more abundant in the seed coat sclereids; a distribution overlapping with that of clusters 2 and 3 from the MidIR analysis (Figure 4D). The highest levels of Zn were seen in the inner and outer layers of the seed coat, as well as the central region of the embryonic axis and outer edge of the cotyledons. The distribution of Zn in the embryonic axis overlapped with the region of the greatest protein content, as seen in the MidIR image (Figure 4B). Regions with high levels of Zn were selected (1-3) for XANES (Figure 4G) and results indicated the presence of different species of Zn across these regions (Figure 4H). Calibration standards were not available so conclusive identification of the species requires future investigation.

Case study 3 involved the spatial localization of P in high phytate (NAM13) and low phytate (NAM45) canola lines using multi-beamline SR imaging. Additional information regarding these genotypes is available in Supplementary Table 5 and in Brown et al., 2024. Phytate analysis demonstrated that NAM13 and NAM45 contain 4.35% and 0.19% phytate, respectively (Table 2). These seeds differed in total oil and protein content with the high phytate NAM13 having 13.6% higher oil and 9.1% lower protein than the low phytate NAM45. Using MidIR spectroscopy, the spatial distributions of oil (lipid) and protein across canola seed subcompartments were interrogated. Comparison of RGB (R, lipid; G, protein; B, carbohydrate) images of the high (Figure 5A) and low (Figure 5B) phytate canola genotypes corroborated the whole seed NIR results. Additionally, MidIR images showed that oil was localized in the cotyledons of the high phytate line with the highest levels in the outer lobes while the low phytate genotype had a more homogeneous distribution of oil. Protein was present in the cotyledons of the high phytate genotype but the highest levels were within the embryonic axis. The low phytate genotype had a relatively homogeneous distribution of protein. Carbohydrate was predominantly seen in the seed coat

10.3389/fpls.2024.1395952

TABLE 2 Canola composition analysis.

	% Oil	% Protein	% Phytate
NAM13	45.83	23.69	4.35 ± 0.30
NAM45	31.53	33.26	0.19 ± 0.02

and lining the outer epidermal edge of the embryo in both genotypes. Since lipid and protein in the cotyledons appeared to differ in their predominant distribution patterns, a high magnification scan of this region (Figure 5C) in the high phytate genotype was performed. This showed protein concentrated in distinct regions of the cell, in a lipid-rich matrix. This was supported by SIR and calculation of the lipid to protein ratio (Figure 5D; visible image in 5E). Within cells, individual protein bodies (green) surrounded by lipid-rich areas (red) could be discerned. Calculation of component ratios is useful for cross-sample comparison as it accounts for sample section thickness variation; this comparison is not as reliable for pure peak integration visualization as presented in the RGB images.

To examine elemental distributions in canola seed with varied phytate content, XRF reconstructed images from adjacent sections of the same seeds were collected. XRF mapping at SXRMB highlighted elemental differences between the NAM13 and NAM45 genotypes (Figures 5F, G). RGB images of Cd (red), P (green) and Ca (blue) showed primary distributions of Cd and Ca in the seed coat in high (Figure 5F) and low (Figure 5G) phytate genotypes. P was distributed throughout the seed tissues in both genotypes however, in the low phytate genotype the distribution appeared consistently homogeneous. This was better illustrated by XRF images showing P distribution without other elements (Figures 5H, I). The high phytate genotype (Figure 5H) showed high levels of P in the embryonic axis, with mid-range levels in the inner lobes of the cotyledon, and low levels of P in the outer cotyledon lobes. In contrast, the low phytate genotype (Figure 5I) showed high levels of P throughout the seed. These MidIR and XRF images showed co-localization of P and protein within the seed.

Given the phytate status of the canola NAM genotypes, an important consideration was to identify P independently of protein, to ultimately identify P in the form of phytate. P K-edge XANES of select regions of the high phytate (Figure 5J) and low phytate (Figure 5K) genotypes confirmed the presence of P, although no speciation difference was detected (Figure 5L). A XANES measurement of the P signal was also important for subsequent spectromicroscopy (SM), to confirm the presence of P, to determine the energy to be used for on/off-resonance measurements, and to minimize beam time.

XRF at the SM beamline reveals element distributions with nano-scale resolution. Mapping of P within the cotyledons of canola (Topas; Figures 6A, C) showed the presence of punctate deposits of P within defined structures, consistent with globoids within protein storage vacuoles. These deposits were also identifiable in the NAM45 low phytate genotype, however, the distribution was more diffuse and indicative of fewer crystallized phytate deposits (Figures 6B, D). Semi-quantitative measures of P content may be obtained by counting the pixels above background that represent the P signal, within defined regions of interest, in this case whole cells (red circles). When normalized for cell size, Topas (Figure 6C) had 5648 counts/1000 pixels, while the low phytate line (Figure 6D) had 4521 counts/1000 pixels. These data support a stronger, localized P signal within subcellular domains of Topas canola cotyledon cells (Figure 6A) than the P in a low phytate canola genotype (Figure 6B). Thus, the distribution of P within the putative protein storage vacuoles made possible by the high-resolution mapping of elements at the SM beamline is critical for the interrogation of phytate deposition within these genotypes.

The power of cross-beam multi-modal imaging for correlating structural features with compositional traits is demonstrated with the use of the same canola seed (Figure 7A) for imaging at multiple beamlines. A false-coloured slice through an SR-μCT image (Figure 7C) can be identified from a 3D rendering of a whole canola seed (Figure 7B; Supplementary Video 4). After freezing and cryo-sectioning of the SR-μCT imaged seed, the same slice can be directly compared to those imaged by light microscopy (Figure 7D), MidIR (Figure 7E) and XRF (Figure 7F) spectroscopy. SR-μCT images show detailed morphology of the cellular composition of the seed (as shown in the pan through in Supplementary Video 5), which can be critical for the interpretation of reconstructed chemical images with lower resolution or limited compositional differences. When imaged with MidIR, the slice of interest contained an embryonic axis rich in protein (green), cotyledons rich in lipids (red) and a seed coat rich in carbohydrate (blue; Figure 7E). XRF spectroscopy in micro-mode (5 μm resolution) showed the distribution of Ca (red), Fe (green) and Zn (blue) within an adjacent section of the same seed (Figure 7F). This demonstrates that the embryonic axis contained high levels of Zn while the seed coat and the cotyledons were rich in Ca. Iron in this canola seed had a discrete denticulate localization within the cotyledons when observed by XRF spectroscopy, which could be precisely mapped to the vasculature region of the cotyledons, and a central domain in the radicle using SR-μCT datasets. With these complimentary SR approaches, observations made in one dataset could be informed by another.

These synchrotron chemical imaging techniques proved to be valuable for identifying the overt differences between dicot and monocot species. MidIR scans of a mature wheat seed (or grain) (Figure 8) in a region without (Figure 8A) and with (Figure 8B) the embryo, followed by a diagrammatic rendering with noteworthy anatomic features labelled (Figure 8C), highlight that the endosperm was predominantly carbohydrate rich, the cotyledon (also called scutellum) was protein- and lipid-rich, and the embryonic axis was primarily comprised of protein. The pericarp was carbohydrate-rich but also had a prevalent localized lipid component (Figure 8A). The seed coat was rich in lipids and the underlying aleurone and sub-aleurone endosperm layers were protein-rich. An XRF (Figure 8D) false-coloured image revealed Ca (red), K (green) and Mn (blue) distribution is the adjacent

section of the same wheat seed. These data showed prominent K in the bran (consisting of pericarp, seed coat and aleurone) and embryonic axis, preferential presence of Mn in association with the endosperm, and Ca in the embryo (Figure 8D).

4 Discussion

Widespread availability of vast genomic resources has ushered in a "post-genomics" era for research in the life sciences, in which a focus on integrating genetic and phenotypic data to functionally analyze and decipher biological phenomena is forefront (Tuggle et al., 2024). Synchrotron imaging offers an array of tools for uniquely accessing phenotypic traits, including comprehensive measures of an organisms' internal architecture and metabolic composition. Here, we featured a panel of seed datasets analyzed at the Canadian Light Source (CLS) to demonstrate the potential for synchrotron technology to reveal the inner workings of seeds. Methods detail seed preparation possibilities for imaging at five synchrotron beamlines (including BMIT-BM, MidIR spectroscopy, BioXAS-Imaging, SXRMB, and SM). Data presented demonstrate the value of imaging tools available at each beamline for studies of seed embryo, endosperm and seed coat anatomy, development, physiology and biochemistry. While each beamline can inform different questions, the use of multimodal imaging with synchrotron technology is not common in the literature, especially in the field of plant science. Indore et al. (2023) demonstrate the use of μCT, MidIR and XRF to determine changes in barley seed during storage, however, where multimodal SR imaging exists, it is generally limited to CT and XRF (e.g Varga et al., 2020; Calo et al., 2022; Nakhforoosh et al., 2024). We demonstrate the power of combining synchrotron hard and soft X-ray modalities to resolve phenotypic and metabolic complexities and correlate structural and chemical datasets within the same seed. Together with genomic insights, synchrotron technologies support comprehensive seed analyses and open doors to informed crop improvement.

SR-μCT imaging (using BMIT-BM beamline) provides superior penetration into specimens and is ideally suited to the study of whole seeds. SR-μCT imaging reveals the complex internal anatomy of seeds with micron and submicron resolutions, allowing cellular boundaries and subcellular features to be distinguished. Further, 3D volume rendering and modeling enable quantitative measures of complex forms to be made accurately, while providing novel views of the spatial relationships and interfaces that link seed subcompartments. SR-μCT imaging allows the extent of necrosis due to bacterial penetration into a seed to be distinguished from surrounding networks of seed vasculature (Figure 2). Localization of bacterial fruit blotch pathogen within watermelon seeds (case study 1) is dependent on the method of infection with embryonic localization resulting from pistil penetration (Dutta et al., 2012). Penetration into the embryo (as seen in the current study), as compared to perisperm-endosperm layer localization following ovary pericarp infection, is associated with increased survival of the bacteria following desiccation and exposure to seed treatments such as peroxyacetic acid and chlorine gas (Dutta et al., 2016). While seed infection status and seed treatment efficacy can be assessed using seedling grow-out assays, SR-μCT allows for visualization of seed health and could serve to increase understanding of pathogenesis and to aid the development of better treatments to control outbreaks. This study also provides a novel proof of concept for segmentation of necrotic regions of seed tissue following infection. Networks of air gaps can similarly be mapped with SR-μCT, to assess the impacts of prolonged dormancy and storage on seed viability (Cloetens et al., 2006). SR-μCT imaging can be applied to immature seeds to gain access to internal developmental processes that occur in each subcompartment. In pea (case study 2), this included precise volumetric size measures for all seed subcompartments and apoplastic spaces, the stage of embryonic development, the degree of endosperm degeneration, and the 3D architecture of inner seed coat layers not present in mature peas (Figure 5; Weber et al., 2005). 3D rendering at this resolution provides complete seed information thus allowing interrogation of relationships not possible with traditional 2D light microscopy. Perhaps most advantageous is the ability the BMIT-BM beamline offers to non-destructively examine mature seeds, bypassing cumbersome preparation efforts (Figure 1) and maintaining the possibility for post-imaging seed analyses, and even seed germination.

SR-μCT imaging presented here employed hard X-rays with 12.6 – 40 keV (BMIT-BM) energy range which is sufficient to visualize the internal structures of valued crop seeds in Canada. However, harder X-rays (available with high-energy BMIT-ID, with 28 – 140 keV photon energies) can accommodate SR-μCT imaging of large seeds and plant structures, as demonstrated for wheat root architecture (Supplementary Figure 2) as well as high density materials. In biological materials, the majority of elements have a low atomic weight, therefore only weakly absorb X-rays and make it difficult to differentiate tissues and structures using absorption-contrast imaging. Phase retrieval in seeds can improve visualization of embryonic cotyledons (Figure 2) and supply information regarding particle size and distribution based on histogram peak shapes and shifts (Figure 3). Phase retrieval is necessary in some instances such as mapping complex networks of vasculature (Figure 2) where conventional attenuation/absorption contrast limits feature demarcations. The use of SR-μCT imaging of seeds, in combination with targeted solvent extractions, can additionally provide indications of chemistry that is localized within tissues and cells (Figure 3).

The ability to obtain μCT images from plant tissues is not unique to synchrotron radiation as laboratory X-ray scanners can also perform this task (Karahara et al., 2023). A recent report indicates that the use of lab-based scanners is approaching that of synchrotron-based scanners (Zwanenburg et al., 2021). However, synchrotron-based scanners offer a number of key advantages, including faster scan times that can protect sample integrity and reduce motion artifacts for hydrated samples, adjustable propagation distances for increased flexibility and control in phase contrast imaging, and energy tunability which allows for optimization of contrast (Garcea et al., 2018; Kopittke et al., 2018). Phase retrieval is possible with some lab-based scanners but, the methodology required reduces the photon flux of these instruments, thus, extending scanning times (Quenot et al., 2022). The parallel, monochromatic, brighter beam with SR also results in a better signal to noise ratio than the divergent, polychromatic beam of lab scanners, yielding better image quality

in images of equal resolution. The energy tunability of synchrotron light also provides an advantage in elemental analysis over lab-based instruments, as SR X-ray absorption spectroscopy can probe distinct species (Kopittke et al., 2018).

Access to synchrotron facilities is a potential hurdle, especially if time is required on multiple beamlines. However, the assistance of expert beamline staff in drafting proposals and designing experiments as well as the subsequent peer-review process (https://www.lightsource.ca/users/getting-started/peer-review-process.php) governing access to beam time ensure high quality research is awarded time. The community of practice further opens access to users who are less familiar with these technologies and those who can't afford to purchase, develop and maintain these technologies. While lab instruments might be more accessible to some researchers, this is not the case for all and access to lab instruments for multiple imaging modalities is likely more rare.

Synchrotron-based imaging and spectroscopy are immensely powerful tools for the study of chemistry in heterogeneous biological materials. The techniques employing hard and soft X-rays in this study, together with Mid-IR spectroscopy, provide broad detection capacities for a range of plant metabolites. These synchrotron imaging tools provide spatial context to biochemical analyses, allowing metabolic landscapes inside seeds to be mapped with micro- and even nano-scale resolutions. Seed nutrition traits can therefore be assigned specifically to the embryo, endosperm or seed coat. Using MidIR spectroscopy, the distribution of protein, lipid and carbohydrate was revealed in protein-rich pea (case study 2; Figure 4), oil-rich canola (case study 3; Figures 5, 7), and carbohydrate-rich wheat (Figure 8) seeds. The use of canola seeds with contrasting phytate levels illustrates the value of spatially-resolved analytics for informed seed variety comparisons. The synchrotron-based imaging tools applied in case study 3 resolved the spatial relationship between oil and protein within the embryo. The sequestration of macromolecules within canola seed embryos is expected based on the presence and organization of protein storage vacuoles and oil bodies. These protein storage vacuoles are the expected location of the phytate crystals in canola seeds. In pea, spectra from seed coat sclerenchyma and parenchyma cell layers of the immature seed coat, as well as features within the embryo (including the radicle and cotyledons) were readily distinguished by distinct macronutrient profiles (Figure 4). The contrasting chemistries of wheat grain subcompartments (Figure 8) were particularly marked, with stark distinctions in the macronutrient reserve. These results highlight the radial gradient of the starchy endosperm with the sub-aleurone layer being rich in protein and relatively limited in starch content (for review see Shewry et al., 2020). The addition of correlative μCT could aid in the identification of cell types with these different distribution patterns throughout the seed.

Conventional MidIR spectroscopy provides a valued overview of the macronutrients within seed subcompartments, resolving organ and tissue chemistries, and prominent cell layers within these subcompartments. IR spectroscopy's detection potentially extends far beyond the macronutrient-focused seed analyses presented here to a range of organic macromolecules (see reviews by Schulz and Baranska, 2007; Dokken et al., 2005; Türker-Kaya and Huck, 2017; Cozzolino, 2012; Abidi, 2021). Of particular interest to plant research, cell wall polysaccharides such as cellulose, hemicelluloses, and pectin (Lahlali et al., 2015; Kumar et al., 2016; Pereira et al., 2018), biopolymers including suberin, lignin, and cutin (Zeier and Schreiber, 1999; Moore and Owen, 2001; Yan et al., 2009; Tanino et al., 2017), and diverse specialized metabolites and natural products (such as phenolic compounds, flavonoids, alkaloids, glucosinolates, essential oils, terpenes and glycosides; see review by Cozzolino, 2009) can be resolved by IR spectroscopy. The majority of MidIR spectroscopy chemical imaging presented here employed a conventional IR source, rather than bright synchrotron light. Synchrotron IR spectroscopy offers improved signal-to-noise over conventional IR, which improves lateral spatial resolutions for spectra that can be distinguished at the cellular level (Figure 5). Sub-diffraction limited IR imaging techniques such as optical photothermal IR (O-PTIR) and nano-scale IR techniques offer additional IR tools for chemically analyzing biological materials with sub-micron and nanometer resolutions, respectively.

Micronutrient maps across seeds can be captured in parallel with macronutrient maps for comprehensive analysis of seed nutrient compositions and localization patterns. XRF employing hard X-rays and soft X-rays was used to illustrate the range of detectable elements that can be mapped across seeds. The CLS BioXAS-Imaging beamline is well suited for XRF elemental mapping of biologically relevant elements (e.g. K, Ca, Mg, Mn, Fe, Zn, Cu) as shown in Figures 4, 7 and 8. Soft X-ray XRF performed at the CLS SXRMB increases the range of detectable elements, being ideally suited to lighter atomic weight elements (e.g. Si, S, P, Cl, K, Ca) as featured in Figure 5. SXRMB detection of some heavier elements, such as Cd, can be accessed through their L-edge emission.

Soft X-ray spectromicroscopy (SM) allows imaging elements with lower atomic weights (e.g. C, N, O, Na, Mg, Si, P) than SXRMB and BioXAS, and provides spatial resolution to probe composition of fine subcellular structures within a seed. Here, we illustrated the value of SM for comparing canola seeds with contrasting phytate levels (Figure 6). SM was able to distinguish punctate P deposits independent of the protein content of the cell. These are indicative of phytate crystals in a high phytate genotype as compared to a generalized P localization in the low phytate line. Using XRF imaging with SM, it is also possible to collect 3D data as demonstrated by Kim et al. (2006). While 3D XRF can provide important spatial context for understanding elemental distribution in samples, there are a number of challenges with 3D XRF in biological tissues. The sample must be stable under extended scan times and 3D rendering for low atomic weight, biologically relevant elements is limited to thin samples due to self-absorption of the XRF signal. Self absorption is also an important consideration for 2D XRF as highlighted by Pushie et al. (2022). In addition to informing elemental distributions in seed cells, SM can be used to determine differences in elemental speciation with high spatial resolution. XANES of select elements (P at the SXRMB beamline, Figure 5; zinc at the BioXAS-Imaging beamline, Figure 4) demonstrates this capacity. XANES spectra for Zn in developing pea seeds show different species of Zn between seed tissues. With the availability of appropriate standards, the specific species could

be identified. XANES at SXRMB was not able to differentiate P species at this resolution, however, in general, speciation detection of P in biological samples is problematic as there are multiple species in a complex matrix and the K-edge spectra have limited distinguishable features (Schmieder et al., 2020). Greater differentiation of spectra is possible using P $L_{2,3}$-edge XANES however, the concentration of P required is significantly greater than that for detecting P with its K-edge (Kruse et al., 2009).

The ability to map metabolic compositions to the seed coat, embryo and endosperm allows spatially resolved insights into macronutrient and micronutrient compositions in seeds. We highlight the flexible and complementary nature of synchrotron imaging tools for seed research. Cross-platform synchrotron imaging, as presented here, strengthens understanding of structural and compositional data. Although technically challenging, obtaining correlated datasets for the same seed across beamline techniques is increasingly feasible. Semi-correlative data presented used neighbouring sections for macro- and micronutrient profiling in pea (Figure 4), and correlative microscopy across synchrotron platforms deciphered macro- and micronutrient compositions and structural landscapes at the same locale within a single canola seed (Figure 7). XRF highlights the abundance of iron in the putative endodermal cells surrounding the vasculature (Figure 7F) as previously demonstrated in arabidopsis (Kim et al., 2006; Grillet et al., 2014) and *Brassica napus* (Ibeas et al., 2017). This demonstrates the value of imaging across beamlines, as the cell types associated with iron signal from XRF could be directly correlated with cell types in the same seed in SR-μCT images. This correlative imaging approach provides novel information not gleaned from either beamline used individually. Continued efforts to support correlative microscopy across synchrotron beamlines promise to facilitate direct comparisons between structural and compositional imaging datasets, for more complete understanding of the physical and chemical properties that function within heterogeneous living systems. By revealing spatiotemporal complexities of the composition and architecture of embryos, endosperm and seed coat, with opportunities for non-destructive and correlated imaging, these methodologies have immense value potential for seed biologists, seed bank curators, and breeders alike.

The computational resources and expertise required to analyze large SR-generated data presents a considerable limitation to the use of SR technology. Enhancements in analysis pipelines, AI and machine learning algorithms promise to reduce bottlenecks in most existing SR data analysis pipelines. For example, Albers et al. (2024) describe a high-throughout tomography platform with an automated data processing pipeline that provides a 3D reconstructed data set less than two minutes after the X-ray scan is complete. Segmentation of 3D μCT data is an additional bottleneck, requiring time, experience and extensive computational resources. Advances in deep learning have provided tools such as Biomedisa (Lösel et al., 2020) and Google's Colaboratory web application (described by Rippner et al., 2022) to increase the speed, computational efficiency and accessibility of segmentation. Modules available in Quasar (Toplak et al., 2021) include machine learning algorithms for the analysis of spectroscopy data and are accessible to non-computer experts. Excellent reviews of these algorithms and advances in this space are provided by Meza Ramirez et al. (2020) and An et al. (2022). As novel tools in this space continue to emerge, accessibility of SR-based technology will become increasingly feasible for researchers. Growth of SR-guided research could enable the integration of anatomic and biochemical phenotypes with the genotypic and transcriptomic datasets more commonly available, to enhance critical understanding of the fundamental biological processes and the mechanisms that underlie seed trait variation. Developments in the field of SR-driven research suggest a bright future for synchrotron-informed research in plant sciences.

Data availability statement

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

Author contributions

PA: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. KYT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. JAS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. JJD: Formal analysis, Writing – review & editing, Methodology. MV: Formal analysis, Investigation, Writing – review & editing. HS: Investigation, Writing – review & editing, Data curation, Formal analysis. SK: Investigation, Resources, Writing – review & editing. KKT: Resources, Writing – review & editing. JW: Resources, Writing – review & editing. SV: Resources, Writing – review & editing. CK: Conceptualization, Resources, Supervision, Writing – review & editing. TDQ: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the Sustainable Protein Production program, in the Aquatic and Crop Resource Development Division of the National Research Council of Canada (NRC DAF# 58477). Following scientific peer-review and selection, all experimental time at the Canadian Light Source was made possible by the CLS General User Access program.

Acknowledgments

A large part of the research herein was performed at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and

Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan. SR-µCT data handling, processing, and analysis was partially supported by the grants and contributions program of the NRC. We thank Canadian Light Source beamline scientists Drs. Sergey Gasilov, Gosia Korbas, Scott Rosendahl, and Viorica (Ibi) Bondici for expert guidance and technical support. Assistance with sample preparations provided by Dr. Guosheng Liu, Department of Biology, University of Saskatchewan, is gratefully acknowledged. We thank Prof. Gregory E. Welbaum, Virginia Tech for sharing watermelon seed data, Dr. Gopalan Selvaraj, NRC, for coordinating and sharing the wheat root architecture study, Charlotte Fleet, McMaster University for assistance with canola seed IR data acquisition and processing and Nirmal Joseph, Saskatchewan Polytechnic for assistance with canola XRF imaging. We thank Dr. Nii Patterson, NRC, for review of the manuscript.

Conflict of interest

Authors KYT, JS, JD, MV, and CK were employed by Canadian Light Source Inc.

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.2024.1395952/full#supplementary-material

SUPPLEMENTARY FIGURE 1
SR-µCT imaging of a developing pea seed showing cellular structure of the seed coat and embryo **(A)** and the localization and structure of the 3D rendered embryo within the seed coat **(B)** and at different orientations **(C, D)**. Scale bar 800 µm.

SUPPLEMENTARY FIGURE 2
SR-µCT imaging of wheat roots in soil, produced at the Canadian Light Source BMIT-ID beamline. Plants were grown in Sunshine Mix #3 in 7.5 cm diameter customized rhizotrons. The top ~15 cm of the 120 cm rhizotron was imaged. Roots were manually segmented from SR-µCT slices using Avizo software **(A)**; red and green are examples of segmented primary roots). Three-dimensional volume rendering of primary roots is shown in **(B)** and secondary roots are shown with the primary roots in **(C)**.

SUPPLEMENTARY VIDEO 1
Sequential slices through SR-µCT imaged watermelon (*Citrullus lanatus*) seeds prior to segmentation. A healthy seed is seen in the left panel while the right panel shows a seed infected with bacterial fruit blotch *Acidovorax citrulli*.

SUPPLEMENTARY VIDEO 2
SR-µCT of bacterial fruit blotch infection (*Acidovorax citrulli*) necrosis (blue) as occurring in sequential slices of a watermelon (*Citrullus lanatus*) seed. See Figure 2D for a single slice (not coloured) selected from this image series.

SUPPLEMENTARY VIDEO 3
Three-dimensional rendering of watermelon (*Citrullus lanatus*) seed showing bacterial fruit blotch (blue) produced by *Acidovorax citrulli* and seed vasculature (red) following SR-µCT. See Figure 6C for a representative plane from this 3D view.

SUPPLEMENTARY VIDEO 4
Three-dimensional rendering of a canola (*Brassica napus*) seed SR-µCT image highlighting the location of the false-coloured slice shown in Figure 7C.

SUPPLEMENTARY VIDEO 5
High resolution, sequential slices through an SR-µCT-imaged canola (*Brassica napus*) seed, prior to segmentation.

References

Abidi, N. (2021). *FTIR microspectroscopy: Selected emerging applications* (Cham, Switzerland: Springer International Publishing). doi: 10.1007/978-3-030-84426-4

Albers, J., Nikolova, M., Svetlove, A., Darif, N., Lawson, M. J., Schneider, T. R., et al. (2024). High throughput tomography (HiTT) on EMBL beamline P14 on PETRA III. *J. Synchrotron Radiat.* 31, 186–194. doi: 10.1107/s160057752300944x

An, D., Zhang, L., Liu, Z., Liu, J., and Wei, Y. (2022). Advances in infrared spectroscopy and hyperspectral imaging combined with artificial intelligence for the detection of cereals quality. *Crit. Rev. Food Sci. Nutr.* 63, 9766–9796. doi: 10.1080/10408398.2022.2066062

Balerna, A., and Mobilio, S. (2014). "Introduction to synchrotron radiation," in *Synchrotron radiation* (Berlin, Heidelberg: Springer Berlin Heidelberg), 3–28. doi: 10.1007/978-3-642-55315-8_1

Bharti, A., and Goyal, N. (2019). "Fundamental of synchrotron radiations," in *Synchrotron radiation - useful and interesting applications* (London, UK: IntechOpen). doi: 10.5772/intechopen.82202

Brown, C. H., Gulden, R. H., Shirtliffe, S. J., and Vail, S. L. (2024). The evaluation of secondary seed dormancy potentials of spring brassica napus l. Genotypes and the relationship with seed germination, vigor, and seed quality traits. *Crop Sci.* 64, 1542–1558. doi: 10.1002/csc2.21205

Burdman, S., and Walcott, R. (2012). *Acidovorax citrulli*: Generating basic and applied knowledge to tackle a global threat to the cucurbit industry. *Mol. Plant Pathol.* 13, 805–815. doi: 10.1111/j.1364-3703.2012.00810.x

Calo, C. M., Rizzutto, M. A., Pérez, C. A., MaChado, R., Ferreira, C. G., Aguero, N. F., et al. (2022). Some notes on dense structures present in archaeological plant remains: X-ray fluorescence computed tomography applications. *Minerals* 12, 1130. doi: 10.3390/min12091130

Carvalho, F. C. Q., Santos, L. A., Dias, R. C. S., Mariano, R. L. R., and Souza, E. B. (2012). Selection of watermelon genotypes for resistance to bacterial fruit blotch. *Euphytica* 190, 169–180. doi: 10.1007/s10681-012-0766-1

Cloetens, P., Mache, R., Schlenker, M., and Lerbs-Mache, S. (2006). Quantitative phase tomography of arabidopsis seeds reveals intercellular void network. *Proc. Natl. Acad. Sci.* 103, 14626–14630. doi: 10.1073/pnas.0603490103

Cozzolino, D. (2009). Near infrared spectroscopy in natural products analysis. *Planta Med.* 75, 746–756. doi: 10.1055/s-0028-1112220

Cozzolino, D. (2012). Recent trends on the use of infrared spectroscopy to trace and authenticate natural and agricultural food products. *Appl. Spectrosc. Rev.* 47, 518–530. doi: 10.1080/05704928.2012.667858

Dokken, K. M., Davis, L. C., and Marinkovic, N. S. (2005). Use of infrared microspectroscopy in plant growth and development. *Appl. Spectrosc. Rev.* 40, 301–326. doi: 10.1080/05704920500230898

Doll, N. M., and Ingram, G. C. (2022). Embryo–endosperm interactions. *Annu. Rev. Plant Biol.* 73, 293–321. doi: 10.1146/annurev-arplant-102820-091838

Dutta, B., Avci, U., Hahn, M. G., and Walcott, R. R. (2012). Location of *Acidovorax citrulli* in infested watermelon seeds is influenced by the pathway of bacterial invasion. *Phytopathology* 102, 461–468. doi: 10.1094/phyto-10-11-0286-r

Dutta, B., Schneider, R. W., Robertson, C. L., and Walcott, R. R. (2016). Embryo localization enhances the survival of *Acidovorax citrulli* in watermelon seeds. *Phytopathology* 106, 330–338. doi: 10.1094/phyto-09-15-0232-r

Dynes, J. J., Regier, T. Z., Snape, I., Siciliano, S. D., and Peak, D. (2015). Validating the scalability of soft x-ray spectromicroscopy for quantitative soil ecology and biogeochemistry research. *Environ. Sci. Technol.* 49, 1035–1042. doi: 10.1021/es505271p

Ebersbach, J., Khan, N. A., McQuillan, I., Higgins, E. E., Horner, K., Bandi, V., et al. (2022). Exploiting high-throughput indoor phenotyping to characterize the founders of a structured *B. napus* breeding population. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.780250

Eckshtain-Levi, N., Munitz, T., Živanović, M., Traore, S. M., Spröer, C., Zhao, B., et al. (2014). Comparative analysis of type III secreted effector genes reflects divergence of *Acidovorax citrulli* strains into three distinct lineages. *Phytopathology* 104, 1152–1162. doi: 10.1094/phyto-12-13-0350-r

Faragó, T., Gasilov, S., Emslie, I., Zuber, M., Helfen, L., Vogelgesang, M., et al. (2022). Tofu: A fast, versatile and user-friendly image processing toolkit for computed tomography. *J. Synchrotron Radiat.* 29, 916–927. doi: 10.1107/s160057752200282x

Feng, X., Liu, N., Zhang, H., and Yu, P. (2020). Chemical imaging of the microstructure of chickpea seed tissue within a cellular dimension using synchrotron infrared microspectroscopy: A preliminary study. *J. Agric. Food Chem.* 68, 11586–11593. doi: 10.1021/acs.jafc.0c04446

Garcea, S. C., Wang, Y., and Withers, P. J. (2018). X-ray computed tomography of polymer composites. *Composites Sci. Technol.* 156, 305–319. doi: 10.1016/j.compscitech.2017.10.023

Gasilov, S., Webb, M. A., Panahifar, A., Zhu, N., Marinos, O., Bond, T., et al. (2024). Hard x-ray imaging and tomography at the biomedical imaging and therapy beamlines of canadian light source. *J. Synchrotron Radiat.* 31, 1346–1357. doi: 10.1107/s1600577524005241

Grillet, L., Mari, S., and Schmidt, W. (2014). Iron in seeds – loading pathways and subcellular localization. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00535

Guendel, A., Rolletschek, H., Wagner, S., Muszynska, A., and Borisjuk, L. (2018). Micro imaging displays the sucrose landscape within and along its allocation pathways. *Plant Physiol.* 178, 1448–1460. doi: 10.1104/pp.18.00947

Hitchcock, A. P. (2023). Analysis of x-ray images and spectra (aXis2000): A toolkit for the analysis of x-ray spectromicroscopy data. *J. Electron Spectrosc. Related Phenomena* 266, 147360. doi: 10.1016/j.elspec.2023.147360

Hossain, M. T., Liyanage, S., and Abidi, N. (2022). FTIR microspectroscopic approach to investigate macromolecular distribution in seed coat cross-sections. *Vibrational Spectrosc.* 120, 103376. doi: 10.1016/j.vibspec.2022.103376

Huss, J. C., and Gierlinger, N. (2021). Functional packaging of seeds. *New Phytol.* 230, 2154–2163. doi: 10.1111/nph.17299

Ibeas, M. A., Grant-Grant, S., Navarro, N., Perez, M. F., and Roschzttardtz, H. (2017). Dynamic subcellular localization of iron during embryo development in brassicaceae seeds. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02186

Indore, N. S., Jayas, D. S., Karunakaran, C., Stobbs, J., Bondici, V. F., Vu, M., et al. (2023). Study of microstructural, nutritional, and biochemical changes in hulled and hulless barley during storage using x-ray and infrared techniques. *Foods* 12, 3935. doi: 10.3390/foods12213935

Indore, N. S., Karunakaran, C., and Jayas, D. S. (2022). Synchrotron tomography applications in agriculture and food sciences research: A review. *Plant Methods* 18, 101. doi: 10.1186/s13007-022-00932-9

Indore, N. S., Karunakaran, C., Jayas, D. S., Stobbs, J., Vu, M., Tu, K., et al. (2024). Characterization of spring and durum wheat using non-destructive synchrotron phase contrast x-ray microtomography during storage. *NPJ Sci. Food* 8, 29. doi: 10.1038/s41538-024-00271-0

Iwai, T., Takahashi, M., Oda, K., Terada, Y., and Yoshida, K. T. (2012). Dynamic changes in the distribution of minerals in relation to phytic acid accumulation during rice seed development. *Plant Physiol.* 160, 2007–2014. doi: 10.1104/pp.112.206573

Karahara, I., Yamauchi, D., Uesugi, K., and Mineyuki, Y. (2023). Three-dimensional visualization of plant tissues and organs by x-ray micro–computed tomography. *Microscopy* 72, 310–325. doi: 10.1093/jmicro/dfad026

Kaznatcheev, K. V., Karunakaran, C.H., Lanke, U. D., Urquhart, S. G., Obst, M., and Hitchcock, A. P. (2007). Soft x-ray spectromicroscopy beamline at the CLS: Commissioning results. *Nucl. Instruments Methods Phys. Res. Section A: Accelerators Spectrometers Detectors Associated Equip.* 582, 96–99. doi: 10.1016/j.nima.2007.08.083

Kim, S. A., Punshon, T., Lanzirotti, A., Li, L., Alonso, J. M., Ecker, J. R., et al. (2006). Localization of iron in arabidopsis seed requires the vacuolar membrane transporter VIT1. *Science* 314, 1295–1298. doi: 10.1126/science.1132563

Kopittke, P. M., Punshon, T., Paterson, D. J., Tappero, R. V., Wang, P., Blamey, F. P. C., et al. (2018). Synchrotron-based x-ray fluorescence microscopy as a technique for imaging of elements in plants. *Plant Physiol.* 178, 507–523. doi: 10.1104/pp.18.00759

Kruse, J., Leinweber, P., Eckhardt, K.-U., Godlinski, F., Hu, Y., and Zuin, L. (2009). Phosphorus $L_{2,3}$-edge XANES: Overview of reference compounds. *J. Synchrotron Radiat.* 16, 247–259. doi: 10.1107/s0909049509000211

Kumar, S., Lahlali, R., Liu, X., and Karunakaran, C. (2016). Infrared spectroscopy combined with imaging: A new developing analytical tool in health and plant science. *Appl. Spectrosc. Rev.* 51, 466–483. doi: 10.1080/05704928.2016.1157808

Lahlali, R., Karunakaran, C., Wang, L., Willick, I., Schmidt, M., Liu, X., et al. (2015). Synchrotron based phase contrast x-ray imaging combined with FTIR spectroscopy reveals structural and biomolecular differences in spikelets play a significant role in resistance to fusarium in wheat. *BMC Plant Biol.* 15, 24. doi: 10.1186/s12870-014-0357-5

Le, T. D. Q., Alvarado, C., Girousse, C., Legland, D., and Chateigner-Boutin, A.-L. (2019). Use of x-ray micro computed tomography imaging to analyze the morphology of wheat grain through its development. *Plant Methods* 15, 84. doi: 10.1186/s13007-019-0468-y

Legland, D., Alvarado, C., Badel, E., Guillon, F., King, A., Le, T. D. Q., et al. (2022). Synchrotron based x-ray microtomography reveals cellular morphological features of developing wheat grain. *Appl. Sci.* 12, 3454. doi: 10.3390/app12073454

Lösel, P. D., Kamp, T., van de, J. A., Ershov, A., Faragó, T., Pichler, O., et al. (2020). Introducing Biomedisa as an open-source online platform for biomedical image segmentation. *Nat. Commun.* 11, 5577. doi: 10.1038/s41467-020-19303-w

Madsen, C. K., and Brinch-Pedersen, H. (2020). Globoids and phytase: The mineral storage and release system in seeds. *Int. J. Mol. Sci.* 21, 7519. doi: 10.3390/ijms21207519

Meza Ramirez, C. A., Greenop, M., Ashton, L., and Rehman, I. (2020). Applications of machine learning in spectroscopy. *Appl. Spectrosc. Rev.* 56, 733–763. doi: 10.1080/05704928.2020.1859525

Montanha, G. S., Romeu, S. L. Z., Marques, J. P. R., Rohr, L. A., Almeida, E., dos, A. R., et al. (2022). Microprobe-XRF assessment of nutrient distribution in soybean, cowpea, and kidney bean seeds: A fabaceae family case study. *ACS Agric. Sci. Technol.* 2, 1318–1324. doi: 10.1021/acsagscitech.2c00260

Moore, A. K., and Owen, N. L. (2001). Infrared spectroscopic studies of solid wood. *Appl. Spectrosc. Rev.* 36, 65–86. doi: 10.1081/asr-100103090

Nakhforoosh, A., Hallin, E., Karunakaran, C., Korbas, M., Stobbs, J., and Kochian, L. (2024). Visualization and quantitative evaluation of functional structures of soybean root nodules via synchrotron x-ray imaging. *Plant Phenomics* 6, 0203. doi: 10.34133/plantphenomics.0203

Pereira, L., Flores-Borges, D. N. A., Bittencourt, P. R. L., Mayer, J. L. S., Kiyota, E., Araújo, P., et al. (2018). Infrared nanospectroscopy reveals the chemical nature of pit membranes in water-conducting cells of the plant xylem. *Plant Physiol.* 177, 1629–1638. doi: 10.1104/pp.18.00138

Preedy, V. R., Watson, R. R., and Patel, V. B. (2010). *Nuts and seeds in health and disease prevention* (San Diego: Academic Press). doi: 10.1016/B978-0-12-375688-6.10141-0

Pushie, M. J., Sylvain, N. J., Hou, H., Hackett, M. J., Kelly, M. E., and Webb, S. M. (2022). X-ray fluorescence microscopy methods for biological tissues. *Metallomics* 14, mfac032. doi: 10.1093/mtomcs/mfac032

Quenot, L., Bohic, S., and Brun, E. (2022). X-ray phase contrast imaging from synchrotron to conventional sources: A review of the existing techniques for biological applications. *Appl. Sci.* 12, 9539. doi: 10.3390/app12199539

Radchuk, V., and Borisjuk, L. (2014). Physical, metabolic and developmental functions of the seed coat. *Front. Plant Sci.* 5. doi: 10.3389/fpls.2014.00510

Ramalho, J. J., Jones, V. A. S., Mutte, S., and Weijers, D. (2021). Pole position: How plant cells polarize along the axes. *Plant Cell* 34, 174–192. doi: 10.1093/plcell/koab203

Ravel, B., and Newville, M. (2005). ATHENA, ARTEMIS, HEPHAESTUS: Data analysis for x-ray absorption spectroscopy using IFEFFIT. *J. Synchrotron Radiat.* 12, 537–541. doi: 10.1107/s0909049505012719

Ren, Z.-W., Yang, M., McKenna, B. A., Lian, X.-M., Zhao, F.-J., Kopittke, P. M., et al. (2022). Fast x-ray fluorescence microscopy provides high-throughput phenotyping of element distribution in seeds. *Plant Physiol.* 191, 1520–1534. doi: 10.1093/plphys/kiac534

Rippner, D. A., Raja, P. V., Earles, J. M., Momayyezi, M., Buchko, A., Duong, F. V., et al. (2022). A workflow for segmenting soil and plant x-ray computed tomography images with deep learning in google's colaboratory. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.893140

Romeu, S. L. Z., Marques, J. P. R., Montanha, G. S., Carvalho, H. W. P., and de Pereira, F. M. V. (2021). Chemometrics unraveling nutrient dynamics during soybean seed germination. *Microchemical J.* 164, 106045. doi: 10.1016/j.microc.2021.106045

Rousseau, D., Widiez, T., Di Tommaso, S., Rositi, H., Adrien, J., Maire, E., et al. (2015). Fast virtual histology using x-ray in-line phase tomography: Application to the 3D anatomy of maize developing seeds. *Plant Methods* 11, 55. doi: 10.1186/s13007-015-0098-y

Sabelli, P. A., and Larkins, B. A. (2009). The development of endosperm in grasses. *Plant Physiol.* 149, 14–26. doi: 10.1104/pp.108.129437

Schmieder, F., Gustafsson, J. P., Klysubun, W., Zehetner, F., Riddle, M., Kirchmann, H., et al. (2020). Phosphorus speciation in cultivated organic soils revealed by p k-edge XANES spectroscopy. *J. Plant Nutr. Soil Sci.* 183, 367–381. doi: 10.1002/jpln.201900129

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089

Schulz, H., and Baranska, M. (2007). Identification and quantification of valuable plant substances by IR and raman spectroscopy. *Vibrational Spectrosc.* 43, 13–25. doi: 10.1016/j.vibspec.2006.06.001

Shewry, P. R., Wan, Y., Hawkesford, M. J., and Tosi, P. (2020). Spatial distribution of functional components in the starchy endosperm of wheat grains. *J. Cereal Sci.* 91, 102869. doi: 10.1016/j.jcs.2019.102869

Sivakumar, C., Stobbs, J. A., Tu, K., Karunakaran, C., and Paliwal, J. (2024). Unravelling particle morphology and flour porosity of roller-milled green lentil flour using scanning electron microscopy and synchrotron x-ray micro-computed tomography. *Powder Technol.* 436, 119470. doi: 10.1016/j.powtec.2024.119470

Solé, V. A., Papillon, E., Cotte, M., Walter, H., and Susini, J. (2007) A multiplatform code for the analysis of energy-dispersive x-ray fluorescence spectra. *Spectrochimica Acta Part B: Atomic Spectrosc.* 62, 63–68doi: 10.1016/j.sab.2006.12.002

Tanino, K., Willick, I., Hamilton, K., Vijayan, P., Jiang, Y., Brar, G. S., et al. (2017). Chemotyping using synchrotron mid-infrared and x-ray spectroscopy to improve agricultural production. *Can. J. Plant Sci.* 97, 982–996. doi: 10.1139/cjps-2016-0376

Thompson, L. U. (1990). "Phytates in canola/rapeseed," in *Canola and rapeseed* (Boston, MA: Springer US), 173–192. doi: 10.1007/978-1-4615-3912-4_10

Tofoni, A., Tavani, F., Persson, I., and D'Angelo, P. (2023). P k-edge XANES calculations of mineral standards: Exploring the potential of theoretical methods in the analysis of phosphorus speciation. *Inorganic Chem.* 62, 11188–11198. doi: 10.1021/acs.inorgchem.3c01346

Toplak, M., Birarda, G., Read, S., Sandt, C., Rosendahl, S. M., Vaccari, L., et al. (2017). Infrared Orange: Connecting hyperspectral data with machine learning. *Synchrotron Radiat. News* 30, 40–45. doi: 10.1080/08940886.2017.1338424

Toplak, M., Read, S. T., Sandt, C., and Borondics, F. (2021). Quasar: Easy machine learning for biospectroscopy. *Cells* 10,2300. doi: 10.3390/cells10092300

Tuggle, C. K., Clarke, J. L., Murdoch, B. M., Lyons, E., Scott, N. M., Beneš, B., et al. (2024). Current challenges and future of agricultural genomes to phenomes in the USA. *Genome Biol.* 25, 8. doi: 10.1186/s13059-023-03155-w

Türker-Kaya, S., and Huck, C. (2017). A review of mid-infrared and near-infrared imaging: Principles, concepts and applications in plant tissue analysis. *Molecules* 22, 168. doi: 10.3390/molecules22010168

van der Ent, J., de Jonge, M. D., de, M. D., Echevarria, G., Aarts, M. G. M., Mesjasz-Przybyłowicz, J., et al. (2022). Multimodal synchrotron x-ray fluorescence imaging reveals elemental distribution in seeds and seedlings of the zn–cd–ni hyperaccumulator *Noccaea caerulescens*. *Metallomics* 14, mfac026. doi: 10.1093/mtomcs/mfac026

Varga, T., Hixson, K. K., Ahkami, A. H., Sher, A. W., Barnes, M. E., Chu, R. K., et al. (2020). Endophyte-promoted phosphorus solubilization in populus. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.567918

Vijayan, P., Willick, I. R., Lahlali, R., Karunakaran, C., and Tanino, K. K. (2015). Synchrotron radiation sheds fresh light on plant research: The use of powerful techniques to probe structure and composition of plants. *Plant Cell Physiol.* 56, 1252–1263. doi: 10.1093/pcp/pcv080

Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi: 10.1146/annurev.arplant.56.032604.144201

Willick, I. R., Stobbs, J., Karunakaran, C., and Tanino, K. K. (2020). "Phenotyping plant cellular and tissue level responses to cold with synchrotron-based fourier-transform infrared spectroscopy and x-ray computed tomography," in *Plant cold acclimation* (New York NY: Humana), 141–159. doi: 10.1007/978-1-0716-0660-5_11

Xiao, Q., Maclennan, A., Hu, Y., Hackett, M., Leinweber, P., Sham, T.-K., et al. (2017). Medium-energy microprobe station at the SXRMB of the. *J. Synchrotron Radiat.* 24, 333–337. doi: 10.1107/s1600577516017604

Xiong, H., Wang, W., and Sun, M.-X. (2021). Endosperm development is an autonomously programmed process independent of embryogenesis. *Plant Cell* 33, 1151–1160. doi: 10.1093/plcell/koab007

Yamauchi, D., Tamaoki, D., Hayami, M., Uesugi, K., Takeuchi, A., Suzuki, Y., et al. (2012). "Extracting tissue and cell outlines of arabidopsis seeds using refraction contrast x-ray CT at the SPring-8 facility," in *AIP conference proceeding* (Tokyo, Japan: AIP), 237–242. doi: 10.1063/1.4742298

Yan, M., Guevara-Oquendo, V. H., Rodríguez-Espinosa, M. E., Yang, J.-C., Lardner, H., Christensen, D. A., et al. (2020). Utilization of synchrotron-based and globar-sourced mid-infrared spectroscopy for faba nutritional research about molecular structural and nutritional interaction. *Crit. Rev. Food Sci. Nutr.* 62, 1453–1465. doi: 10.1080/10408398.2020.1843397

Yan, H., Hua, Z., Qian, G., Wang, M., Du, G., and Chen, J. (2009). Analysis of the chemical composition of cotton seed coat by Fourier-transform infrared (FT-IR) microspectroscopy. *Cellulose* 16, 1099–1107. doi: 10.1007/s10570-009-9349-2

Zeier, J., and Schreiber, L. (1999). Fourier transform infrared-spectroscopic characterisation of isolated endodermal cell walls from plant roots: Chemical nature in relation to anatomical development. *Planta* 209, 537–542. doi: 10.1007/s004250050758

Zuo, C., Li, J., Sun, J., Fan, Y., Zhang, J., Lu, L., et al. (2020). Transport of intensity equation: A tutorial. *Optics Lasers Eng.* 135, 106187. doi: 10.1016/j.optlaseng.2020.106187

Zwanenburg, E. A., Williams, M. A., and Warnett, J. M. (2021). Review of high-speed imaging with lab-based x-ray computed tomography. *Measurement Sci. Technol.* 33, 012003. doi: 10.1088/1361-6501/ac354a

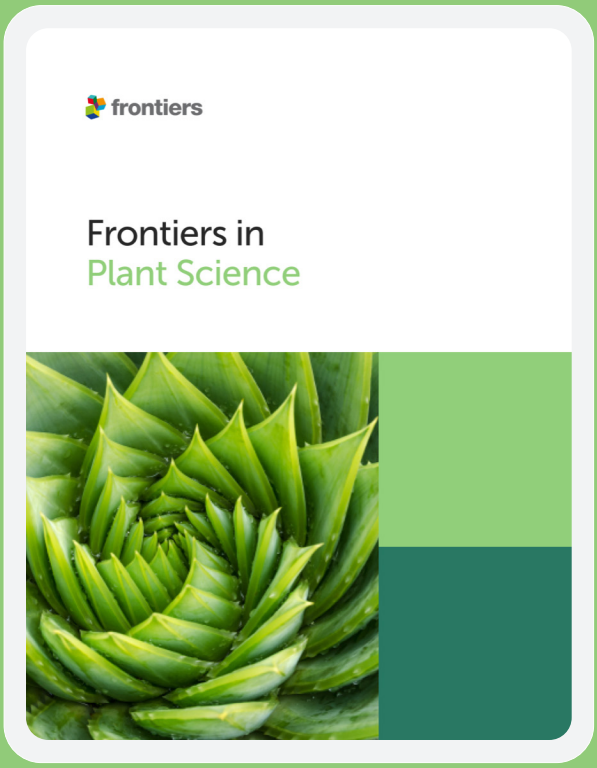
Frontiers in Plant Science

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