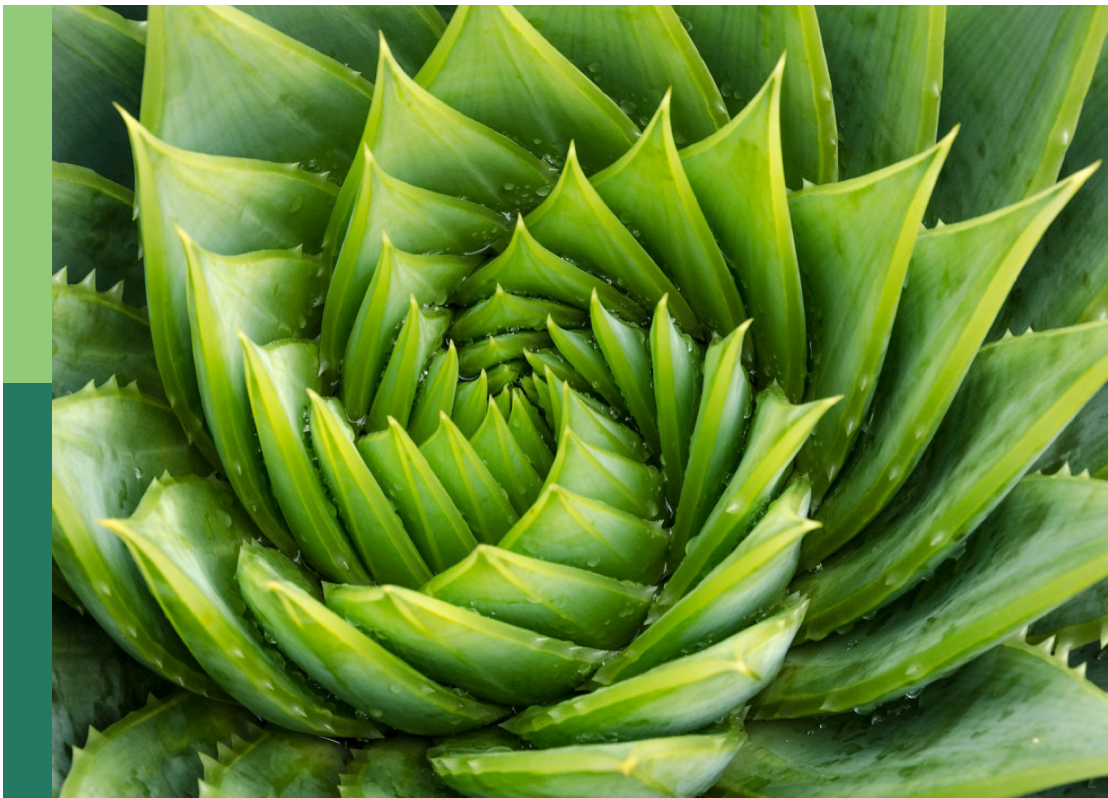
frontiers | Research Topics

Research advances on forest tree functional genomics and breeding

Edited by
Juan Du, Meng-Zhu Lu, Quanzi Li, Jack P. Wang, Yicun Chen and Jiehua Wang

Coordinated by
Yi An

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-5790-7
DOI 10.3389/978-2-8325-5790-7

About Frontiers

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

Frontiers journal series

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

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Research advances on forest tree functional genomics and breeding

Topic editors
Juan Du — Zhejiang University, China
Meng-Zhu Lu — Zhejiang Agriculture & Forestry University, China
Quanzi Li — Chinese Academy of Forestry, China
Jack P. Wang — North Carolina State University, United States
Yicun Chen — Chinese Academy of Forestry, China
Jiehua Wang — Tianjin University, China

Topic coordinator
Yi An — Zhejiang Agriculture & Forestry University, China

Citation
Du, J., Lu, M.-Z., Li, Q., Wang, J. P., Chen, Y., Wang, J., An, Y., eds. (2024). *Research advances on forest tree functional genomics and breeding*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5790-7

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 04 December 2024
DOI 10.3389/fpls.2024.1508507

OPEN ACCESS

EDITED AND REVIEWED BY
Huihui Li,
Chinese Academy of Agricultural Sciences, China

*CORRESPONDENCE
Juan Du
djuan@zju.edu.cn
Quanzi Li
20240082@zafu.edu.cn

RECEIVED 09 October 2024
ACCEPTED 14 November 2024
PUBLISHED 04 December 2024

CITATION
Du J, Ye T, An Y, Chen Y, Wang J, Wang J, Lu M and Li Q (2024) Editorial: Research advances on forest tree functional genomics and breeding.
Front. Plant Sci. 15:1508507.
doi: 10.3389/fpls.2024.1508507

Editorial: Research advances on forest tree functional genomics and breeding

Juan Du[1,2]*, Tianqi Ye[1], Yi An[3], Yicun Chen[4], Jack Wang[5], Jiehua Wang[3,6], Mengzhu Lu[3] and Quanzi Li[3]*

[1]State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, China, [2]Institute of Fundamental and Transdisciplinary, Zhejiang University, Hangzhou, Zhejiang, China, [3]College of Forestry and Biotechnology, Zhejiang A & F University, Hangzhou, China, [4]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing, China, [5]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, United States, [6]School of Environmental Science and Engineering, Tianjin University, Tianjin, China

KEYWORDS

forest tree, woody plants, functional genomics, breeding, genome editing

Editorial on the Research Topic
Research advances on forest tree functional genomics and breeding

Forest ecosystems—one of the biggest carbon sinks in the world—play a key role in terrestrial biodiversity and carbon sequestration. As important sustainable resources, trees serve as rich sources of agronomic and economic traits, providing wood, pulp and paper, fiber-related products, energy, and chemical products. Over the past few decades, conventional crossbreeding has helped generate plant varieties with improved agronomic and economic traits. However, conventional crossbreeding in forestry is time-intensive and has reached a bottleneck. Thus, improvement in growth and agronomic and economically important traits in tree species requires attention. Biotechnology has recently resulted in great progress in crop breeding, owing to the development of high-quality genome assembly and annotation tools, gene identification techniques, and efficient gene editing. Nonetheless, compared to crop species, extensive efforts are needed for the assembly and annotation of high-quality genomes, identification of key genes regulating agronomic and economically important traits, and highly efficient gene editing in tree species that exhibit high heterozygosity.

This Frontiers Research Topic aims to present the latest fundamental discoveries in the field of forest tree genomics, including genetic studies focusing on the genes and pathways associated with key agronomic and economically important traits, molecular mechanisms underlying secondary growth regulation, and the potential utilization of biotechnology in genetic improvement of woody plant species.

This volume is organized into the following sections: (1) genome assembly and annotation; (2) functional identification of key genes regulating tree growth, vascular development, and stress response; and (3) genetic transformation and gene editing in woody plants.

Genome assembly

The first tree species with a completely sequenced genome was *Populus trichocarpa* (Tuskan et al., 2006). With the increase in sequencing depth and length and decrease in sequencing cost, the genomes of several woody plants, including poplar, have been sequenced (Liu et al., 2019; Liu et al., 2020; Huang et al., 2021; Niu et al., 2022; Zhang et al., 2024). Researchers have also been able to sequence a nearly gapless, highly contiguous genome of a doubled haploid line of *P. ussuriensis*, whose 19 chromosomes have been assembled from telomere to telomere (Liu et al., 2024; Niu and Li, 2024; Shi et al., 2024). Additionally, Sahu et al. have assembled the genome of Indian rosewood at the chromosome level, providing the foundation for studying tree growth and wood formation in this species. The assembly of this reference genome is likely to promote research on the formation of heartwood, which is an economically important wood.

Functional identification of key genes regulating the growth, vascular development, and stress response of tree species

Tree biomass is determined by longitudinal and lateral growth. The cultivation of fast-growing trees is of great interest to produce adequate biomass in a shorter time. In this regard, researchers have attempted to identify the key genes regulating the growth rate of tree species for genetic improvement. Akutsu et al. used genomic selection (GS) for growth characteristics in an open-pollinated breeding population of Korean red pine (*Pinus densiflora*) and concluded that the trained GS model was more effective than traditional breeding methods. Kang et al. conducted GS prediction on a half-sib progeny population of *Shorea macrophylla* using six methods, and showed that GS with GWAS-based SNP selection was useful for breeding tree species.

Wood formation, or secondary growth, is a biological process specific to woody plants. The continuous activities of the cambium, comprising cambial cell proliferation, cell expansion, secondary wall thickening, and programmed cell death (PCD), produce xylem (wood) (Du et al., 2023). Key regulators of cambial cell proliferation and secondary wall thickening in poplar trees have been identified using forward and reverse genetics (Hu et al., 2022; Kim et al., 2022; Luo and Li, 2022; Dai et al., 2023; Li et al., 2024; Zhang et al., 2024; Zhou et al., 2024; Zhu and Li, 2024). However, studies on PCD in xylem cells are still limited.

Liu et al. stained different xylem cells using different dyes and sorted living cells, early PCD cells, and late PCD cells using flow cytometry. This method can also improve the analysis of gene expression dynamics along the continuous developmental stages of PCD. Guérin et al. generated dominant repressor poplar lines for *PtaPLATZ18*, which encodes an A/T-rich and zinc-binding protein. *PtaPLATZ18-SRDX* transgenic lines exhibited wider xylem compared to the wildtype plants, with a higher lignin content in transgenic wood. Moreover, the transgenic plants exhibited significantly increased height, suggesting the potential of this gene in promoting tree growth and wood production.

Tree growth is immensely affected by environmental factors such as drought and high temperatures (Jiang et al., 2023), which induce several physiological and molecular processes, such as abscisic acid synthesis in roots and leaves. Yu et al. analyzed the transcriptomic profile of *Phoebe bournei*, an important afforestation tree species in the subtropical region of China, exposed to drought stress. Through gene co-expression network analysis, they identified two core transcription factors, TGA4 and APRR2, involved in drought. Nonetheless, further genetic experiments are required to determine the functions of the candidate genes.

Genetic transformation and gene editing in woody plants

Genetic transformation is vital to characterize the function of a gene. In addition to genome assembly, successful genetic transformation using stems (Song et al., 2006) and leaves as explants (Li et al., 2017) makes *P. trichocarpa* a model species. As *P. trichocarpa* is an endemic species, researchers have resort to other poplar species, such as *P. alba*, *P. alba* × *P. glandulosa*, and *P. tremula*. Transformation in poplar, which relies on tissue culture, is mostly mediated by *Agrobacterium tumefaciens*, although *A. rhizogenes*-mediated transformation is independent of tissue culture and produces hairy roots, from which plants can regenerate. In this regard, Ying et al. reviewed recent advances in *A. rhizogenes*-mediated transformation and Ri breeding in woody plants and emphasized its potential application in the difficult-to-propagate woody species.

Gene editing has been increasingly applied for the genetic improvement of plant species because of its high precision (Borthakur et al., 2022). Fan et al. first used the type II clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein (Cas9) system in *P. tomentosa* (Fan et al., 2015). Furthermore, the CRISPR-Cas12 system has been used to knock out multiple targets of *Phytoene desaturase 8* in poplar (An et al., 2020). In this study, An et al. also evaluated the effects of temperature on gene editing efficiency and observed that the majority of editing included large-fragment deletions. Similarly, Movahedi et al. examined the effects of three factors on editing and showed that high *Agrobacteria* concentration, increased DDT number, and optimized homologous arm length resulted in efficient homology-directed repair. Currently, the CRISPR-Cas system is most commonly used for gene editing in poplar, but different woody plant species may need different CRISPR-Cas systems. Thus, increased efforts are needed to test the efficiency of gene editing in woody plants other than poplar.

This Research Topic incorporates articles on genome assembly and annotation; functional identification of key genes regulating tree growth, vascular development, and stress response; and genetic transformation and gene editing in woody plants, and serves as a valuable resource for studying forest tree breeding and functional genomics.

Author contributions

JD: Writing – review & editing, Writing – original draft. YA: Validation, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Visualization, Writing – review & editing. YC: Validation, Writing – review & editing. JPW: Writing – review & editing. JW: Validation, Writing – review & editing. ML: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing. QL: Writing – original draft, Writing – review & editing. TY: Data curation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is supported by the National Natural Science Foundation of China (32071792), National Key Program on 2023YFE124800, and the Key program of the National Science Foundation of Zhejiang province (LZ22C160002) to JD, Zhejiang University.

Conflict of interest

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

Publisher's note

References

An, Y., Geng, Y., Yao, J., Fu, C., Lu, M., Wang, C., et al. (2020). Efficient genome editing in populus using CRISPR/Cas12a. *Front. Plant Sci.* 11, 593938. doi: 10.3389/fpls.2020.593938

Borthakur, D., Busov, V., Cao, X. H., Du, Q., Gailing, O., Isik, F., et al. (2022). Current status and trends in forest genomics. *Forestry Res.* 2, 11. doi: 10.48130/FR-2022-0011

Dai, X., Zhai, R., Lin, J., Wang, Z., Meng, D., Li, M., et al. (2023). Cell-type-specific PtrWOX4a and PtrVCS2 form a regulatory nexus with a histone modification system for stem cambium development in Populus trichocarpa. *Nat. Plants* 9, 96–111. doi: 10.1016/j.molp.2023.03.005

Du, J., Wang, Y., Chen, W., Xu, M., Zhou, R., Shou, H., et al. (2023). High-resolution anatomical and spatial transcriptome analyses reveal two types of meristematic cell pools within the secondary vascular tissue of poplar stem. *Mol. Plant* 16, 809–828. doi: 10.1016/j.molp.2023.03.005

Fan, D., Liu, T., Li, C., Jiao, B., Li, S., Hou, Y., et al. (2015). Efficient CRISPR/cas9-mediated targeted mutagenesis in populus in the first generation. *Sci. Rep.* 5, 12217. doi: 10.1038/srep12217

Hu, J., Su, H., Cao, H., Wei, H., Fu, X., Jiang, X., et al. (2022). AUXIN RESPONSE FACTOR7 integrates gibberellin and auxin signaling via interactions between DELLA and AUX/IAA proteins to regulate cambial activity in poplar. *Plant Cell* 34, 2688–2707. doi: 10.1093/plcell/koac107

Huang, X., Chen, S., Peng, X., Bae, E.-K., Dai, X., Liu, G., et al. (2021). An improved draft genome sequence of hybrid Populus alba × Populus glandulosa. *J. Forestry Res.* 32, 1663–1672. doi: 10.1007/s11676-020-01235-2

Jiang, L., Zhang, X., Zhu, J., Wei, X., Chen, B., Liu, J., et al. (2023). Environmental heterogeneity determines beta diversity and species turnover for woody plants along an elevation gradient in subtropical forests of China. *Forestry Res.* 3. doi: 10.48130/FR-2023-0026

Kim, M. H., Bae, E. K., Lee, H., and Ko, J. H. (2022). Current understanding of the genetics and molecular mechanisms regulating wood formation in plants. *Genes* 13. doi: 10.3390/genes13071181

Li, S., Zhen, C., Xu, W., Wang, C., and Cheng, Y. (2017). Simple, rapid and efficient transformation of genotype nisqually-1: a basic tool for the first sequenced model tree. *Sci. Rep.* 7, 2638. doi: 10.1038/s41598-017-02651-x

Li, W., Lin, Y. J., Chen, Y. L., Zhou, C., Li, S., De Ridder, N., et al. (2024). Woody plant cell walls: Fundamentals and utilization. *Mol. Plant* 17, 112–140. doi: 10.1016/j.molp.2023.12.008

Liu, W., Liu, C., Chen, S., Wang, M., Wang, X., Yu, Y., et al. (2024). A nearly gapless, highly contiguous reference genome for a doubled haploid line of Populus ussuriensis, enabling advanced genomic studies. *Forestry Res.* 4, e019. doi: 10.48130/forres-0024-0016

Liu, Y. J., Wang, X. R., and Zeng, Q. Y. (2019). *De novo* assembly of white poplar genome and genetic diversity of white poplar population in irtysh river basin in china. *Sci. China. Life Sci.* 62, 609–618. doi: 10.1007/s11427-018-9455-2

Liu, P. L., Zhang, X., Mao, J. F., Hong, Y. M., Zhang, R. G., E, Y., et al. (2020). The tetracentron genome provides insight into the early evolution of eudicots and the formation of vessel elements. *Genome Biol.* 21, 291. doi: 10.1186/s13059-020-02198-7

Luo, L., and Li, L. (2022). Molecular understanding of wood formation in trees. *Forestry Res.* 2. doi: 10.48130/FR-2022-0005

Niu, S., and Li, Q. (2024). From haploid to reference: a new milestone in poplar genomics. *Forestry Res.* 4. doi: 10.48130/forres-0024-0018

Niu, S., Li, J., Bo, W., Yang, W., Zuccolo, A., Giacomello, S., et al. (2022). The chinese pine genome and methylome unveil key features of conifer evolution. *Cell* 185, 204–217.e214. doi: 10.1016/j.cell.2021.12.006

Shi, T. L., Jia, K. H., Bao, Y. T., Nie, S., Tian, X. C., Yan, X. M., et al. (2024). High-quality genome assembly enables prediction of allele-specific gene expression in hybrid poplar. *Plant Physiol.* 195, 652–670. doi: 10.1093/plphys/kiae078

Song, J., Lu, S., Chen, Z. Z., Lourenco, R., and Chiang, V. L. (2006). Genetic transformation of Populus trichocarpa genotype Nisqually-1: a functional genomic tool for woody plants. *Plant Cell Physiol.* 47, 1582–1589. doi: 10.1093/pcp/pcl018

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, populus trichocarpa (Torr. & gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691

Zhang, Y., Chen, S., Xu, L., Chu, S., Yan, X., Lin, L., et al. (2024). Transcription factor PagMYB31 positively regulates cambium activity and negatively regulates xylem development in poplar. *Plant Cell* 36. doi: 10.3389/fpls.2024.1343312

Zhou, H., Song, X., and Lu, M. Z. (2024). Growth-regulating factor 15-mediated vascular cambium differentiation positively regulates wood formation in hybrid poplar (Populus alba × p. glandulosa). *Front. Plant Sci.* 15, 1343312. doi: 10.3389/fpls.2024.1343312

Zhu, Y., and Li, L. (2024). Wood of trees: Cellular structure, molecular formation, and genetic engineering. *J. Integr. Plant Biol.* 66, 443–467. doi: 10.1111/jipb.13589

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fpls.2023.1159615

OPEN ACCESS

EDITED BY
Quanzi Li,
Chinese Academy of Forestry, China

REVIEWED BY
Vijay Sheri,
East Carolina University, United States
Ulhas S. Kadam,
Gyeongsang National University,
Republic of Korea

*CORRESPONDENCE
Chen Xu
xuchenidea@hotmail.com

SPECIALTY SECTION
This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 06 February 2023
ACCEPTED 23 March 2023
PUBLISHED 17 April 2023

CITATION
Movahedi A, Wei H, Kadkhodaei S, Sun W, Zhuge Q, Yang L and Xu C (2023) CRISPR-mediated genome editing in poplar issued by efficient transformation.
Front. Plant Sci. 14:1159615.
doi: 10.3389/fpls.2023.1159615

CRISPR-mediated genome editing in poplar issued by efficient transformation

Ali Movahedi[1,2], Hui Wei[3], Saeid Kadkhodaei[4], Weibo Sun[1], Qiang Zhuge[1], Liming Yang[1] and Chen Xu[5]*

[1]College of Biology and the Environment, Nanjing Forestry University, Nanjing, China, [2]College of Arts and Sciences, Arlington International University, Wilmington, DE, United States, [3]Key Laboratory of Landscape Plant Genetics and Breeding, School of Life Sciences, Nantong University, Nantong, China, [4]Agricultural Biotechnology Research Institute of Iran, Isfahan Branch, Agricultural Research, Education and Extension Organization, Isfahan, Iran, [5]Nanjing Key Laboratory of Quality and Safety of Agricultural Product, Nanjing Xiaozhuang University, Nanjing, China

Background: CRISPR has been increasingly used for plant genetic improvements because of its high efficiency and precision. Recently, the authors have reported the possibility of homology-directed repair (HDR) using CRISPR/Cas9 through woody plants such as poplar. HDR often replaces nucleotides with one donor DNA template (DDT), including homologous sequences.

Methods: CRISPR–Cas9 was recruited, and three variables, Agrobacteria inoculator concentration, pDDT/pgRNA ratio, and homologous arm length, were designed to integrate *nptII* and 2XCamV 35S into the *MKK2* promoter zone.

Results: Here, we showed that recovered poplars on kanamycin-supplemented media exhibited enhanced expression of *MKK2* affected by the precise integration of 2XcamV 35S and *nptII*, improving biochemical and phenotypic properties. Our findings confirmed that *Agrobacterium* inoculator OD_{600} = 2.5, increased DDT numbers during cell division to 4:1 pDDT/pgRNA, and optimized homologous arms 700 bp caused efficient HDR and increased *MKK2* expression.

Conclusion: Efficient transformations resulted from optimized variables, directly affecting the HDR efficiency through woody plants such as poplar.

KEYWORDS

CRISPR, efficient transformation, efficient HDR, DDT, poplar

Introduction

Genetic improvements and RNA-guided genome editing of plants are currently possible using bacterial type II (*Streptococcus pyogenes*) clustered regularly interspaced short palindromic repeats (CRISPR) associated with a potential nuclear-localized endonuclease (Cas) and gRNA. In the CRISPR/Cas9 system, one chimeric gRNA molecule includes CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA), which drive Cas9 to cleave target sequences followed by at least 15-nucleotide (without mismatch at the 5′-end) spacer sequences complemented by target DNA. The Cas9 protein contains two DNA cleavage

domains, HNH and RuvC, that cleave DNA to form a double-stranded break (DSB), typically 3 bp upstream of the protospacer adjacent motif (PAM) on the target DNA, especially 5′-NGG or 5′-NAG sequences. DNA system repairs include homology-directed repair (HDR) (Rouet et al., 1994; Bibikova et al., 2003) and non-homologous end-joining (NHEJ) (Bibikova et al., 2002).

HDR often replaces nucleotides with one donor DNA template, including homologous sequences, while NHEJ repairs damaged DNA by inserting or deleting (indel) nucleotides in DSB regions. Recently, researchers have indicated the abilities of stable transformation procedures using CRISPR/Cas9 (Di Fan et al., 2015; Howells et al., 2018). In addition, some plant genomes, such as wheat (Wang et al., 2014), *Arabidopsis thaliana* (Feng et al., 2013), *Nicotiana benthamiana* (Nekrasov et al., 2013), sorghum (Jiang et al., 2013), maize (Char et al., 2017), and poplar (Di Fan et al., 2015; Zhou et al., 2015), have been edited by the CRISPR–Cas9 system. Recently, the authors have shown that a poplar genome could be edited precisely by CRISPR–Cas9 under optimized conditions (Movahedi et al., 2022). In this published paper, the authors revealed how they applied the optimized delivery conditions (presented in the current paper) to overcome poplar genome editing through precise nucleotide replacements using CRISPR–Cas9. Moreover, the poplar genome has shown that CRISPR/Cas9 is a highly efficient tool for generating homozygous mutations and stable transformants through the first generation (Di Fan et al., 2015). It has been reported that virus-based replicons and particle bombardment could increase this translocation regarding the difficulty of donor DNA pattern (DDP) delivery to the cell nucleus (Gil-Humanes et al., 2017). The expanding number of cells containing DDTs at S/G2 cell division phases is essential for enhancing HDR efficiency (Yang et al., 2016). Attempts have been made to develop *Agrobacterium* transformation efficiency and to reveal that Agrobacteria inoculator OD directly affects transformation efficiency (Movahedi et al., 2014). However, still, there is no statement on enhancing the Agrobacteria method transfer to increase the efficiency of carrying DDP and the recovery of DSBs as HDR. *Mitogen-activated protein kinases* (*MAPKs*) are specific transferase kinase enzymes for threonine and serine amino acids. Recently, scientists have proved that *MAPKs* direct cellular responses to heat shock and osmotic stresses (Mohanta et al., 2015).

MAPK kinase (MAPKK or MKK) is a dual-specific protein kinase and needs phosphorylation to activate the MAPK pathway. In *Arabidopsis*, *MAPKK2* (*MKK2*) are transcriptional regulator genes stimulated by environmental stresses such as salinity and cold to promote plant resilience (Teige et al., 2004). *MKK4* and *MKK5*, which are more closely linked, share over 75% of their amino acids and seem to stimulate *MPK3* and *MPK6* in response to stress (Asai et al., 2002). *MKK5* displays upregulated expressions in the *MKK2* overexpression lines in response to stress regarding the activation of *MPK3* and *MPK6* (Asai et al., 2002). Upregulation of *MKK5* in the *MKK2* overexpression lines indicates the existence of a crosstalk between these pathways (Teige et al., 2004). Therefore, we focused on *MKK5* in this study as the downstream regulation of *MKK2* to investigate the *MKK2* expression.

Poplar is a fast-growing woody species planted extensively worldwide with significant ecological and economic value. Broad poplar genomic resources are available for studying functional genomics and performing model tree analyses. According to Polle et al. (2013), functional genomics studies in woody plants are problematic because of low-efficiency genetic transformations, few mutants, and a long growth period. We tried to integrate an exogenous promoter 2XcaMV 35S into the predicted promoter zone of the *MKK2* gene in the poplar genome for the first time. Moreover, we optimized a stable Agrobacteria transformation method and Cas9 nuclease to build DSBs at the specific sites on the *MKK2* gene. We recruited an HDR repair system to analyze the expression changes of the desired gene against salt stress and prove the role of the *MKK2* gene in poplar resistance.

Materials and method

MKK2 locus identification, target detection, and line generation

Identification and target detection

The *MAPKK* 2-like (*MKK2*) gene from *Populus euphratica* (XM_011018250.1) was searched to find similar sequences in *Populus trichocarpa* (*P. trichocarpa*) (XM_002324230.2) using the BLAST database of the National Center for Biotechnology Information (NCBI) (https://blast.ncbi.nlm.nih.gov/). Blasted sequences were aligned to reveal similarities between the *MKK2* gene from 14 plant species that exhibited highly conserved domains. The comparative analysis revealed that *MAPK* is a highly conserved gene family through plant species (Liu, 2012; Jiang and Chu, 2018). The phylogenetic tree was calculated to prove the similarities when comparing far genome distances with 1,000 replicates through bootstrap analysis (Supplementary Figures 1, 2). The *MKK2* locus proximal promoter area was analyzed and confirmed by BDGP (Neural Network Promoter Prediction). Then, degenerate primers (F- and R-degeneracy; Supplementary Table 1) were designed using Geneious Prime® 2021 (Biomatters Ltd., New Zealand) to isolate consensus DNA sequences from *P. trichocarpa*, which were then ligated into the pEASY-TA cloning vector and verified by Sanger sequencing (GenScript, Nanjing, China). Verified sequences were submitted to NCBI as *Populus trichocarpa* MAP kinase kinase family protein under accession number MG871465.

Line generation

To generate mutated *mkk2* lines or negative control (NCT), we applied Geneious Prime® 2021 to analyze all the potential targets on the *MKK2* gene resulting in the detection of one conserved CRISPR target from both alleles of *MKK2* (chromosomes 6 and 18: NC_037290.1: 12,509,702–>12,513,458 and NC_037302.1: 4,954,681–>4,950,600), which was scored according to the on-target activity of 0.662 (Doench et al., 2014; Doench et al., 2016) and specificity score of 71.43% (Hsu et al., 2013) in blasting across the whole genome of *P. trichocarpa*. The detected CRISPR target was then applied to design oligos with *Bsa*I and ligate into the pGREB31 vector to construct pGREB-*mkk2* (*mkk2_F* and *mkk2_R*; Supplementary Table 1) (Xie et al., 2014).

To generate the *MKK2* overexpressed lines or positive control (PCT), we designed primers to PCR-amplified 2-kb genomic proximal promoter area from *P. trichocarpa* with *Bam*HI and *Kpn*I (Promo_F and Promo_R; Supplementary Table 1). The full-length CDS of *MKK2* was PCR-amplified from *P. trichocarpa* cDNA with *Bam*HI and *Pst*I to ligate with the proximal promoter and cloned into the pCAMBIA1300 vector to construct pCAM-*MKK2* (*MKK2*_full_F and *MKK2*_full_F; Supplementary Table 1).

To generate the *MKK2* edited lines (Ed), we analyzed the *MKK2* proximal promoter area on chromosome 18 to detect the potential CRISPR target, which was scored according to the on-target activity of 0.469 (Doench et al., 2014; Doench et al., 2016) and the specificity score of 83.33% (Hsu et al., 2013) in blasting across the whole genome of *P. trichocarpa* (Oligo-F and Oligo-R; Supplementary Table 1). A pair of oligos flanked by *Bsa*I adaptors were designed to anneal and phosphorylate for targeting. The synthesized oligos were ligated into digested pRGEB31 vectors by the *Bsa*I restriction enzyme to form pgRNA (Xie and Yang, 2013). All vectors were then transferred into *Escherichia coli* (DH5α) and propagated. Afterward, vectors were extracted using the plasmid midi kit (Qiagen, USA), and all ligated oligo sequences were confirmed by Sanger sequencing (GeneScript, Nanjing, China).

We transfected wild-type poplars to generate control lines (CT) using a pRGEB31 vector with no designed target gRNA scaffold (empty vector).

Transformation

Control line transformation (NCT, PCT, and CT)

Populus trichocarpa seedlings were grown in a phytotron at 23°C ± 2° C under a 16/8 light/dark period. All tissue culture and transformation techniques were performed according to a recently published paper (Movahedi et al., 2015). The plasmids pGREB-*mkk2* for NCT, pCAM-*MKK2* for PCT, and empty vector for CT were then transferred into *Agrobacterium tumefaciens*. The *A. tumefaciens*-triggered explants were co-cultivated in a dark room for 2 days. Afterward, explants were cultured into a selection of semi-solid woody plant medium (WPM) containing 0.05 mg/L of indole-3-butyric acid (IBA), 0.004 mg/L of thidiazuron (TDZ), 400 mg/L of cefotaxime, 8 mg/L of hygromycin, and 0.8% (w/v) agar. After 4 weeks, the regenerated buds were subcultured in media with 0.05 mg/L of IBA, 0.001 mg/L of TDZ, 200 mg/L of cefotaxime, 8 mg/L of hygromycin, and 0.8% (w/v) agar, to shoot. The 8-week-old prolonged shoots were then introduced into the media with 100 mg/L of cefotaxime and 8 mg/L of hygromycin and allowed to root. The rooted explants were then assumed to be successfully transformed. Approximately 80 ± 10 independent biological replicates were generated for each NCT, PCT, and CT line.

Edited line transformation

To generate Ed lines, *P. trichocarpa* seedlings were used as the template for the above transformation techniques. Three Agrobacteria concentrations (OD_{600} = 0.5, 1.5, and 2.5), three ratios of pDDT (plasmid including donor DNA template) and pgRNA (plasmid including target scaffold) (1:4, 1:1, and 4:1), and four different lengths of homologous arms to generate pDDT (500, 600, 700, and 800 bp) were selected as the variables in this study and classified into nine groups (Supplementary Table 2). Approximately 80 ± 10 independent biological replicates were generated for each variable (Supplementary Table 2). The 8-week-old prolonged shoots were then transferred into selective media with 200 mg/L of cefotaxime and 25 mg/L of kanamycin and allowed to root. Moreover, the CT lines were assumed as 0 bp of homologous arms and applied separately for each group as the control to subculture into selective media with 200 mg/L of cefotaxime and 25 mg/L of kanamycin. All groups were then analyzed to calculate and compare the transformation efficiency (Supplementary Tables 2, 3).

Verification of negative and positive lines and survival rates

To verify the NCT lines, we used *mkk2* mutants as a template resource for *Agrobacterium* transformation and then transferred pCAM-*MKK2* into *mkk2* explants regarding the above methods leading to reactive *MKK2* expression and generating mutant-overexpressions (M-OEs). Approximately 80 ± 5 M-OEs were approved for hygromycin screening. Eight-week-old recovered explants from the NCT, PCT, M-OEs, and CT lines were transferred to MS media supplemented with 0, 25, and 50 mM of NaCl (20 independent biological replicates for each NaCl concentration) and allowed to grow for 3 weeks. After that, we synthesized cDNAs from all transformed NCT, PCT, M-OEs, and CT lines and used the Applied Biosystems real-time PCR (Applied Biosystems, USA) to analyze the expression of the *MKK2* gene (*MKK2*_full_F and *MKK2*_full_R; Supplementary Table 1) via NaCl treatments for comparison and verification (three technical repeats for each biological replicate).

In addition, we allowed the NaCl-treated poplars to recover for 1 week by regulatory watering. Then, all 4-week-old transformed explants were investigated for survival rates [(Survived living explants: Total seedling transformed) × 100].

MKK5 expression analysis

To verify the NCT lines, we further analyzed the *MKK5* (chromosome 8; NC_037292.1) expression that was downstream-regulated by the *MKK2* gene using designed primers by Geneious Prime® (*MKK5*_F and *MKK5*_R; Supplementary Table 1) to isolate 1,071 bp from *MKK5* CDS. All transformed explants were introduced to different NaCl treatments. Then, the synthesized cDNAs from all transformed NCT, PCT, M-OEs, and CT lines were used for real-time PCR (three technical repeats for each biological replicate).

Plasmid design and cassettes

The 2XCaMV 35S promoter was isolated from the pCre vector (DQ645631) using 2X-F-pCre and 2X-R-pCre primers (Supplementary Table 1). The 2X-F primer was used to add 30 bp

homology sequences from the 5′ region of *nptII*, and the 2X-R primer was used to add 40 bp homology sequences from the 3′ region of the upstream homologous (UH) arm (Supplementary Table 1; Supplementary Figure 3A). The neomycin phosphotransferase (*nptII*) CDS was selected as a resistant gene against kanamycin and isolated from the PBI121 (AF485783) plasmid using *nptII*-F-PBI and *nptII*-R-PBI primers (Supplementary Table 1). The *nptII*-F primer was then used to add 40 bp homology sequences from the 5′ region of the downstream homologous (DH) arm (Supplementary Figure 3B). The PCR products of the overhung 2XCaMV 35S and *nptII* were then ligated using PCR to form a pre-cassette (Supplementary Figure 3C). To ligate products using PCR, we need a 30–40-bp overhang on one of the PCR products complementary to the end of the adjacent PCR product (using initial primers). It is necessary to allow the PCR products to anneal at a higher temperature to avoid non-specific hybridization through the long PCR products, which are non-complementary, and then fill up the PCR tubes with the final primers to make a double-stranded product. Therefore, three cycles of annealing (68°C) and extension (74°C) were arranged, and the desired primers were supplemented to the distal ends of the fragments using a regular PCR. The primers Back-F and Back-R (Supplementary Table 1) were then applied to form the cassette backbone and were confirmed by sequencing (Supplementary Figure 3D). To optimize the length of the homologous arms, nucleotides of different lengths from both sides of the selected target (500, 600, 700, and 800 bp) were used to construct cassettes including different homologous arms. The final constructed cassettes were cloned into the pGEM-T easy vector (Promega, New Zealand) for verification by sequencing. Briefly, the DH homologous arm was isolated from genomic DNA using 500-DH-F and 500-DH-R (Supplementary Table 1). To add the T-DNA left border, we used the previous PCR product as the template and amplified that again using T500-DH-F and 500-DH-R primers (Supplementary Table 1; Supplementary Figure 4A). The UH homologous arm also was isolated from genomic DNA using 500-UH-F and 500-UH-R (Supplementary Table 1; Supplementary Figure 4B). As described, two isolated homologous arms and the cassette backbone were then ligated by PCR to form the cassette (Supplementary Figure 4C). All lengths of the homologous arms were made with the same method as described and used appropriate primers (Supplementary Table 1). The verified cassettes were then ligated into the pRGEB31 plasmid using the restriction cloning technique to create pDDT (Supplementary Figure 4D).

Genomic DNA and total RNA extraction and cDNA synthesis

Genomic DNA was extracted from CT, NCT, and edited events grown on kanamycin using the CTAB method. Briefly, 80-100 µg of young leaves were quickly ground in liquid nitrogen, and 200 µl of 65° C preheated CTAB was added to the samples, followed by five inversions. Then, the samples were incubated at 65°C for 30 min, supplanted by adding 200 µl of chloroform, incubated at room temperature (RT) for 5 min, and centrifuged at 16,000*g* for 10 min at 4°C. The supernatant was then shifted to a new tube, mixed with 300 µl of isopropanol, and incubated on ice for 30 min. The precipitated genomic DNA was then separated via 16,000*g* centrifugation for 20 min at 4°C. The pellet was then washed twice with 500 µl of 70% ethanol. The semidried pellet was dissolved in 100 µl of double-distilled water and stored at 4°C. The combination of the obtained genomic DNA was measured using the BioDrop spectrophotometer at 1,300 ng/µl. Total RNA (100 ng/ml) was also extracted from the young transformant leaves grown on kanamycin and WT applying TRIzol. According to the manufacturer, we then adopted reverse transcription using total RNA and oligo-dT primers to synthesize the first cDNA strand (PrimeScript One-Step RT-PCR Kit Ver.2, Takara Biotechnology, Dalian, China).

Integration confirmation and off-target activity

To verify *nptII* integration and off-target activity incidents mediated by CRISPR genome editing, we used extracted genomic DNA from CT, NCT, and edited events (Ed_23_1, Ed_31_1, Ed_34_1, Ed_35_1, Ed_35_2, Ed_35_3, and Ed_36_1) to amplify 1,457 bp of a PCR fragment from *nptII* as the reverse primer and nucleotides out of the 3′ homologous arm (downstream homologous) on the second intron as the forward primer (5,552 F, 7,008 R; Supplementary Table 1). ImageJ ver.2 was applied to measure on- and off-target intensities.

TaqMan real-time PCR for HDR efficiency

To test the proper direction of combining exogenous *nptII*, we used extracted genomic DNA as the template to run the TaqMan assay applying dye labels such as FAM and VIC using Applied Biosystems real-time PCR (Applied Biosystems, USA) (Movahedi et al., 2022). The Ed_23_1, Ed_31_1, Ed_34_1, Ed_35_1, Ed_35_2, Ed_35_3, and Ed_36_1 recovered events were used in this method. Thus, we designed primers to probe three 90-bp fragments: FAM1, 2, and 3 (Supplementary Table 1). These primers were planned to probe 52 bp of downstream homologous and 38 bp from the 3′ end of *nptII* (FAM1), 44 bp of *nptII* and 46 bp of 2XCamV 35 (FAM2), and 37 bp of 2XCamV 35 and 53 bp of upstream homologous (FAM3). In addition, the primers probed one 106 bp fragment VIC on the *β-actin* gene as the reference with a stable copy number (Supplementary Table 1, 1117-F and 1222-R). All samples were analyzed in quadruplicates. TaqMan uses a fluorogenic probe to bind 20 single-stranded nucleotides only between forward and reverse primers. Therefore, only completed strands could be assigned by this probe and generate the fluorescent signals in TaqMan real-time PCR. The signals detected from each FAM were compared to VIC signals to calculate $\Delta\Delta C_t$ and HDR efficiency (%).

Polymorphisms

NHEJ is characterized by introducing irregular small indels into the targeted site. However, regardless of this mutagenic potential

and its propensity for error, NHEJ plays an active role in repairing genome integrity and suppressing chromosomal translocations and in the bulk of repair events in the genome (Ceccaldi et al., 2016).

We analyzed the variant genotypes within UH, DH, and knocked-in fragments from recovered events resulting from CRISPR activity to test the effect of HDR promotions resulting from optimized transformation on polymorphisms. Five polymorphism varieties were detected, including deletions, insertions, SNP transitions (A to C or G to T and vice versa), SNP transversions (purines to pyrimidines or vice versa), and substitutions. All genomic DNA extracted from the recovered events was then isolated by the designed primers (Seq_F and Seq_R; Supplementary Table 1), cloned into a pEASY-TA cloning vector, and sequenced (GeneScript, Nanjing, China) to detect polymorphisms that occur using Geneious Prime 2022.

Phenotypic properties

It has been shown that *MAPK* genes direct cellular responses against abiotic stresses such as salinity (Teige et al., 2004; Mohanta et al., 2015; Sun et al., 2015) and regulate cell growth and death, differentiation, and the cell cycle (Jonak, 2002; Yang et al., 2017). Regarding these findings, we measured chlorophyll and carotenoid contents and root length before and after salt stress to prove the proper HDR and evaluate the effect of 2XCaMV 35S on *MKK2* expression among CT, NCT, and recovered events.

Statistical analysis

All data were analyzed using one-way ANOVA with Tukey's *post-hoc* comparisons calculated by GraphPad Prism 9.3 (GraphPad Software, LLC). Differences were analyzed when the confidence intervals presented no overlap of the mean values with an error value of 0.05.

Results

Mutant lines were confirmed via salt treatments

The *mkk2* mutant lines (NCT) were investigated for lacking *MKK2* expression (%) through NaCl treatments and in comparison with the CT lines (Figure 1A). While the NCT lines revealed no *MKK2* expression, the CT and PCT lines indicated incremental expressions via increasing salt treatments. The M_OE lines only revealed increased *MKK2* expression (%) in the 50-mM NaCl concentration. We analyzed survival rates (%) to further confirm the proper *mkk2* mutants compared with the CT lines (Figure 1B). While the CT, PCT, and M_OE lines revealed acceptable survival rates (%) among various NaCl treatments, the NCT lines were negatively affected by the high salt concentration of 50 mM. It has been proved that the expressed *MKK2* activated by cold and salt stresses upregulates the *MKK5* expressions (Teige et al., 2004).

Regarding these findings, the lack of *MKK2* expression causes the downregulation of *MKK5* under salt stress. We analyzed the *MKK5* expressions affected by *MKK2* silencing to further verify the proper *mkk2* mutant lines compared with the CT lines (Figure 1C). The results revealed significant incremental *MKK5* expression in the CT lines compared with the NCT lines under 25 and 50 mM of NaCl concentrations. The M_OE and PCT lines also revealed acceptable increased *MKK5* expression under 25 and 50 mM of NaCl concentrations.

The efficient transformation resulted in the generation of recovered events

A schematic diagram of the designed experiments showed the target site, which was in the promoter zone of the *MKK2* locus (Figure 1D). This target was replaced by different homologous arms associated with the *nptII* CDS and 2XCamV 35S promoter among different ratios of pDDT/pgRNA (Figure 1D). The arranged groups, including different variables (Supplementary Table 2), revealed various regenerated edited events (Eds) on the screen media supplemented by hygromycin. Most regenerated Eds were observed through groups 6, 8, and 9 with Agrobacteria inoculators OD_{600} = 2.5, 1.5, and 2.5, respectively. In addition, groups 8 and 9 exhibited more regenerated Eds than group 6, with a pDDT/pgRNA ratio of 4:1 (Supplementary Table 2; Supplementary Figure 5). These results revealed recovered events (rooted in supplemented media with kanamycin) only in groups 6, 8, and 9 (Supplementary Figure 5). In addition, only recovered Eds with 700 bp homologous arms (group 9) showed significant differences compared with the CT lines (Supplementary Figure 5). The comparisons of all the lengths of homologous arms through groups 6, 8, and 9 revealed that Eds with 700 bp homologous arms were significantly more than the other events with 0 (CT), 500, 600, and 800 bp (Figure 1E). Concerning all variables and Eds, group 9 revealed significantly more transformation efficiency (%) (recovered Eds) than the other groups (Figure 1F; Supplementary Table 3). In total, the results showed that 700 bp of homologous arms were associated with high concentrations of Agrobacteria inoculation (OD_{600} = 2.5) and pDDT/pgRNA ratio of 4:1, significantly improving transformation efficiency (%) to recover edited events.

HDR was confirmed by recombinant mRNA

The supposed recombinant mRNA resulting from successful HDR was assessed using *MKK2* and *nptII* fusion (Figure 2A). Regarding this hypothesis, *MKK2* and *nptII* mRNAs should be fused to each other and tracked by one shared probe (Figure 2A). Designed primers (FUS_F and FUS_R; Supplementary Table 1) were applied to detect the expression of this fusion by real-time PCR from synthesized cDNA. All recovered Eds revealed fusion expression compared with the CT lines (Figure 2B). Regarding these results, Ed_35_1, Ed_35_2, and Ed_35_3 exhibited significant fusion expressions compared with the other Eds. To further prove

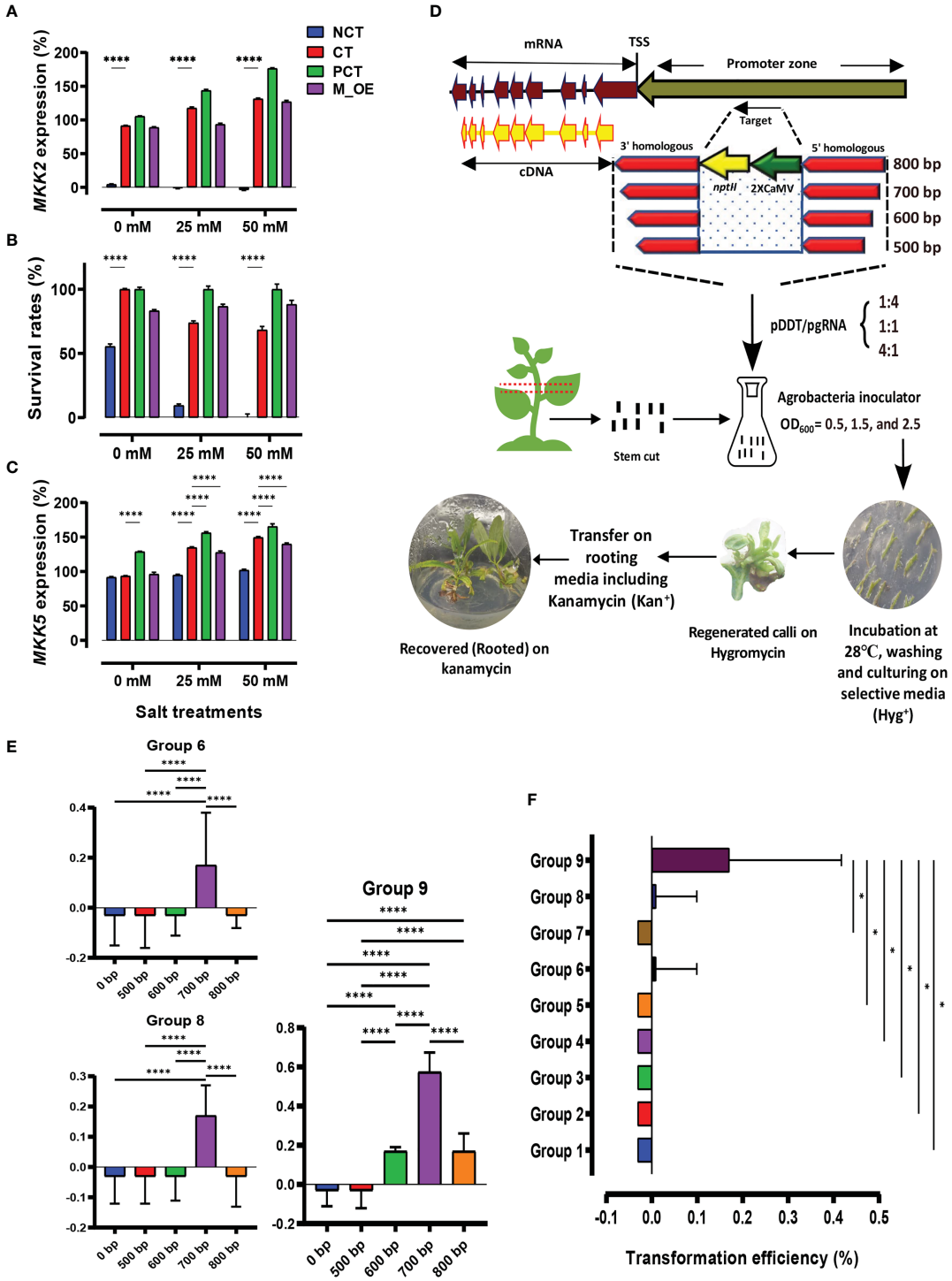


FIGURE 1
Graphical summary shows the designed experiments for transformation assessment and homology-directed repair (HDR) efficiency. **(A–C)** Control (CT) and negative control (NCT) verification among *MKK2*, survival treatments, and *MKK5* expression. **(D)** Different homologous arm lengths, Agrobacteria inoculators, and the proportion of pDDT/pgRNA caused efficient transformation screened by media-supplemented kanamycin. Transformed explants were regenerated on hygromycin-selective media and rooted in kanamycin-selective media. **(E)** Mean comparison of recovered events influenced by different homologous arms. **(F)** Transformation efficiency (%) evaluation through designed groups; error bars represent SE; asterisks represent *p*-values as * ≤0.05, and **** ≤0.0001.

proper HDR through recovered Eds, the complete CDS of *MKK2* and *nptII* was real-time PCR-amplified using designed primers (MKK2_full_F and MKK2_full_R and nptII_F and nptII_R; Supplementary Table 1) to evaluate their expressions (Figure 2C). All the recovered Eds showed *MKK2* expression (%) more than the CT lines affected by the precise integration of the 2XCaMV 35S promoter under 0 mM of NaCl treatment (Figure 2D). Ed_35_2 and Ed_35_3 exhibited significant expressions compared with the CT

lines. Increasing the NaCl concentration to 25 mM increased *MKK2* expression, in which Ed_35_1, Ed_35_2, and Ed_35_3 exhibited significant enhanced *MKK2* expression against the CT lines (Figure 2D). Moreover, the *nptII* expression analyses in 0 and 25 mM of NaCl proved its proper expression in recovered Eds compared with the CT lines with no expression (Figure 2D). The downstream *MKK5* regulated by *MKK2* expression was further analyzed under 0 and 25 mM of NaCl treatment (Figure 2D). The results revealed significant *MKK5* expression in Ed_35_1, Ed_35_2, and Ed_35_3 compared with the CT lines affected by more *MKK2* expression by integrating 2XCamV 35S under 0 mM of NaCl concentration. Increasing salt stress to 25 mM resulted in an enhanced *MKK5* expression through edited events compared with the CT lines. NCT lines revealed decreased *MKK5* expression through increasing NaCl concentrations, which was influenced by the lack of *MKK2* expression in *mkk2* mutant lines. Concerning the

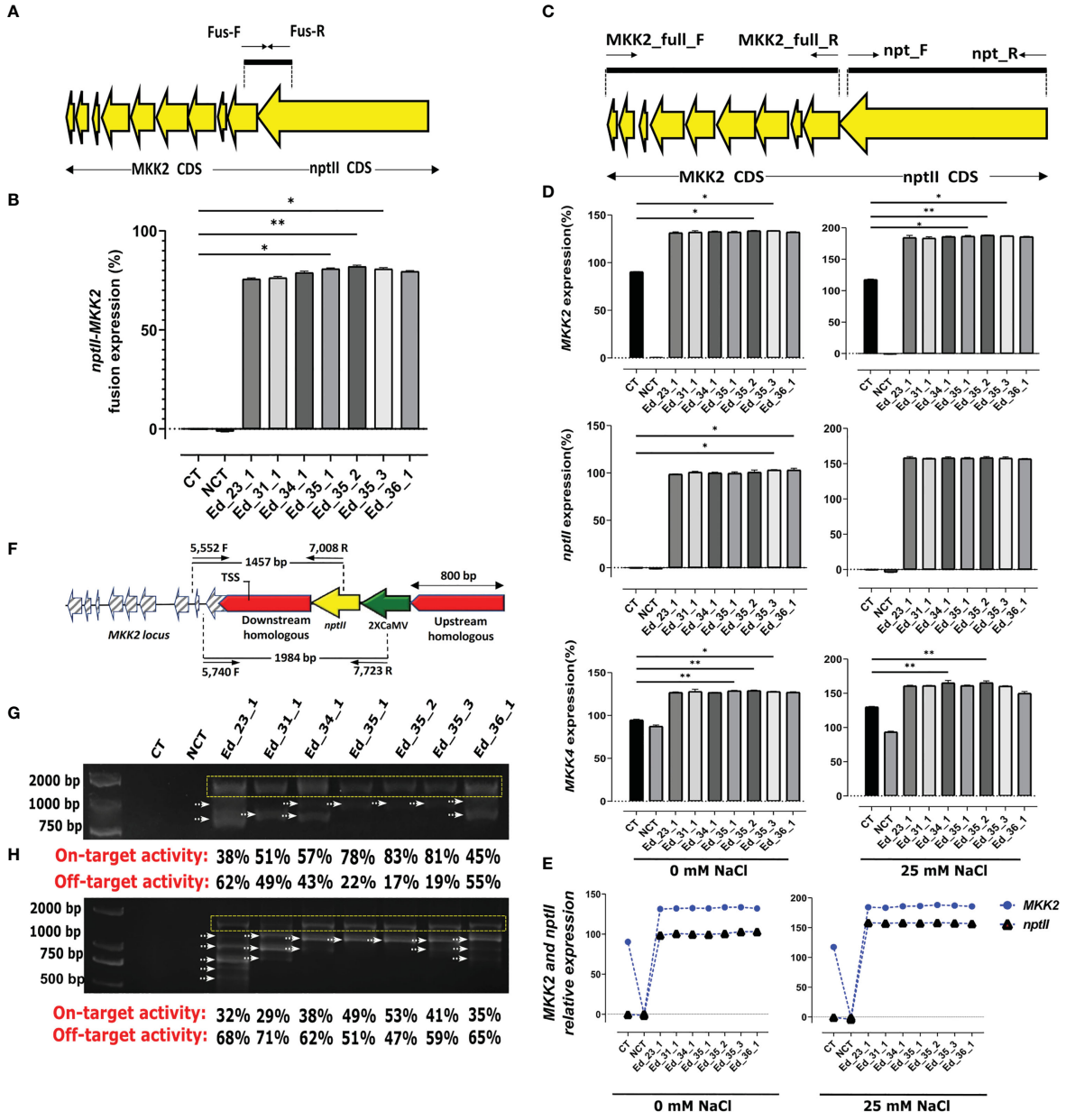


FIGURE 2
Recovered edited events exhibited nptII–MKK2 fusion and proper exogenous and exopromoter integration. **(A)** Schematic of *nptII–MKK2* mRNA fusion. **(B)** nptII and MKK2 fusion mRNA revealed more expression via recovered events, especially Ed_35_1, Ed_35_2, and Ed_35_3, compared with the CT line. **(C)** Schematic of complete CDS of nptII and *MKK2* through proper HDR. **(D)** The recovered edited events showed *MKK2*, *nptII*, and the downstream *MKK5* more expressions than the CT line under 0 and 25 mM of NaCl treatments. **(E)** *nptII* and *MKK2* exhibited relative expressions under 0 and 25 mM of NaCl treatments. **(F)** Schematic of properly integrating *nptII* and 2XCaMV 35S into *MKK2* promoter zone and overexpressed *MKK2* driven by 2XCaMV 35S. **(G)** PCR-amplified verification of *nptII* proper integration and on- and off-target evaluation. **(H)** PCR-amplified validation of 2XCaMV 35S proper integration and on- and off-target evaluation. Error bars represent SE; asterisks represent *p*-values as *≤ 0.05 and ** ≤0.01.

MKK2 and *nptII* fusion mRNA, we evaluated the relative expressions among the recovered Eds and under 0 and 25 mM of salt stress. We found constant relative expressions between endogenous *MKK2* and integrated exogenous *nptII*, resulting in the exact HDR happening through the recovered Eds (Figure 2E).

Recombinant genomic DNA has confirmed on-target activity

Specific primers were designed from the *nptII* and *MKK2* locus only to amplify the 1,457-bp fragment from successful HDR in recovered edited events (Figure 2F). To evaluate the recombinant genome, we used extracted genomic DNA from all recovered events to amplify PCR (5,552F and 7,008R; Supplementary Table 1) (Figure 2F). The recovered Eds revealed exactly 1,457-bp bonds, resulting in the precise *nptII* and *MKK2* locus integrations. Furthermore, several unexpected bonds were attended as off-target happenings. The specific and unspecific bonds were then analyzed to measure their intensities and determine on- and off-target activities through the recovered Eds (Figure 2G; Supplementary Figure 6A). To further verify the precise integration of 2XCamV 35S with the *MKK2* locus, specifically designed primers (5,740F and 7,723R; Supplementary Table 1) were applied to PCR amplify several nucleotides from 2XCamV 35S to intron 2 of the *MKK2* locus (Figure 2F). All the recovered Eds revealed 1,984-bp-specific bonds (Figure 2H; Supplementary Figure 6B). In addition, Ed_35_1, Ed_35_2, and Ed_35_3 showed lesser off-target activities but more on-target activities than the other recovered Eds.

TaqMan real-time PCR revealed efficient HDR

We applied FAM1, 2, and 3 to evaluate HDR efficiency (Figure 3A). FAM1 was primed (6,593F and 6,682R; Supplementary Table 1) to estimate HDR efficiency through the downstream homologous arm and *nptII*. FAM1 $\Delta\Delta C_t$ analyses revealed significant homology nucleotides in Ed_35_2 and Ed_35_3 compared with the CT and NCT lines (Figure 3B). Sanger sequencing among FAM1 further exhibited several mutant nucleotides within Ed_23_1, Ed_34_1, Ed_35_1, Ed_35_3, and Ed_36_1, except Ed_35_2 with no mutant nucleotides (Figure 3C). To estimate HDR efficiency between *nptII* and 2XCamV 35S, FAM2 was designed to prime (7,395F and 7,484; Supplementary Table 1). FAM2 $\Delta\Delta C_t$ analyses revealed homology nucleotides, resulting in sufficient FAM2 signals from all the recovered Eds (Figure 3B). Only Ed_35_2 showed significantly more signals compared with the CT and NCT lines. In addition, Sanger sequencing confirmed FAM2-sufficient homology nucleotides with no mutation (Figure 3C). To evaluate the upstream homologous arm and 2XCamV 35S homology nucleotides, FAM3 was primed (8,183F and 8,272R; Supplementary Table 1). While FAM3 $\Delta\Delta C_t$ analyses exhibited significant HDR efficiency through Ed_35_2 compared with the CT and NCT lines, Ed_35_1 and Ed_35_3 exhibited significantly more FAM3 signals than the NCT lines only (Figure 3B). While Sanger sequencing further confirmed FAM3 $\Delta\Delta C_t$ to show several mutant nucleotides through Ed_23_1, Ed_31_1, Ed_34_1, and Ed_36_1, Ed_35_1, Ed_35_2, and Ed_35_3 exhibited no mutant nucleotides (Figure 3C). HDR efficiency (%) was then evaluated by calculating all achieved FAM signals from the recovered Eds except for the CT and NCT lines with no HDR efficiency (Figure 3D; Supplementary Table 4). While Ed_31_1, Ed_34_1, and Ed_36_1 exhibited acceptable HDR efficiency compared with Ed_23_1 with the lowest efficient HDR, Ed_35_1, Ed_35_2, and Ed_35_3 exhibited more significant efficient HDR (Figure 3D).

Efficient transformed edited events revealed minor polymorphisms

Total polymorphism analyses exhibited significantly lesser polymorphisms through the Ed_35_2 event than Ed_23_1 and Ed_36_1 with the most polymorphisms (Figure 3E; Supplementary Table 5). Moreover, Ed_35_1 and Ed_35_2 exhibited lower polymorphisms than Ed_23_1, Ed_31_1, Ed_34_1, and Ed_36_1 (Figure 3E). More detailed analyses revealed that maximum deletions occurred in Ed_23_1 genome editing, and minimum deletions happened in Ed_35_1, Ed_35_2, and Ed_35_3 (Figure 3F; Supplementary Figure 7). Ed_35_1, Ed_35_2, and Ed_35_3 revealed lower insertions through HDR happenings, while Ed_23_1, Ed_31_1, and Ed_36_1 exhibited higher insertion nucleotides (Supplementary Figure 7; Supplementary Table 6).

Biochemical and phenotypic assessments proved efficient HDR through efficient transformants

Since *MKK2* is a positive regulator of poplar defense against environmental stresses, it is supposed that targeted *MKK2* over NHEJ may decrease plant resistance to salinity. Targeted *MKK2* over the HDR system to insert the exogenous 2XCaMV 35S promoter produced overexpressed *MKK2* (Figure 2D) compared with the CT and NCT lines and may promote resistance to salt stress. Regarding the direct positive effect of *MKK2* expression on phytohormones such as chlorophyll and carotenoids (Mao et al., 2017) and on plant growth and development (Nishihama et al., 1995), we analyzed all the recovered Eds for the contents of these phytohormones and further root development under salt treatments.

Total chlorophyll was detected in higher contents through all the recovered Eds than in the CT and NCT lines under 0 mM of NaCl concentration (Figure 3G). NaCl (25 mM) increased chlorophyll contents affected by overexpressed *MKK2* through 2XCamV 35S integration among the recovered Eds, but Ed_35_2 and Ed_35_3 exhibited significantly more contents than the NCT line (Figure 3G). To confirm the effect of enhanced *MKK2* expression resulting from the precise integration of HDR, we

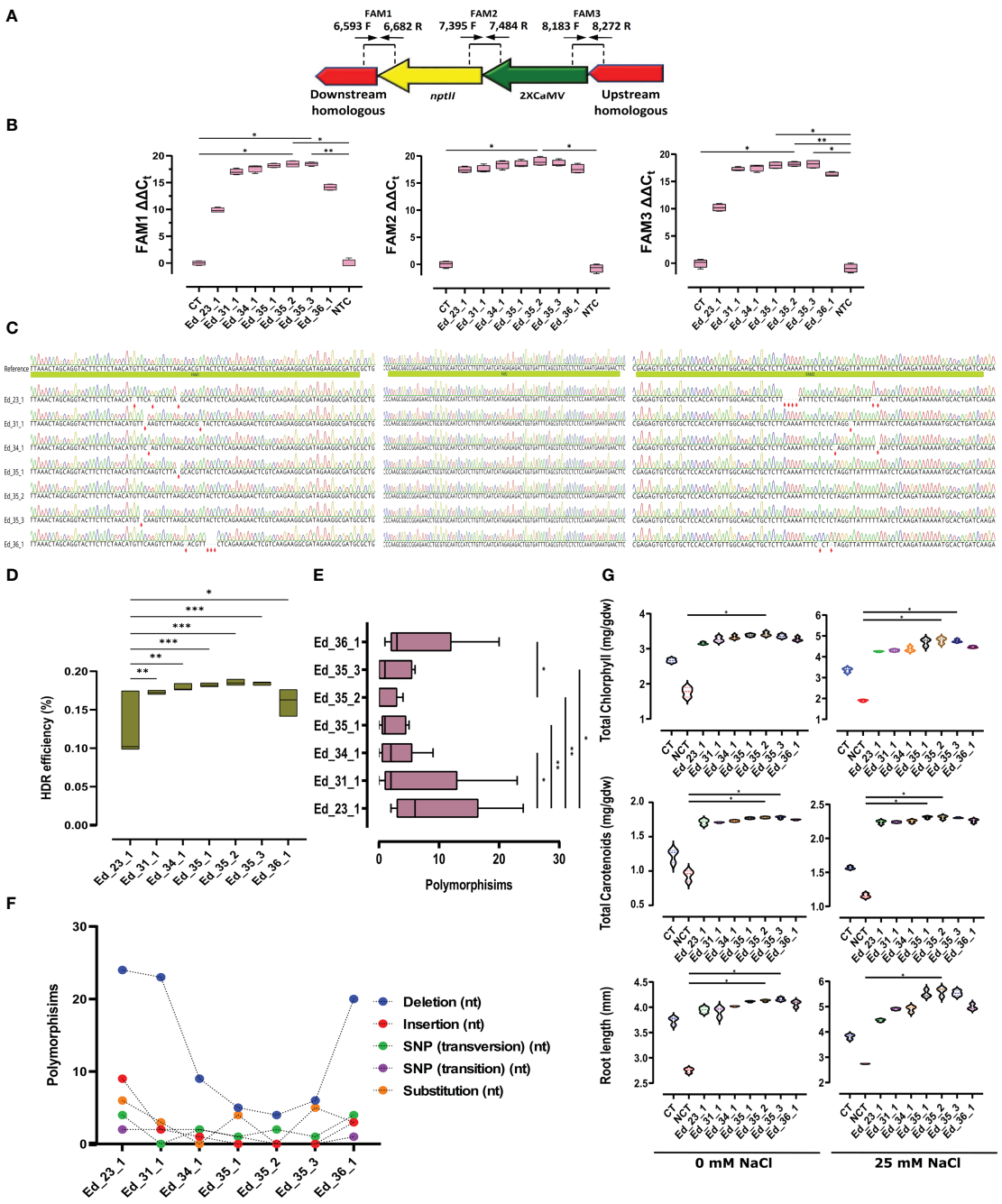


FIGURE 3
TaqMan real-time PCR exhibited efficient HDR through efficient transformant events. **(A)** Schematic FAM design to evaluate HDR efficiency. **(B)** Ed_35_1, Ed_35_2, and Ed_35_3 exhibited more FAM signals than the CT and NCT lines. Especially Ed_35_2 revealed significantly more expression signals within FAM1, 2, and 3 than the recovered edited events. **(C)** Sanger sequencing to verify results achieved by FAM signals. **(D)** HDR efficiency (%) evaluation calculated from the achieved FAM signals. Ed_35_1, Ed_35_2, and Ed_35_3 revealed more HDR efficiency. **(E)** Mean comparison of polymorphisms that happened through recovered edited events. **(F)** Polymorphisms exhibited more insertions and deletions among Eds except for Ed_35_1, Ed_35_2, and Ed_35_3 events. **(G)** Total chlorophyll and carotenoids calculated from the recovered edited events exhibited more through Ed_35_1, Ed_35_2, and Ed_35_3 than the other Eds under 0 and 25 mM of NaCl treatments. In addition, root length measurement confirmed more growth among Ed_35_1, Ed_35_2, and Ed_35_3 than the other Eds under 0 and 25 mM of NaCl treatments. Error bars represent SE; asterisks represent *p*-values as * ≤0.05, ** ≤0.01, and *** ≤0.001.

further analyzed carotenoid contents and found similar results of chlorophyll under 0 and 25 mM of salt stress (Figure 3G). The higher 25-mM NaCl concentration caused carotenoids to increase slightly in the CT and NCT lines but significantly increased carotenoids among Ed_35_1 and Ed_35_2. Finally, all the recovered Eds exhibited enhanced growth in root lengths under 0 mM of NaCl treatment. A high NaCl concentration of 25 mM revealed an acceptable enhancement in root length in Ed_35_1 and Ed_35_3 but caused significantly more root growth in Ed_35_2 compared with the NCT line (Figure 3G).

Discussion

MPK pathways are involved in different physiological reactions and are conserved genetically in eukaryotes (Davis, 2000). Activating MPK pathways in response to stressors promotes plant stress tolerance (Sinha et al., 2011). In reaction to stress, *MKK4* and *MKK5* overexpression has been shown to promote *MPK3* and *MPK6* expression (Asai et al., 2002). MPK3 and MPK6 effectively phosphorylated IAA15 at the Ser-2 and Thr-28 residues (Kim et al., 2022), decreasing lateral root growth due to increased IAA15 accumulation, leading to inhibit the formation of lateral roots in response to drought. Moreover, increased *MKK5* expression caused by *MKK2* overexpression has been shown to be effective in triggering *MPK3* and *MPK6* in response to stresses (Asai et al., 2002), resulting in the significance of *MKK2* as a transcriptional regulator gene that makes plants more resistant to salt and cold (Teige et al., 2004).

CRISPR–Cas9 has been used for biotechnology, functional biology, and genetic medicine. *Agrobacterium rhizogenes* and *A. tumefaciens* have been used to transfer CRISPR-gRNA sites into plant cells (Jia and Wang, 2014; Di Fan et al., 2015). This research revealed that increasing OD_{600} to 2.5 improved HDR efficiency, while groups 6 and 9 revealed recovered events. In addition, OD_{600} = 0.5 exhibited no recovered events regardless of the pDDT/pgRNA ratio. Increasing OD_{600} to 1.5 caused recovery events within group 8. The other variable was the number of DDTs in the S/G2 cell division phases. According to Yang et al. (2016), the number of DDTs is critical in enhancing HDR efficiency. Our analyses exhibited more recovered events via a pDDT/pgRNA ratio of 4:1 between groups 8 and 9. This ratio caused to increase pDDT dramatically for efficient HDR. The homologous arm lengths were evaluated as the third variable in this study. It has been indicated that the lengths of the homologous arms are vital to making efficient HDR. Our experiments evaluated the homologous arm lengths that exhibited optimized 700 bp compared with the other lengths (Figure 1E) (Song and Stieger, 2017). Group 9, with a pDDT/pgRNA ratio of 4:1, Agrobacterium inoculator with OD_{600} = 2.5, and optimized 700 bp homologous arm length, revealed significantly more recovered events than the other groups (Figure 1E; Supplementary Figure 5). Regarding these findings, Eds_35 transformed with OD_{600} = 2.5, pDDT/pgRNA ratio of 4:1, and 700 bp homologous arm length exhibited a significantly higher expression integrating *nptII* and *MKK2* fusion mRNA expression (Figure 2B). Moreover, Eds_35 exhibited the expressions of *MKK2* and its downstream *MKK5* affected by the integrated 2XCamV 35S (Figure 2D). Predictably, the fused *MKK2* and exogenous *nptII* mRNA exhibited a relative expression, resulting in an efficient HDR (Figure 2E). TaqMan real-time PCR proved that the most HDR efficiency occurred in Ed_35_1, Ed_35_2, and Ed_35_3 (Figure 3D). Efficient HDR happened through Ed_35 events causing lower or non-mutant nucleotides leading to significantly decreased polymorphisms in HDR efficiency. The other recovered events from groups 6 and 8 exhibited more polymorphisms (Figure 3E).

On the other hand, there were minimum deletion and no insertion mutations through Eds_35 (Figure 3F). Further analysis of biochemical and phenotypic changes revealed significantly more total chlorophyll and carotenoids in the Ed_35 recovered events than the other recovered events under 0 and 25 mM of NaCl treatments (Figure 3G). The investigation of root length also revealed significantly more growth within Eds_35 than the other recovered edited events under salt stress. Altogether, efficient HDR happened through an efficient transformation with a high concentration of *Agrobacterium* inoculator OD_{600} = 2.5, an increased number of DDTs with a pDDT/pgRNA ratio of 4:1, and a further optimized 700 bp homologous arm length. We concluded that efficient transformation significantly directs efficient HDR, decreasing polymorphisms.

Conclusion

We have shown that a high concentration of *Agrobacterium* inoculator, homologous arm length optimization, and increasing the number of DDTs via cell division lead to efficient HDR through poplar genome editing. We also proved that a significant reduction in CRISPR-induced polymorphisms could be achieved by following these guidelines. This breakthrough technology will probably encourage biotechnological research, breeding programs, and conservation of tree species and developed crops.

Data availability statement

All data supporting the findings of this study are available in the article and its Supplementary Figures and Tables. Raw Sanger sequencing data are available on the Mendeley dataset with DOI: 10.17632/5tp44bnczy.3.

Author contributions

AM: conceptualization, software, formal analysis, writing—original draft, visualization, and project administration. HW: methodology, formal analysis, writing—review and editing, and data curation. SK: validation and writing—review and editing. WS, QZ, and LY: formal analysis and visualization. CX: formal analysis, visualization, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Nanjing Key Laboratory of Quality and Safety of Agricultural Products (NJGS2021-16), Technological Innovation Team of Colleges and Universities of Jiangsu Province (SUJIAOKE 2021), and Key Subject of Ecology of Jiangsu Province (SUJIAOYANHAN 2022).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1159615/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Alignment of Mitogen-activated protein kinase kinase 2 from 14 species to reveal similarities and identify desired sequences in *P. trichocarpa* using Geneious Prime 2022. *Hevea brasiliensis* - XM_021794488.1, *Manihot esculenta* - XR_002487451.1, *Jatropha curcas* - XM_020680666.1, *Populus euphratica* - XM_011018250.1, *Populus trichocarpa* - XM_002324230.2, *Nicotiana tomentosiformis* - XM_009605449.2, *Durio zibethinus* - XM_022899101.1, *Gossypium arboreum* - XM_017790462.1, *Gossypium hirsutum* - NM_001327410.1, *Ziziphus jujuba* - XM_016033841.1, *Juglans regia* - XM_018990646.1, *Betula platyphylla* - KJ459948.1, *Eucalyptus grandis* - XM_010049224.2, *Citrus sinensis* - XM_006475450.2.

SUPPLEMENTARY FIGURE 2
Phylogenetic tree of Mitogen-activated protein kinase kinase 2; Bootstrap analysis was performed using 1,000 replicates to evaluate the reliability of the various phylogenetic groups.

SUPPLEMENTARY FIGURE 3
The schematic figures of cassette backbone preparation

SUPPLEMENTARY FIGURE 4
The schematic figures of homologous arms and pDDT preparation for 500 bp homology sequences

SUPPLEMENTARY FIGURE 5
Mean comparison of regenerated and recovered events through different designed groups.

SUPPLEMENTARY FIGURE 6
The raw document used in this research. **(A)** 1457 bp of recombinant genomic DNA has been amplified to prove the proper HDR happenings through recovered events isolated from nptII and *MKK2* locus intron 2. **(B)** 1984 bp of amplified fragments selected from 2XCaMV35S and *MKK2* locus intron 1. On-target and off-target activities are shown as the specific and unspecific bonds, respectively; CT was used as the control. NCT was used as the negative control. Yellow dashed lines indicate cut gels presented in .

SUPPLEMENTARY FIGURE 7
Sanger sequences, alignment, and sequence analysis from recovered edited events reveal polymorphisms.

References

Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W.-L., Gomez-Gomez, L., et al. (2002). MAP kinase signalling cascade in arabidopsis innate immunity. *Nature* 415, 977–983. doi: 10.1038/415977a

Bibikova, M., Beumer, K., Trautman, J. K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. *Science* 300, 764. doi: 10.1126/science.1079512

Bibikova, M., Golic, M., Golic, K. G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in drosophila using zinc-finger nucleases. *Genetics* 161, 1169–1175. doi: 10.1093/genetics/161.3.1169

Ceccaldi, R., Rondinelli, B., and D'Andrea, A. D. (2016). Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* 26, 52–64. doi: 10.1016/j.tcb.2015.07.009

Char, S. N., Neelakandan, A. K., Nahampun, H., Frame, B., Main, M., Spalding, M. H., et al. (2017). An agrobacterium -delivered CRISPR /Cas9 system for high-frequency targeted mutagenesis in maize. *Plant Biotechnol. J.* 15, 257–268. doi: 10.1111/pbi.12611

Davis, R. J. (2000). "Signal transduction by the JNK group of MAP kinases," in *Inflammatory processes: Molecular mechanisms and therapeutic opportunities*. Eds. L. G. Letts and D. W. Morgan (Basel: Birkhäuser Basel), 13–21.

Di , Fan, Liu, T, Li, C., Jiao, B., Li, S., Hou, Y., et al. (2015). Efficient CRISPR/Cas9-mediated targeted mutagenesis in populus in the first generation. *Sci. Rep.* 5, 12217. doi: 10.1038/srep12217

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi: 10.1038/nbt.3437

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., et al. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267. doi: 10.1038/nbt.3026

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., et al. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res.* 23, 1229–1232. doi: 10.1038/cr.2013.114

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., et al. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR /Cas9. *Plant J.* 89, 1251–1262. doi: 10.1111/tpj.13446

Howells, R. M., Craze, M., Bowden, S., and Wallington, E. J. (2018). Efficient generation of stable, heritable gene edits in wheat using CRISPR/Cas9. *BMC Plant Biol.* 18. doi: 10.1186/s12870-018-1433-z

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., et al. (2013). DNA Targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832. doi: 10.1038/nbt.2647

Jia, H., and Wang, N. (2014). Targeted genome editing of sweet orange using Cas9/sgRNA. *PloS One* 9, e93806. doi: 10.1371/journal.pone.0093806

Jiang, M., and Chu, Z. (2018). Comparative analysis of plant MKK gene family reveals novel expansion mechanism of the members and sheds new light on functional conservation. *BMC Genomics* 19, 407. doi: 10.1186/s12864-018-4793-8

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D. P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 41, e188–e188. doi: 10.1093/nar/gkt780

Jonak, C. (2002). Complexity, cross talk and integration of plant MAP kinase signalling. *Curr. Opin. Plant Biol.* 5, 415–424. doi: 10.1016/S1369-5266(02)00285-6

Kim, S. H., Bahk, S., Nguyen, N. T., Le Pham, M. A., Kadam, U. S., Hong, J. C., et al. (2022). Phosphorylation of the auxin signaling transcriptional repressor IAA15 by MPKs is required for the suppression of root development under drought stress in arabidopsis. *Nucleic Acids Res.* 50, 10544–10561. doi: 10.1093/nar/gkac798

Liu, Y. (2012). Roles of mitogen-activated protein kinase cascades in ABA signaling. *Plant Cell Rep.* 31, 1–12. doi: 10.1007/s00299-011-1130-y

Mao, J., Li, W., Mi, B., Dawuda, M. M., Calderón-Urrea, A., Ma, Z., et al. (2017). Different exogenous sugars affect the hormone signal pathway and sugar metabolism in "Red globe" (Vitis vinifera l.) plantlets grown *in vitro* as shown by transcriptomic analysis. *Planta* 246, 537–552. doi: 10.1007/s00425-017-2712-x

Mohanta, T. K., Arora, P. K., Mohanta, N., Parida, P., and Bae, H. (2015). Identification of new members of the MAPK gene family in plants shows diverse conserved domains and novel activation loop variants. *BMC Genomics* 16. doi: 10.1186/s12864-015-1244-7

Movahedi, A., Wei, H., Zhou, X., Fountain, J. C., Chen, Z.-H., Mu, Z., et al. (2022). Precise exogenous insertion and sequence replacements in poplar by simultaneous

HDR overexpression and NHEJ suppression using CRISPR-Cas9. *Horticulture Res.* 9, uhac153. doi: 10.1093/hr/uhac154

Movahedi, A., Zhang, J., Amirian, R., and Zhuge, Q. (2014). An efficient agrobacterium-mediated transformation system for poplar. *IJMS* 15, 10780–10793. doi: 10.3390/ijms150610780

Movahedi, A., Zhang, J., Gao, P., Yang, Y., Wang, L., Yin, T., et al. (2015). Expression of the chickpea CarNAC3 gene enhances salinity and drought tolerance in transgenic poplars. *Plant Cell Tiss Organ Cult* 120, 141–154. doi: 10.1007/s11240-014-0588-z

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 691–693. doi: 10.1038/nbt.2655

Nishihama, R., Banno, H., Shibata, W., Hirano, K., Nakashima, M., Usami, S., etal. (1995). Plant homologues of components of MAPK (mitogen-activated protein kinase) signal pathways in yeast and animal cells. *Plant Cell Physiol.* 36, 749–757. doi: 10.1093/oxfordjournals.pcp.a078818

Polle, A., Janz, D., Teichmann, T., and Lipka, V. (2013). Poplar genetic engineering: promoting desirable wood characteristics and pest resistance. *Appl. Microbiol. Biotechnol.* 97, 5669–5679. doi: 10.1007/s00253-013-4940-8

Rouet, P., Smih, F., and Jasin, M. (1994). Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 91, 6064–6068. doi: 10.1073/pnas.91.13.6064

Sinha, A. K., Jaggi, M., Raghuram, B., and Tuteja, N. (2011). Mitogen-activated protein kinase signaling in plants under abiotic stress. *Plant Signal Behav.* 6, 196–203. doi: 10.4161/psb.6.2.14701

Song, F., and Stieger, K. (2017). Optimizing the DNA donor template for homology-directed repair of double-strand breaks. *Mol. Ther. - Nucleic Acids* 7, 53–60. doi: 10.1016/j.omtn.2017.02.006

Sun, W., Chen, H., Wang, J., Sun, H. W., Yang, S. K., Sang, Y. L., et al. (2015). Expression analysis of genes encoding mitogen-activated protein kinases in maize provides a key link between abiotic stress signaling and plant reproduction. *Funct. Integr. Genomics* 15, 107–120. doi: 10.1007/s10142-014-0410-3

Teige, M., Scheikl, E., Eulgem, T., Dóczi, R., Ichimura, K., Shinozaki, K., et al. (2004). The MKK2 pathway mediates cold and salt stress signaling in arabidopsis. *Mol. Cell* 15, 141–152. doi: 10.1016/j.molcel.2004.06.023

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., et al. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* 32, 947–951. doi: 10.1038/nbt.2969

Xie, K., Minkenberg, B., and Yang, Y. (2014). Targeted gene mutation in rice using a CRISPR-Cas9 system. *Bio-protoc.* 4 (17). doi: 10.21769/BioProtoc.1225

Xie, K., and Yang, Y. (2013). RNA-Guided genome editing in plants using a CRISPR-cas system. *Mol. Plant* 6, 1975–1983. doi: 10.1093/mp/sst119

Yang, D., Scavuzzo, M. A., Chmielowiec, J., Sharp, R., Bajic, A., and Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci. Rep.* 6. doi: 10.1038/srep21264

Yang, C., Wang, R., Gou, L., Si, Y., and Guan, Q. (2017). Overexpression of populus trichocarpa mitogen-activated protein kinase Kinase4 enhances salt tolerance in tobacco. *IJMS* 18. doi: 10.3390/ijms18102090

Zhou, X., Jacobs, T. B., Xue, L.-J., Harding, S. A., and Tsai, C.-J. (2015). Exploiting SNP s for biallelic CRISPR mutations in the outcrossing woody perennial populus reveals 4-coumarate:CoA ligase specificity and redundancy. *New Phytol.* 208, 298–301. doi: 10.1111/nph.13470

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 May 2023
DOI 10.3389/fpls.2023.1193065

OPEN ACCESS

EDITED BY
Dayong Li,
Beijing Vegetable Research Center, China

REVIEWED BY
Zhiqian Pang,
University of Florida, United States
Jianbo Xie,
Beijing Forestry University, China

*CORRESPONDENCE
Quanzi Li
liqz@caf.ac.cn
Hui Li
huil1984@163.com

RECEIVED 24 March 2023
ACCEPTED 26 April 2023
PUBLISHED 30 May 2023

CITATION
Wei M, Li H, Wang Q, Liu R, Yang L and Li Q (2023) Genome-wide identification and expression profiling of B3 transcription factor genes in *Populus alba* × *Populus glandulosa*. *Front. Plant Sci.* 14:1193065. doi: 10.3389/fpls.2023.1193065

Genome-wide identification and expression profiling of B3 transcription factor genes in *Populus alba* × *Populus glandulosa*

Mingke Wei[1], Hui Li[2]*, Qiao Wang[3], Rui Liu[1], Linxi Yang[1] and Quanzi Li[1]*

[1]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing, China, [2]Guangzhou Institute of Forestry and Landscape Architecture, Guangzhou, Guangdong, China, [3]College of Resources and Environment, Qingdao Agricultural University, Qingdao, Shandong, China

B3-domain containing transcription factors (TFs) are well known to play important roles in various developmental processes, including embryogenesis, seed germination, etc. Characterizations and functional studies of the B3 TF superfamily in poplar are still limited, especially on their roles in wood formation. In this study, we conducted comprehensive bioinformatics and expression analysis of B3 TF genes in *Populus alba* × *Populus glandulosa*. A total of 160 B3 TF genes were identified in the genome of this hybrid poplar, and their chromosomal locations, syntenic relationships, gene structures, and promoter cis-acting elements were analyzed. Through domain structure and phylogenetic relationship analyses, these proteins were classified into four families LAV, RAV, ARF, and REM. Domain and conservation analyses revealed different gene numbers and different DNA-binding domains among families. Syntenic relationship analysis suggested that approximately 87% of the genes resulted from genome duplication (segmental or tandem), contributing to the expansion of the B3 family in *P. alba* × *P. glandulosa*. Phylogeny in seven species revealed the evolutionary relationship of B3 TF genes across different species. B3 domains among the eighteen proteins that were highly expressed in differentiating xylem had a high synteny, suggesting a common ancestor for these seven species. We performed co-expression analysis on the representative genes in two different ages of poplar, followed by pathways analysis. Among those genes co-expressed with four B3 genes, 14 were involved in lignin synthases and secondary cell walls biosynthesis, including *PagCOMT2*, *PagCAD1*, *PagCCR2*, *PagCAD1*, *PagCCoAOMT1*, *PagSND2*, and *PagNST1*. Our results provide valuable information for the B3 TF family in poplar and show the potential of B3 TF genes in engineering to improve wood properties.

KEYWORDS

B3, poplar, xylem development, co-expressed, lignin

1 Introduction

The plant-specific B3 superfamily constitutes one of the largest transcription factor (TF) families in plants. All members of the B3 superfamily contain an approximately 110 amino acid region called the B3 domain (Swaminathan et al., 2008). This domain is initially named because it is the third basic domain in the maize protein VIVIPAROUS1(VP1) (McCarty et al., 1991). The B3 domain of VP1 has a sequence-specific DNA binding activity (Suzuki et al., 1997). Subsequently, the B3 domain has been found in 118 genes of *Arabidopsis* and in 91 genes of rice. B3 TF genes are also present in green algae, mosses, liverworts, ferns, and gymnosperms (Marella et al., 2006).

The B3 superfamily encompasses several distinct gene families, including LAV (LEAFY COTYLEDON2 [LEC2]-ABSCISIC ACID INSENSITIVE3 [ABI3]-VAL) (Wang et al., 2012), ARF (AUXIN RESPONSE FACTOR) (Ulmasov et al., 1997), RAV (RELATED TO ABI3 and VP1) (Kagaya et al., 1999) and REM (REPRODUCTIVE MERISTEM) (Romanel et al., 2009). B3 TFs are involved in a variety of biological processes, such as seed development, embryonic development, etc., generally through polycomb silencing in plants. In *Arabidopsis*, the VAL subfamily genes *VAL1/HIGH-LEVEL EXPRESSION OF SUGAR-INDUCIBLE GENE 2* (*HSI2*), *VAL2/HSI2-Like 1* (*HSL1*), and *VAL3/HSI2-Like 2* (*HSL2*) are expressed in many organs throughout plant development. *val1 val2* double mutant exhibits embryonic callus formation on both roots and shoots, suggesting a role of *VAL1* and *VAL2* in repressing the embryonic developmental program. Germination is delayed in this double mutant, and the analysis shows that HSI2-dependent silencing of *DOG1* promotes the early release of seed dormancy (Chen N et al., 2020). LAV family member LEAFY COTYLEDON 1 (LEC1), which is the Hap3 (HEMEACTIVATED PROTEIN 3) sub-unit of the CAAT-binding factor, controls various aspects of seed development from early embryogenesis to late seed maturation (Baumbusch et al., 2004). ABI3, LEC2, and FUSCA 3 (FUS3) also play key roles in the control of seed maturation in cooperation with LEC1 and abscisic acid (ABA) (Tsukagoshi et al., 2005). In *Vitis Vinifera L.*, overexpression of *VvFUS3* in tomatoes causes a reduction in total cell area and cell number, while the cell size of the fruit pericarp is increased. These results suggest that *VvFUS3* has a role in seed development by influencing the ABA signaling (Ahmad et al., 2022). ARF family member AtARF1 binds the upstream promoter regions of many auxin response genes (Ulmasov et al., 1997). Mutation of *ARF7* causes nearly an absence of lateral root formation, lacking of phototropism and auxin responses (Liscum & Briggs, 1995; Ruegger et al., 1997; Watahiki and Yamamoto, 1997). Both *ARF4* and *ARF3* regulate the abaxial-adaxial polarity in leaves and flowers in *Arabidopsis* (Pekker et al., 2005). RAV1 functions as a negative regulator of lateral root and rosette leaf development, and it is downregulated by the plant steroid hormone epibrassinolide (Hu et al., 2004). *GmRAV* negatively regulates short day (SD)-mediated flowering and hypocotyl elongation, and its overexpression in SD inhibits the growth of soybean leaves, roots, and stems (Zhao et al., 2008). The B3 gene also promotes flowering and functions in the maintenance of the vernalization response in *Arabidopsis*. For example, *FLC* (*FLOWERING LOCUS C*), a gene delays flowering, is transiently downregulated in the *vrn1* (*VERNALIZATION1*) mutant that has vernalization treatment, but levels of *FLC* RNA and protein are increased when the mutant is moved to normal temperature (Levy et al., 2002; Sung and Amasino, 2004). *Brassica oleracea BoREM1*, a putative ortholog of *AtREM34*, is specifically expressed in the cauliflower curd and involved in the determination of floral meristem identity (Franco-Zorrilla et al., 1999). In short, the B3 proteins of the LAV, RAV, ARF, and REM families are mainly involved in the signaling pathways of hormones including auxin, ABA, and brassinosteroid.

Wood formation is a specific biological process in trees. The major components of wood are lignin, cellulose, and hemicelluloses, which are deposited in the secondary cell walls (SCWs) of fiber cells and vessels in the xylem. These three polymers maintain the overall structure and strength of plant SCWs (Wessels et al., 2019). After years of efforts, the mechanism of lignification and SCW deposition has been preliminarily analyzed, and regulatory networks that regulate wood formation have been proposed in some review articles (Zheng-Hua and Zhong, 2015; Borthakur et al., 2022). The regulatory network for SCW thickening includes many NAC and MYB genes, such as NST1 to NST3/SND1 (Mitsuda et al., 2005; Zhong and Ye, 2006; Zhong et al., 2007), VND1 to VND7 (Kubo et al., 2005; Yamaguchi et al., 2008; Zhou et al., 2014), and PtrWND2B/6B (Zhong et al., 2011), are master regulators of xylem differentiation and SCW thickening. However, there are still more upstream regulatory genes for these master regulators, which need to be revealed. *VNI2* (*VND-INTERACTING2*) and *VND7* were co-expressed in the elongated ductal precursors of the root. Overexpression of *VNI2* significantly inhibited the differentiation of normal vessel cells. Transient transformation of the reporter gene showed that VNI2 is a transcriptional repressor, which can inhibit the expression of VND7-regulated vessel-specific genes. Sumoylation of LBD30 in poplar affected the expression of *SND1* and *NST1* in the transient transformation of protoplast systems. As LBD30 is sumoylated by SIZ1, this protein modification activates the regulatory network of SND1/NST1-mediated secondary cell wall formation in fibers. In addition, lignin is the most direct material for the study of wood formation. Lignin accounts for ~20% of SCWs, and it is polymerized by three hydroxy cinnamyl alcohols, including sinapyl alcohol, coniferyl alcohol, and *p*-coumaryl alcohol (Sarkanen and Hergert, 1971; Watahiki and Yamamoto, 1997). These alcohol precursors are termed the syringyl (S), guaiacyl (G), and hydroxyphenyl (H) lignin subunits, respectively. In angiosperms, lignin is polymerized primarily from S and G monolignols and trace amounts of the H monolignol (Wang et al., 2021). The biosynthesis of monolignols lignin occurs in several consecutive reactions involving 11 enzyme families and 24 metabolites, in a branched grid-like pathway (Vanholme et al., 2013; Wang et al., 2014; Sulis and Wang, 2020). These families include phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), *p*-coumaroyl-CoA 3-hydroxylase (C3H), *p*-coumarate CoA ligase (4CL), hydroxycinnamoyl transferase (HCT), caffeoyl shikimate esterase (CSE), caffeoyl-CoA *O*-methyltransferase (CCoAOMT), cinnamoyl-CoA reductase

(CCR), conifer aldehyde 5-hydroxylase (CAld5H, first named F5H, ferulate 5-hydroxylase), caffeic acid 3-*O*-methyltransferase (COMT) and cinnamyl alcohol dehydrogenase (CAD) (Zhao and Dixon, 2011). In this pathway, C3H catalyzes the hydroxylation of coumaric acid to form caffeic acid through the 3-carbon hydroxylation of aromatic rings of various phenol intermediates (Palafox-Carlos et al., 2014). The reduction of feruloyl-CoA to conifer aldehyde is mediated by CCR (Li et al., 2005). Coniferaldehyde is catalyzed by CAD to produce confieryl alcohol, G monolignols (Baucher et al., 1996). Coniferaldehyde can also be converted to sinapyl alcohol by CAld5H/F5H, COMT, and CAD, to produce S monolignols (Lapierre et al., 1999). Overexpression of *CAld5H* have a 64% increase in the S/G ratio in poplar, and co-transformation of *F5H* and *COMT* results in 2- to 3- times higher S/G ratio than *F5H* alone (Li et al., 2003). The lignin S/G ratio varies among different tree species, and it is a more important factor affecting the pulp yield than the lignin content (Del Río et al., 2005; Studer et al., 2011). High pulp yield is correlated with a high S/G ratio in wood biomass, so genes controlling S/G ratio in tree have great promising to breed pulp-specific trees.

Poplar is an ideal species to study the formation of wood, the important biomass energy material in nature. Although the pathway enzyme genes of cellulose, hemicellulose, and lignin biosynthesis and many upstream TF regulators have been identified, these studied regulators are only a small proportion of TFs. Among B3 TF family, PagVAL2-B1 is identified as an candidate regulator involved in wood formation, because yeast one hybridization shows that it can bind to *PagCAld5H2* promoter (Wang et al., 2021). Many remaining genes that are specifically expressed during wood formation may have potential functions in regulating specific wood properties. The involvement of B3 TF family genes in wood formation and their specific regulation on wood property formation needs investigation. In this study, we focused on the characterization of the B3 family in *P. alba* × *P. glandulosa*, and identified potential key genes involved in wood formation.

2 Materials and methods

2.1 Materials

Samples were collected from six-month-old trees that were grown in a greenhouse. After the bark was peeled, the xylem and phloem were scratched by single-end razors, and frozen immediately in liquid nitrogen. The tissues of leaf, xylem and phloem tissue from 10-year-old trees were from the stock in liquid nitrogen in a previous study (Li et al., 2021a).

2.2 Identification and annotation of B3 TF genes in *P. alba* × *P. glandulosa*

P. alba × *P. glandulosa* genome resource was downloaded from the Figshare database (https://figshare.com/articles/dataset/84K_genome_zip/12369209). To identify B3 family members, the HMM profile (PF02362) of the B3 DNA-binding domain from the Pfam database (http://pfam.sanger.ac.uk/) was used as input to perform an HMMER search in the *P. alba* × *P. glandulosa* genome with an E-value cut-off of 1e-3 following the HMMER guider (Horn et al., 2005). In addition, manual annotation is used to correct any inconsistencies between the predicted gene and its actual chromosome location.

All coding sequences (CDSs) were translated into amino acid sequences, and alignment of all B3-domain proteins from Arabidopsis and *P. alba* × *P. glandulosa* was conducted by MAFFT (L-INS-I algorithm) using the B3 domain amino acid sequences (Rozewicki et al., 2019). Subsequently, a phylogeny was generated using IQ-TREE Linux software and Model Finder (Chernomor et al., 2016).

2.3 The naming of B3 family genes

All identified B3 genes were named with consistent patterns, based on their subfamily association and phylogenetic relationships, as well as their subgenome locations (A, B). Each gene is named by starting with the abbreviation for the species name *Pag* (*P. alba* × *P. glandulosa*) followed by the name of the *Arabidopsis* gene with the most updated naming system from this subfamily (e.g. *VAL1* for *VAL-like genes*). The gene names including an A or B indicate the subgenome they are located in: for example, *PagVAL2-B1* and *PagVAL2-A1*. Putative alleles have identical gene names but with different subgenome identifiers (e.g. *PagVAL2-A1*, *PagVAL2-B1*). Genes belonging to one subgenome but at different chromosome locations were consecutively numbered (e.g. *PagVAL2-A1* and *PagVAL2-A2*).

2.4 Phylogenetic tree construction and location of B3 family genes in chromosome

The full length of protein sequences and conservation regions of all the identified B3 family members was aligned by MAFFT (E-INS-I algorithm). The alignments were trimmed using trimAl software, and then a phylogeny tree was inferred under maximum likelihood with IQ-TREE (Capella-Gutiérrez et al., 2009; Chernomor et al., 2016). The substitution model was determined using Model Finder, which is integrated into IQ-TREE. The best-fit model, JTT+F+G4, was selected based on the Bayesian information criterion (Capella-Gutiérrez et al., 2009). To evaluate the reliability of the phylogenetic estimate, Ultrafast bootstraps and a Shimodaira-Hasegawa approximate likelihood ratio test were performed with 1000 replicates each. The phylogenetic tree was visualized on the iTOL web browser (https://itol.embl.de/) and Fig Tree V1.4.4 software.

Gene collinearity analysis and visualization within the B3 genome were conducted by TBtools (Chen C et al., 2020). Detailed analysis methods are described in Li et al., 2022. The chromosome data and gene family file were mapped to their respective locus in the *P. alba* ×

P. glandulosa genome in the acicular diagram using shinyCircos in RStudio (Capella-Gutiérrez et al., 2009).

2.5 Analyses of motifs, domains, gene structures, and cis-acting elements B3 genes

The upstream (2 kb) regions of all identified B3 genes were extracted using the Fasta Extract software (Chen C et al., 2020). Cis-acting elements were analyzed in promoter sequences through the PlantCARE online server (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Conserved motifs of proteins were analyzed using MEME Suit 5.5.1 (http://meme-suite.org/tools/meme), searching up to 28 conserved motifs. Protein domains were identified using the Conserved Domain Database on NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). The gff3 file provided by the genome was used to display the gene structure, including the exon, intron, and UTR regions using TBtools (Chen C et al., 2020). The Gene Structure View module in TBtools was used to visualize the motif and domain of B3 genes.

2.6 Conservation region of the B3 proteins

A BLASTp was performed of all B3 proteins to *A. thaliana* with the threshold of 1e-5 to divide them into four families. The protein sequences in each family were firstly aligned by clustal in MAFF software; then the aligned fasta file was inputted into GeneDoc software (Nicholas and Nicholas, 1997) to show the conserved region of each type of B3 proteins. The conserved regions were then inputted into Jalview software (Waterhouse et al., 2009) to show the conserved sequences and seqlogo graphs of each type of B3 protein.

2.7 Collinearity and phylogeny analyses of B3 genes among different species

The multiple Chr layout, gene link, and gff3 files between two species in seven species were produced by the One Step MCScanX-Super Fast module in TBtools with the *E*-value of 1e-3, including *A. thaliana*, *P. trichocarpa*, *Gnetum montanum*, *Alsophila spinulosa*, *Physcomitrella patens*, *Spirogloea muscicola* and *P. alba* × *P. glandulosa*. The three files generated from two datasets were merged using the File Merge module of MCScan-X software, and homologous genes among different species were obtained from the merged genelink file. By extracting protein sequences of all these homologous genes, a phylogenetic tree among seven species was constructed using IQ-TREE by the maximum likelihood method with Ultrafast bootstraps (1000 replicates). The collinearity plots among different species were visualized using the Multiple Synteny Plot (MSP) module in TBtools software. The CDSs of B3 domain-containing proteins were identified from the genome and transcriptome databases of green algae, moss, fern, gymnosperm, and angiosperm (Supplementary Table S1) with BLASTP with the threshold of 1e-5 using B3s protein from *A. thaliana* as separate queries.

2.8 RNA-seq analysis

The expression patterns of target genes in three tissues, including xylem, phloem and leaf, were analyzed using the two RNA-seq data from ten-year-old trees (Li et al., 2021a) and six-month-old trees (Li et al., 2021b), respectively. Expression levels of 160 B3 genes and 86 midnight-blue module genes were analyzed by the DESeq2 software. The expression levels were quantified as $\log_2$ (transcripts per million) ($\log_2$tpm), and a heatmap was generated using MORPHEUS (http://software.broadinstitute.org/). The genes with undetectable expression levels were removed, and the remaining genes were clustered according to their expression correlation using Hierarchical (Metric: One minus Pearson correlation, Linkage method: Complete).

2.9 Reverse transcription-quantitative PCR

Total RNAs were extracted using the RNeasy Plant Mini Kit (QIAGEN, CA, USA), and reverse-transcribed to cDNA using a Prime Script TM RT Reagent Kit with gDNA Eraser (TaKaRa, Dalian, China). PCR was conducted using the Green Premix ExTagII (TaKaRa, Dalian, China) and detected by the Roche Light Cycler 480 II. Actin was used as the reference. The primers used are shown in Supplementary Table S2.

2.10 Gene co-expression networks and pathways analyses

By utilizing the WGCNA package in R, we were able to detect gene co-expression through analyzing the gene networks. The different model results were then used to visualize by Cytoscape. KEGG enrichment analysis was carried out to screen out the modules of genes related to wood formation. Among them, the midnight-blue enrichment module was selected for analysis. The correlation coefficient of genes in the network was calculated using the Pearson algorithm. The top eighteen genes with the highest weights were selected for drawing the co-expression networks. Eighty-six genes in the modules related to wood formation were used for co-expression analysis.

3 Results

3.1 Evolution relationship of B3 TFs and chromosome distribution of B3 genes *in P. alba* × *P. glandulosa*

To identify B3 genes in the *P. alba* × *P. glandulosa* genome, we performed the HMM search using a Hidden Markov Model algorithm with the conserved B3 domain model (PF02362) as an inquiry. After protein sequences were trimmed based on the alignment, the protein sequences of 160 B3 proteins in *P. alba* × *P. glandulosa* were used to construct the phylogenetic tree, with B3 proteins from *Arabidopsis* (Figure 1).

These 160 B3 proteins were classified into four families, LAV, RAV, ARF, and REM, which contained 10, 20, 56, and 74 members, respectively. The LAV family in *Arabidopsis* consists of two subgroups, the LEC2-ABI3 subgroup and the VAL subgroup. However, the LAV family in *P. alba* × *P. glandulosa* only had one, the VAL subgroup. The gene number of RAV and ARF families in *P. alba* × *P. glandulosa* was similar to that in *Arabidopsis*. Among the four families, the REM family had the most varied gene numbers between *P. alba* × *P. glandulosa* and *Arabidopsis*. The results show that the LAV and REM families are more diverse than the other two families, which have constant gene numbers.

We detected the distribution of the identified B3 genes in chromosomes. They were distributed unevenly on chromosomes (Figure 2). Chromosomes Chr04, Chr14, and Chr15 contained significantly more B3 genes than others did. The B3 genes in these three chromosomes were 29, 18, and 19, respectively, occupying 18.1%, 11.3%, and 11.8% of B3 genes on chromosomes. In contrast, no B3 gene was present in Chr02B, Chr07A, Chr12A, Chr17B, and Chr18B. There was no correlation between the length of the chromosome and the distribution of B3 genes on the chromosome.

3.2 Motifs, domains, gene structures, and promoter cis-acting elements analyses

To gain insight into the identified 160 B3 gene members, we performed domain, motif and gene structure analyses. The motif analysis showed that most B3 TFs contained motifs 1, 2, and 4, indicating that the structures of B3s are conserved (Figure 3, and Supplementary Figures S1, 2). In the REM family, 44 B3 proteins had motifs 1 and 4, whereas the remaining 30 B3 proteins had only motif 1. Motif 25, which corresponded to the zinc finger CW domain (zf-CW), was found in nearly all members in the VAL subgroup. The ARF family is more conservative in terms of the number and distribution of motifs. According to domain analyses, we found that a number of the B3 proteins contained two B3-specific domains (B3 and B3_DNA), indicating specific characteristics of the B3s proteins. The domain analysis indicates that all 160 members belong to the B3 gene family. However, there were a number of variations in the domain number and types between the different B3 proteins. For example, in the ARF family, 24 proteins had three conserved domains (B3, Auxin_resp, and AUX_IAA superfamily), and 32 proteins had only two conserved domains (B3 and Auxin_resp) (Supplementary Figure S3). Among 10 VALs, 9 contained a zf-CW domain and a B3_DNA domain (Figure 4A). In the RAV family, 8 members (PagRAV2-A1/2, PagRAV2-B1, PagRAV3-B, PagTEM2-A1/2, and PagTEM2-B1/2) contained B3 and AP2 domains, whereas the rest of the members had only one B3 domain (Figure 4B). REM proteins showed more variations in domain number and types than the other family B3 proteins. For example, 31 REM members contained multiple B3 domains (Supplementary Figure S4). Gene structure analysis showed that 15% of B3 genes in *P. alba* × *P. glandulosa* contained only one exon, with incomplete untranslational region (UTR), and 3.8% had very long transcripts, potentially resulting from suboptimal genome annotation.

The promoter region among different B3 genes showed variable cis-acting elements. These cis-acting elements were involved in

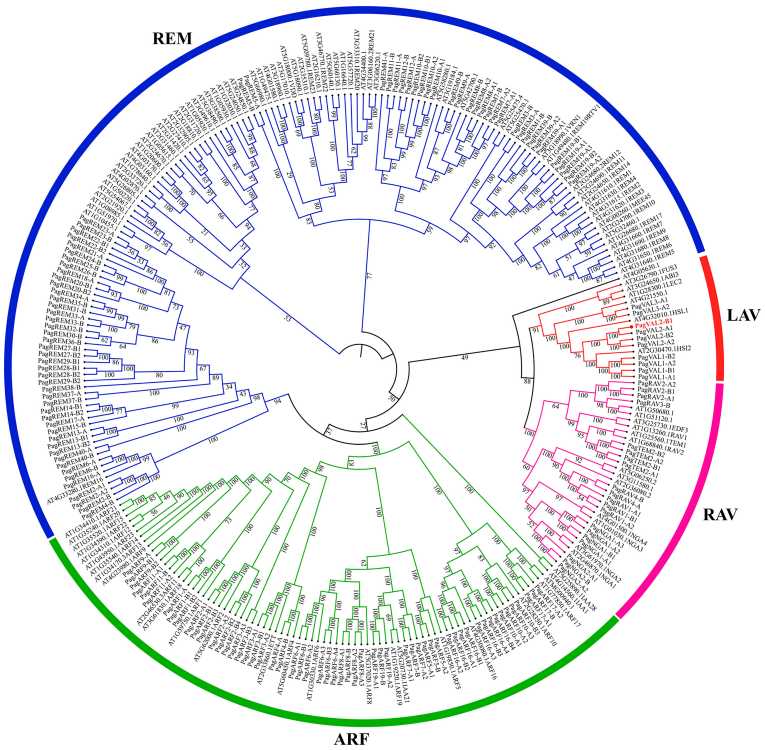


FIGURE 1
Phylogeny analysis of B3 family members. Colors represent different clades of B3 genes. The number on the tree branches represents the bootstrap value.

10.3389/fpls.2023.1193065

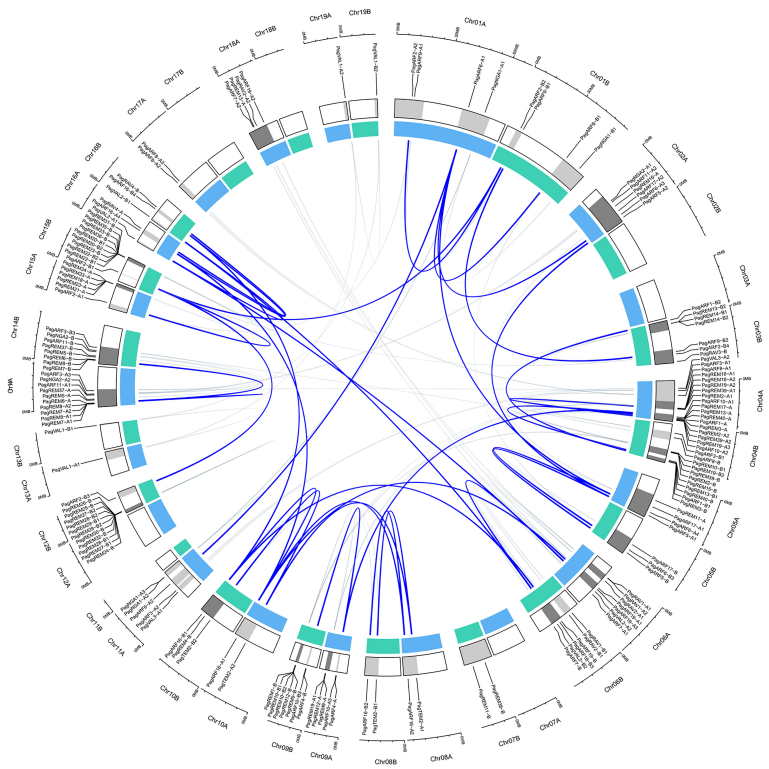


FIGURE 2
Distributions of B3 genes in the chromosomes of *P. alb* × *P. glandulosa*. The genes were mapped to their respective locus in chromosomes in a circular diagram using shinyCircos. Subgenomes are indicated by different shades of blue (inner track), and chromosomal segments are indicated by shades of gray (outer track). Homoeologous genes were inferred by phylogeny and linked with a grey line (inside of the circle). The highlighted blue line within the circus means the gene paralogous relationship of eighteen genes was highly expressed in the xylem.

hormone metabolism, stress response, flavonoid biosynthesis, and so forth (Figure 3; Supplementary Figures S1, 2), such as some hormone metabolism-related motifs, including TGA-element with the AACGAC binding site, TCA-element with the TCAGAAGAGG binding site, MeJA-responsiveness (methyl jasmonate) with the TGACG/CGTCA binding site, and ABRE element with the ACGTG binding site. Among them, ~25% (40 out of 160) B3 genes harbored TGA elements, ~58% (89 out of 160) contained TCA elements, ~70% (112 out of 160) contained MeJA elements, and ~75% (125 out of 160) contained ABRE elements. Cis-acting elements involved in auxin responsiveness (AuxRR-core) were found only in the promoter regions of the ARF (10 out of 56) and REM (7 out of 74) families (Supplementary Figure S5).

3.3 Expansion patterns of B3 genes in *P. alba* × *P. glandulosa*

Gene duplication events contribute to gene proliferation in the plant kingdom and often evolve into partition existing functions to

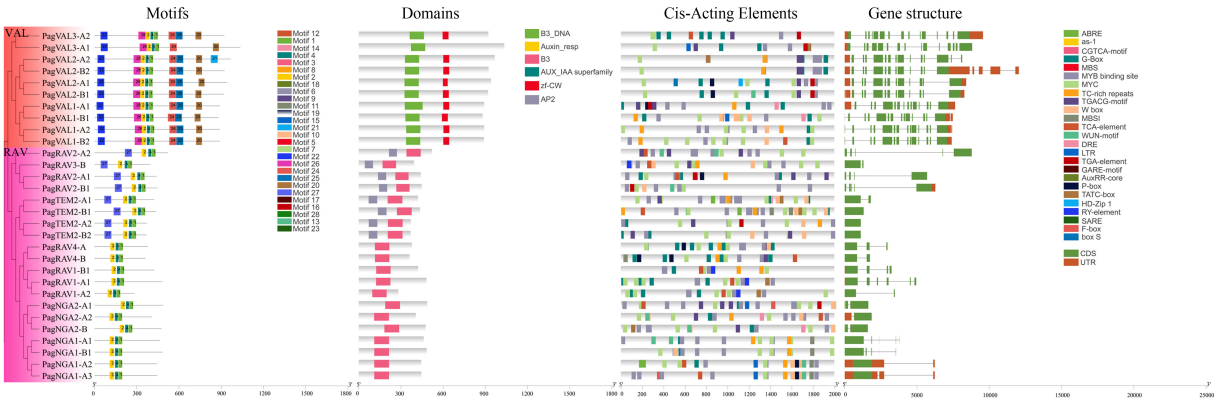


FIGURE 3
Motifs, domains, cis-acting elements, and gene structures of the LAV and RAV families. Different colors represent different motifs, domains, cis-acting elements, and gene structures. The ARF and REM results are in Supplementary Figures S1, S2.

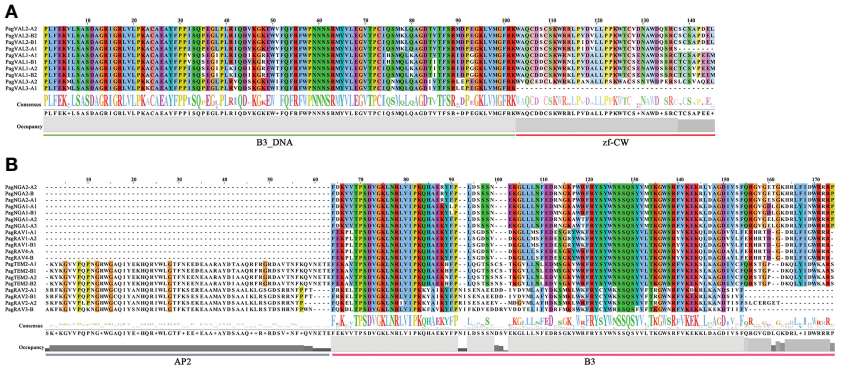


FIGURE 4
Sequence alignment and seqlogo of B3 gene families. **(A)** VAL **(B)** RAV. The same color in the column and the big size of the letters mean a highly conservative region. The ARF and REM domain sequence alignment are in Supplementary Figures S4, S5. The line underneath of sequences indicates residues of the domain consensus block.

form sub-functionalizations or neo-functionalizations (Airoldi and Davies, 2012; Lee & Irish, 2011). According to these criteria, we utilized the shinyCircos software to perform a genome-level collinearity analysis of the B3 genes (Yu et al., 2018). Figure 2 (inside of the circle) showed the associated gene pairs of B3 genes. Out of 160 genes, 141 were discovered to possess paralogous genes, which suggests that B3 genes have undergone significant gene duplication occurrences. The B3 genes, which exist widely on 19 pairs of homoeologous chromosomes, are broadly distributed. Moreover, two or more genes found on the same chromosome within a 200 kb region were likely to be the result of tandem duplication (Houb, 2001). We observed that 44 B3 genes were clustered by 13 tandem duplication events. Among them, 7 chromosomes (Chr11A, Chr06A, Chr14A, Chr14B, Chr04B, Chr15A and Chr03B) had one duplication and 3 chromosomes (Chr4A, Chr15B, and Chr12B) had two duplications. It is surprising that the majority of genes belong to the REM family, suggesting that REM has undergone a greater degree of evolutionary change when compared to other B3 families. In addition to tandem duplication, many B3 genes resulted from segmental duplication events, such as PagVAL1-A1 and PagVAL1-A2 being located in Chr16A and Chr06A, respectively.

3.4 Expression patterns of B3 genes in three tissues of *P. alba* × *P. glandulosa*

To understand the expression patterns of these 160 B3 genes, we quantified their expression levels in three tissues, including leaf, xylem, and phloem, using the RNA-seq data from 6-month-old and 10-year-old trees. Four VAL subfamily genes, *PagVAL2-A1*, *PagVAL2-A2*, *PagVAL2-B1*, and *PagVAL2-B2*, were highly expressed in the differentiating xylem at two age trees, and the expression level increased with age, indicating that they may play important roles in xylem development. Two RAV family genes, *PagTEM2-B2* and *PagRAV4-A*, were also highly expressed in differentiating xylem of two age trees at the same level. The other 18 genes in the RAV family were preferentially expressed in leaves, such as *PagNGA2-B*, *PagNGA2-A2*, and *PagNGA1-A1*. In addition, many genes were not expressed in the 6-month-old trees but were highly expressed in 10-year-old trees, such as *PagVAL3-A2*, *PagARF2-B3*, and *PagREM2-A1* (Supplementary Figures S6–8).

3.5 Collinearity and evolutionary relationships of xylem formation-related B3 genes among seven species

Eukaryotic genomes vary in their level of synteny and collinearity, referring to the degree to which genes are retained on the same chromosomes and in the same order, respectively (Tang et al., 2008). To investigate the origin, evolutionary history, and potential functions of the *P. alba* × *P. glandulosa* B3 genes, we examined genomic synteny of *P. alba* × *P. glandulosa* with *A. thaliana* (herb), *P. trichocarpa* (woody), *G. montanum* (woody), *A. spinulosa* (fern), *P. patens* (moss) and *S. muscicola* (green algae). Based on expression pattern results, 18 genes that were highly expressed in the differentiating xylem of six-month-old trees were used to perform BLASTp with proteins in the above seven species to find their homologs (Figure 5A) (Supplementary Table S3). We obtained a total of 54 homologous proteins that corresponded to the 15 B3 proteins in *P. alba* × *P. glandulosa*. These 54 proteins included 8 in *P. patens*, 10 in *A. spinulosa*, 13 in *P. trichocarpa*, 9 in *Arabidopsis*, 9 in *G. montanum*, and 5 in *S. muscicola* (Figure 5B; Supplementary Table S3). We observed that most paralogous genes in *P. alba* × *P. glandulosa* corresponded to one gene in *Arabidopsis*: for example, *AT5G62000* was orthologous to *PagARF2-B2* and *PagARF2-B4*, and *AT1G59750* was orthologous to *PagARF1-B1* and *PagARF1-B2*, supporting the whole-genome duplication in trees (Tuskan et al., 2006). On the other hand, *AT5G62000* and *AT1G59750* corresponded to the same gene *TnS000811647t03* in *A. spinulosa*, indicating gene extraction in poplar. In all, our result suggests that they shared a common ancestor during evolution.

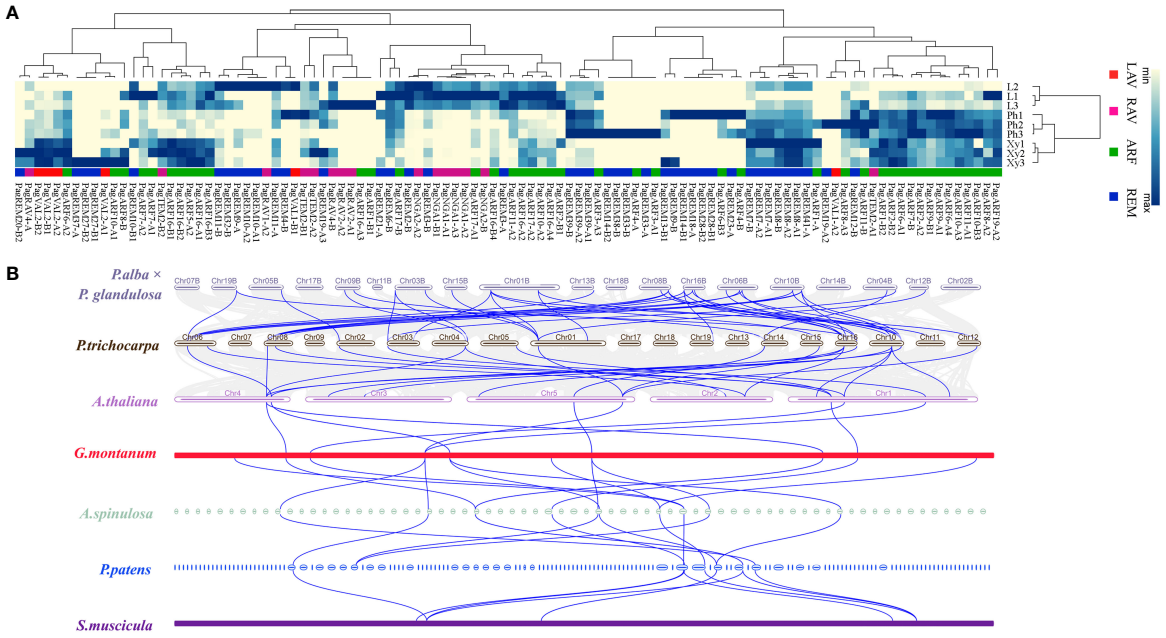


FIGURE 5
Expression analysis of B3 family genes in three tissues and collinearity analysis among different species. **(A)** A heatmap shows the expression level of 104 genes in different subfamilies (rows columns). L, leaf; Ph, phloem; Xy, xylem. **(B)** Multi-collinearity analysis of seven species. A blue line between two species means collinearity genes among different species.

3.6 Gene co-expression analysis

Gene co-expression analysis can help identify genes with similar expression patterns. These genes may interact and co-regulate each other, exhibit similar functions, or be involved in the same signaling pathways or physiological processes. To further understand the correlation of these 160 B3 genes in wood formation, we conducted Weighted Correlation Network Analysis (WGCNA) using the RNA-Seq data of leaf, xylem, and phloem from 6-month-old and 10-year-old trees to construct a co-expression network centered on the 6 tissue-differentially expressed genes. As shown in Figure 6A, we obtained a total of 6 cluster dendrogram modules. Among them, the network in the same module has a strong correlation (Figure 6B). One of the modules centered on four B3 genes (*PagVAL2-A1*, *PagVAL2-B1*, *PagARF3-A1*, *PagARF3-B1*) is related to the wood formation (826 genes). KEGG enrichment analysis showed that these genes shared 20 significantly enriched pathways (Figure 6C). The shared pathways include biosynthesis secondary metabolites pathway, metabolic pathway, and phenylpropanoid biosynthetic pathway.

To investigate the effects of B3 genes on the lignin biosynthetic pathway, we examined the expressions of the 86 genes involved in the secondary cell wall and phenylpropanoid biosynthetic pathways during wood formation. Among them, 36 genes were highly expressed in the differentiating xylem of 6-month-old trees, and most genes were highly expressed in the differentiating xylem of ten-year-old trees. There were fourteen genes co-expressed with these four B3 genes, such as lignin synthase genes *PagCOMT2*, *PagCAD1*, *PagCCR2*, *PagCAD1*, *PagCCoAOMT1*, and SCW TF genes *PagSND2* and *PagNST1* (Figures 7A, B). To further validate the results of RNA-Seq, we used RT-qPCR to quantify expression levels of the eighteen lignin biosynthetic pathway genes in leaf, phloem, and xylem of the two ages poplar. Among these 86 genes that were highly expressed in the xylem, 4 belonged to B3 genes, and these four genes (*PagVAL2-A1*, *PagVAL2-B1*, *PagARF3-A1*, and *PagARF3-B1*) were highly expressed in the differentiating xylem of two ages, indicating that they may play important roles in xylem development Figure 7C. Among the 14 genes that were co-expressed with the four B3 gene, 12 were highly expressed in the developing xylem of two ages trees Figure 8. Among the 12 genes, *PagNST1*, *PagCCoAOMT1*, and *PagCCR2* were significantly increased in 10-year-old tree compared to 6-month, while the remaining nine genes showed opposite expression trends.

4 Discussion

The plant-specific superfamily of B3 transcription factors is distinguished by the existence of one or more B3 domains, or a fusion of B3 domains with supplementary domains like AP2 (APETALA2), AUX_IAA, and zf-CW (Peng and Weselake, 2013). The B3 transcription factor superfamily, comprises four families, among which the LAV(LEC2/VP1/VAL) and ARF families are well-studied in *Arabidopsis*, characterized by diverse functions in plant growth and development, and extensively studied in *Arabidopsis* (Yamasaki et al., 2004; Swaminathan et al., 2008; Yamasaki et al., 2008; Agarwal et al., 2011). In contrast, the identification and characterization of B3 genes in woody plants is limited. New insight into this superfamily can be obtained through genome analysis. In this study, we performed a comprehensive

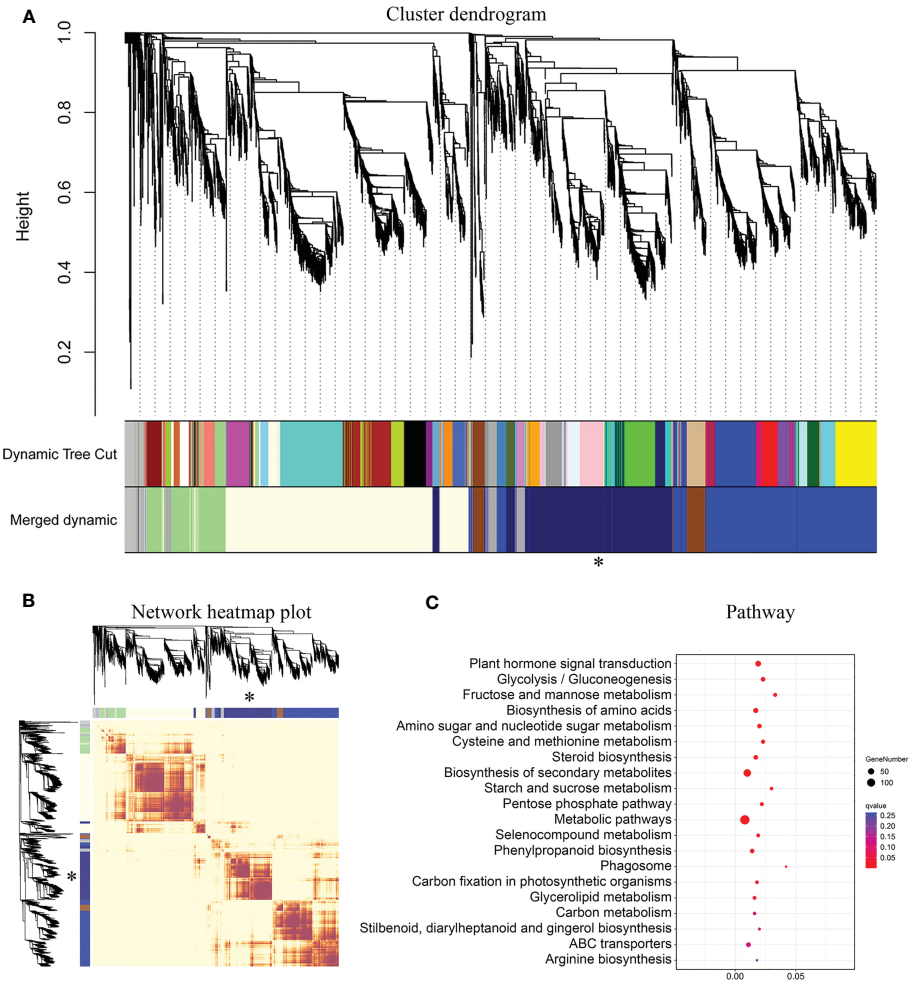


FIGURE 6
Result of WGCNA analysis and KEGG analysis of midnight-blue module. **(A)** Hierarchical clustering gene module in two different ages poplar. **(B)** Gene co-expression network heatmap of different modules. **(C)** KEGG analysis for the midnight-blue module. *, The key module related to wood formation. The different colors in A and B represent different modules. The heatmap describes adjacencies among genes in analysis. The dark color means a high correlation between two genes. Light-luminance color means a low correlation between two genes. The size of the pot means gene number, and darkblue means significantly enriched pathways.

characterization of B3 genes in *P. alba* × *P. glandulosa* and investigated their expression profiles in leaf, phloem, and xylem. We identified 160 putative B3 genes in the genome of *P. alba* × *P. glandulosa*, showing a discrepancy of gene number in each family with *Arabidopsis*. Different species are expected to have varying numbers of B3 genes in each family, and the number of B3 genes identified in different studies may also be influenced by the use of different database sources, methods, and parameters (Romanel et al., 2009; Peng and Weselake, 2013). For example, previous studies have classified B3 proteins lacking typical AP2 domains as members of the RAV family (Magnani et al., 2004; Kim et al., 2006). Similarly, in the LAV family, we identified only 10 VAL subgroup members, while LEC2-ABI3 was not identified according to our analysis; this could be due to the protein sequence variation from species to species. In addition to the B3 domain, the VAL members contain a zf-CW domain, with the one exception of PagVAL3-A1. For the RAV family, our results are consistent with previous results showing that some proteins contain the B3 and AP2 domains (Alvarez et al., 2009). Based on the evolutionary relationships between RAV proteins in *Arabidopsis* and *P. alba* × *P. glandulosa*, it was found that 8 out of 20 RAV proteins had AP2 domains. In the ARF family, 24 out of 56 members contained three domains (B3, Auxin_resp, and AUX_IAA), while the remaining 32 contained only two domains (B3 and Auxin_resp). The same observations have been reported in other plants, such as *Arabidopsis*, apple, and papaya (Li et al., 2015; Peng et al., 2020). We speculated that loss of the zf-CW, AP2, and AUX_IAA domains occurs during evolution. In a gene family, the emergence of new members can occur through domain duplication or loss. For example, in the MYB family, the plant-specific R2R3 organization is thought to have evolved from an R1R2R3-type ancestral gene from which the first repeat domain R1 motif was lost (Braun and Grotewold, 1999; Riechmann et al., 2000; Dias et al., 2016).

The domain arrangement observed in REM proteins is more intricate compared to other families. Thirty-one members of the REM family contain multiple B3 domains: for example, PagREM6-A/

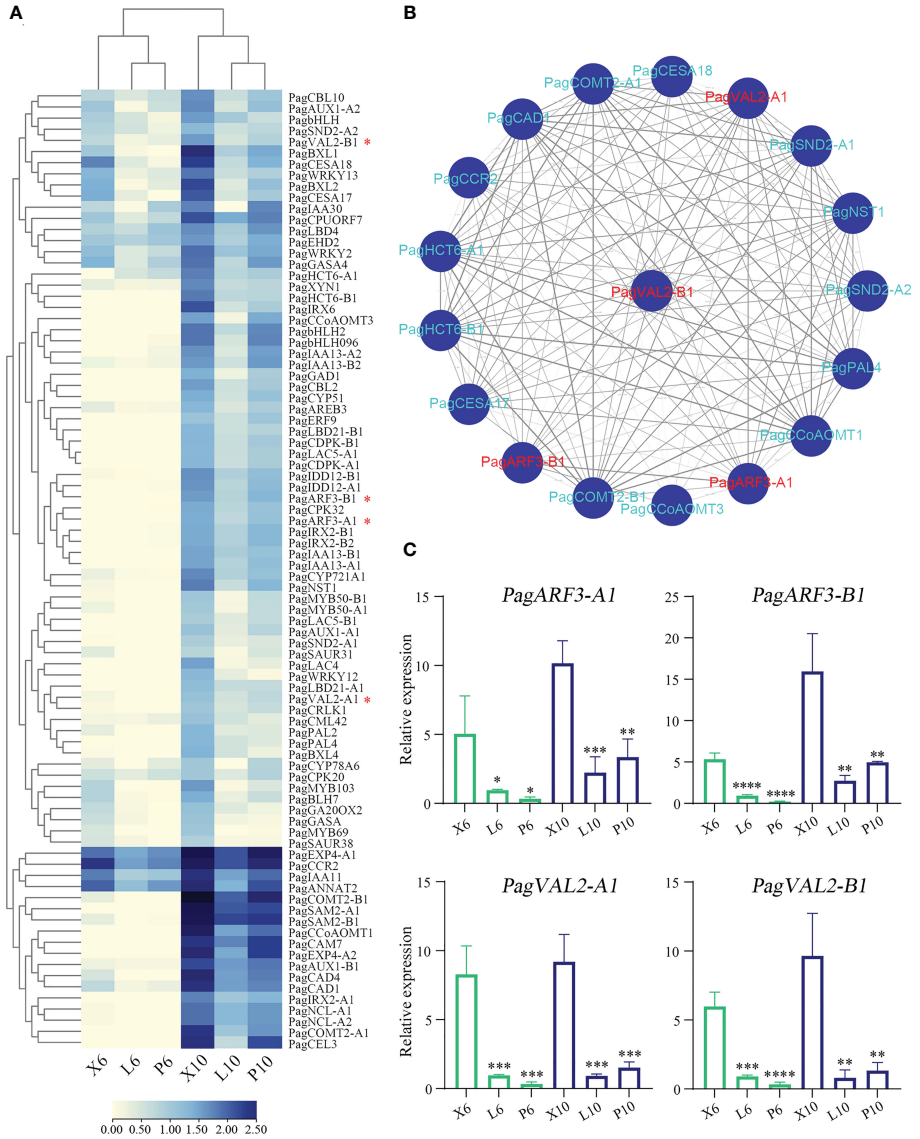


FIGURE 7
Gene expression profile in a wood formation-related module in poplar and co-expression analysis of B3 family genes. **(A)** Heatmap shows the expression of DEGs associated with wood formation. **(B)** Co-expression network of one tissue differentially expressed B3 gene *PagVAL2-B1*. **(C)** Four B3 family gene expression levels based on RT-qPCR. X6, L6, and P6 represent the xylem, leaf, and phloem of 6-month-old trees. X10, L10, and P10 represent the xylem, leaf, and phloem of 10-year-old trees. The relative expression levels are shown as the means ± SD of three biological replicates. The significance between xylem and phloem and between xylem and leaf was analyzed using Dunnett's multiple comparisons test (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.001$).

B contained the highest number of five (Supplementary Figure S4). The complexity of domain architecture in REM proteins suggests that they may have evolved from one or more domain duplication events, which can lead to functional diversification over the course of evolution. However, the study of genes with domain duplications is often challenging as the functional redundancy of many proteins can make it difficult to ascertain their specific roles (Waltner et al., 2005). Previously, domain duplication has been reported in the RAV family as well as other families of transcription factors. For example, in the liverwort *Marchantia polymorpha*, a RAV protein has been identified to contain two B3 domains, indicating the occurrence of domain duplication in this protein family (Swaminathan et al., 2008). Among the 167 basic helix-loop-helix (bHLH) proteins that have been identified in the rice genome, it has been discovered that one particular protein (OC_Os01g09930) contains two bHLH domains that are duplicated (Li et al., 2006).

Gene duplication is the primary source of the evolution of genes and gene families (Cannon et al., 2004). Previous research has demonstrated that the expansion of the B3 superfamily in *Brassica rapa* is primarily driven by tandem duplication, as evidenced by the presence of both two-copy and multi-copy tandem-arrayed B3 genes in its genome (Peng and Weselake, 2013). Duplicated B3 genes and other transcription-factor genes have also been found in the *Arabidopsis* genome (Riechmann et al., 2000; Romanel et al., 2009). Similar findings have been reported in *Cicer arietinum*, where *CaARF4/CaARF5* and *CaARF21/CaARF22* are duplicated

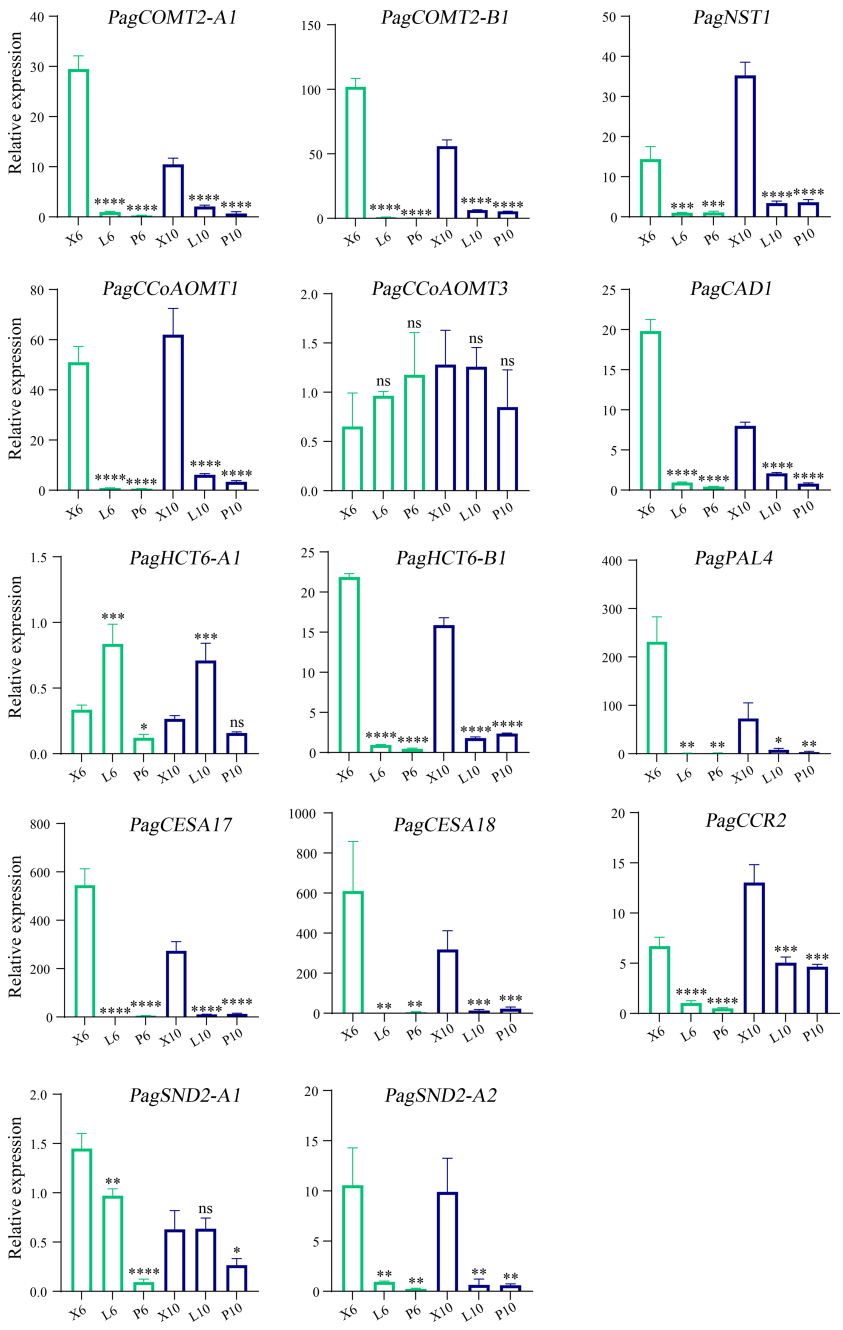


FIGURE 8
Gene expression levelsGene expression levels determined by RT-qPCR. X6, L6, and P6 denote the xylem, leaf, and phloem, respectively. The green and blue columns represent six-month-old and 10-year-old poplar trees. X6, L6, and P6 represent the xylem, leaf, and phloem of 6-month-old trees. X10, L10, and P10 represent the xylem, leaf, and phloem of 10-year-old trees. The relative expression levels are shown as the means ± SD of three biological replicates. The significance between xylem and phloem and between xylem and leaf was analyzed using Dunnett's multiple comparisons test (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.001$). ns, not significant.

in tandem but exhibit distinct structural differences (Die et al., 2018). Likewise, the REM family we identified is distributed on many chromosomes (Chr04A/B, Chr09A/B, Chr14A/B, and Chr15A/B). Tandem duplications appear to have occurred in 40 genes among 74 members of the REM family. In addition, only 44 of all the 160 B3 genes are products of tandem duplication events. Consistent with this, we found that the rest of the gene belongs to segmental duplication. This analysis suggests that segmental duplication may happen more frequently than tandem duplication in *P. alba* × *P. glandulosa*, and that numerous duplicated chromosomal blocks are retained in their genome after duplications (Cannon et al., 2004).

The B3 domain is not specific for angiosperms. It is also present in genes from gymnosperms, ferns, mosses, liverworts, and green

algae (Marella et al., 2006). Evolution and genomic synteny can provide clues to gene function, and can be an effective tools for gene family analysis in the species with available whole-genome sequences (Lyons et al., 2008; Zhang et al., 2012). There is only a single B3 gene in *Chlamydomonas reinhartii*, a single-celled green algae with available complete genome sequence, and an ortholog of this gene has been identified in green algaes *Mesostigma viride* and *S. muscicola*, suggesting that the B3 domain arose before the development of multicellularity in the plant lineage (Merchant et al., 2007). This single B3 gene in *C. reinhartii* is similar in structure to the existing VAL subgroup in the most recent common ancestor of green algae and higher plants. Thus it is hypothesized that the *C. reinhartii* B3 gene is the ortholog of the VAL subgroup (Swaminathan et al., 2008). In the metaphyte lineage, a series of gene duplication events occurred before the speciation event between mosses and vascular plants. The first led to the LEC2-ABI3 subgroup, and subsequent gene duplications led to the RAV, ARF, and REM families (Swaminathan et al., 2008). Because of the evolutionary diversity and differences in genome polyploidy between different species from green algae to vascular plants, we selected seven species for collinearity analysis using orthologous genes. Eighteen B3 TF genes that were highly expressed in the xylem were selected for collinearity analysis. These genes can be found to be homologous in seven species and have collinearity on chromosomes, suggesting that these B3 genes were already present in early subaerial/terrestrial plants (Wang et al., 2020). For example, *SM000256S08685* gene in *S. muscicola*, *PAC:32923104* gene in *P. patens*, *Aspi01Gene22980.t1* gene in *A. spinulosa*, and *AT4G32010.1* gene in *A. thaliana* are orthologs of *Pag.B16G000101* (*PagVAL2-B1*) gene in *P. alba* × *P. glandulosa*, belonging to the VAL subgroup. These findings suggest that the VAL genes have undergone a complex evolutionary history.

The B3 domain is highly conserved in B3 TFs, and it consists of 2 α-helices and 7 β-strands. Differences in amino acid residues may lead to different DNA binding ability or binding sequences of B3 proteins (Yamasaki et al., 2004; Waltner et al., 2005). The B3 domain of *Arabidopsis* VAL protein can recognize and bind to Sph/RY elements (CATGCA), where histone H3 lysine 27 trimethylation (H3K27me3) is marked and suppresses the transcription of downstream genes. When the B3 domain in VAL protein is destroyed, the expression of a downstream gene *DOG1* is significantly upregulated, indicating that VAL proteins function through B3 domain-mediated binding (Chen N et al., 2020). In addition, The AP2 and B3 domains in RAV1 protein can bind to the CAACA and CACCTG motifs, respectively, promoting the high-affinity activity of RAV1 in binding DNA sequences (Kagaya et al., 1999). In the ARF protein, the DNA-binding domain (B3 domain and Auxin-resp domain) is in the N-terminal region, while the protein-protein interaction domain (AUX_IAAs) is in the C-terminal portion (Wang et al., 2007). The promoter regions of B3 genes contain hormone-responsive elements such as ABA response elements (ABREs) and GA response elements (TCA motif and GARE sequences), and consistent some B3 TF genes are induced by hormone. The study on the ABA-dependent activation of the *CRC* (*CRUCIFERIN C*) promoter by the LEC1-NF-YC2 trimer indicates that ABRE motif is a necessary cis-element. Moreover, a seed-specific ABRE-binding protein, bZIP67, can functionally interact with LEC1-NF-YC2 to enhance the activation, while bZIP67 alone cannot activate the promoter. These results indicate that ABRE motif plays an important role in B3 protein-mediated regulation. (Curaba et al., 2004; Gazzarrini et al., 2004; Yamamoto et al., 2009).

The patterns of gene expression can serve as a critical indicator of the function of genes. In this study, some B3 genes exhibit tissue-specific expression patterns in *P. alba* × *P. glandulosa*: for example, *PagARF10-A2*, *PagARF16-A2*, *PagARF17-A1*, and *PagARF11-A2* were found preferentially expressed in leaves, and *PagVAL2-A1/A2/B1/B2* genes were preferentially expressed in xylems. In addition, 31 ARF genes were found to be expressed in all three tissues, among which nine were preferentially expressed in the differentiating xylem. Moreover, it seems that many ARFs function redundantly, thus single loss-of-function mutants do not exhibit growth and developmental phenotypes. In *Arabidopsis*, the *arf7 arf19* double mutant has a strong auxin-related phenotype, including severely impaired lateral root formation and abnormal gravitropism in both hypocotyls and roots, which are not observed in the *arf7* and *arf19* single mutants (Okushima et al., 2005).

The VAL family genes *PagVAL2-B1* and *PagVAL2-A1* were highly expressed in the differentiating xylem of two age trees. Its specific expression in the xylem is consistent with our previous identification that it is a putative upstream of *PagCAld5H2*. Although studies have shown the essential role of VAL genes in regulating the transition from seed maturation to seedling growth, few of them have been studied for their functions in the process of tree development. As one of the most environmentally cost-effective and renewable sources, poplar wood has been widely used for timbering, paper making, and many other commercial applications (Wang et al., 2013). In trees, the secondary growth of stem and the lignin biosynthesis gives rise to wood formation (Zheng et al., 2021; Luo and Li, 2022). To understand their function, we investigated the co-expression gene network analysis, focusing on the one key module related to wood formation we identified, followed by pathways enrichment analysis. Our results from pathways analyses indicated that the differentially expressed genes (DEGs) across tissues and their corresponding network genes are enriched in many different pathways. Among which, *PagVAL2-B1* and that corresponding network genes are enriched in regulation of SCWs biosynthesis and phenylpropanoid biosynthesis. This result suggests that *PagVAL2-B1* can positively regulate secondary growth and lignin biosynthesis in poplar.The members belonging to the plant-specific B3-domain transcription factor family exhibit significant and diverse roles, particularly concerning vegetative and reproductive growth. We performed a detailed genome-wide analysis of 160 B3 family members based on the HMMER search and BLASTP method in *P. alba* × *P. glandulosa.* Motif, domain, promoter, and gene structure analyses were conducted for further identification. Gene duplication events and evolutionary analyses of B3 members were performed to illustrate the relationship within the *P. alba* × *P. glandulosa* and seven species genomes. Among the four

B3 genes that were highly expressed in mature xylem, one (PagVAL2-B1), was studied for its co-expression relationships with other wood formation-related genes by WGCNA. Our analysis suggests that PagVAL2 family genes are involved in wood formation. Further using genetic transformation to study their specific roles in wood formation could help in the design to improve wood properties. Among the 160 B3 TF genes, 15% had incomplete coding region due to the annotation quality, and this may affect the gene structure analysis. The annotation needs further improvement to help the design in genetic transformation and CRISPR-based gene editing.

Data availability statement

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

Author contributions

MW, HL, and QL conceived the project. MW, HL, and QW performed the experiments. MW, HL, RL, and LY ran the program and visualized the data. MW wrote the draft, HL and QL edited the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32071787 to QL).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1193065/full#supplementary-material

References

Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *BioEssays* 33 (3), 189–202. doi: 10.1002/bies.201000107

Ahmad, B., Zhang, S., Yao, J., Chai, S., Yadav, V., Athar, H., et al. (2022). Ectopic expression of *VvFUS3*, B3-domain transcription factor, in tomato influences seed development *via* affecting endoreduplication and hormones. *Hortic. Plant J.* 8 (3), 351–360. doi: 10.1016/j.hpj.2020.12.009

Airoldi, C. A., and Davies, B. (2012). Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genomics* 39 (4), 157–165. doi: 10.1016/j.jgg.2012.02.008

Alvarez, J. P., Goldshmidt, A., Efroni, I., Bowman, J. L., and Esheda, Y. (2009). The NGATHA distal organ development genes are essential for style specification in. *Arabidopsis. Plant Cell* 21 (5), 1373–1393. doi: 10.1105/tpc.109.065482

Baucher, M., Chabbert, B., Pilate, G., Van Doorsselaere, J., Tollier, M. T., Petit-Conil, M., et al. (1996). Red xylem and higher lignin extractability by down-regulating a cinnamyl alcohol dehydrogenase in poplar. *Plant Physiol.* 112 (4), 1479–1490. doi: 10.1104/pp.112.4.1479

Baumbusch, L. O., Wayne, H. D., Galau, G. A., and Jakobsen, K. S. (2004). *LEC1*, *FUS3*, *ABI3* and *Em* expression reveals no correlation with dormancy in *Arabidopsis*. *J. Exp. Bot.* 55 (394), 77–87. doi: 10.1093/jxb/erh014

Borthakur, D., Busov, V., Cao, X. H., Du, Q., Gailing, O., Isik, F., et al. (2022). Current status and trends in forest genomics. *Forestry Res.* 2 (1), 0–0. doi: 10.48130/fr-2022-0011

Braun, E. L., and Grotewold, E. (1999). Newly discovered plant c-myb-like genes rewrite the evolution of the plant myb gene family. *Plant Physiol.* 121 (1), 21–24. doi: 10.1104/pp.121.1.21

Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4, 1–21. doi: 10.1186/1471-2229-4-10

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25 (15), 1972–1973. doi: 10.1093/bioinformatics/btp348

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8):1194–1202 . doi: 10.1016/j.molp.2020.06.009

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 (3), 840–856. doi: 10.1111/nph.16559

Chernomor, O., Von Haeseler, A., and Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65 (6), 997–1008. doi: 10.1093/sysbio/syw037

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 (3), 3660–3669. doi: 10.1104/pp.104.047266

Del Río, J. C., Gutiérrez, A., Hernando, M., Landín, P., Romero, J., and Martínez, Á.T. (2005). Determining the influence of eucalypt lignin composition in paper pulp yield using py-GC/MS. *J. Anal. Appl. Pyrolysis* 74 (1-2), 110–115. doi: 10.1016/j.jaap.2004.10.010

Dias, A. P., Braun, E. L., Mcmullen, M. D., Grotewold, E., Dias, A. P., Braun, E. L., et al. (2016). Recently duplicated maize R2R3 myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol.* 131 (2), 610–620. doi: 10.1104/pp.012047.610

Die, J. V., Gil, J., and Millan, T. (2018). Genome-wide identification of the auxin response factor gene family in *Cicer arietinum*. *BMC Genomics* 19 (1), 1–15. doi: 10.1186/s12864-018-4695-9

Franco-Zorrilla, J. M., Fernández-Calvín, B., Madueño, F., Cruzalvarez, M., Salinas, J., and Martínez-Zapater, J. M. (1999). Identification of genes specifically expressed in cauliflower reproductive meristems. molecular characterization of BoREM1. *Plant Mol. Biol.* 39 (3), 427–436. doi: 10.1023/A:1006130629100

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 (3), 373–385. doi: 10.1016/j.devcel.2004.06.017

Horn, D. R., Houston, M., and Hanrahan, P. (2005). In *ClawHMMER: A streaming HMMer-search implementation. SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing*, Seattle, WA, USA. 11–11. doi: 10.1109/SC.2005.18

Houb, E. B. (2001). The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* 2 (7), 516–527. doi: 10.1038/35080508

Hu, Y. X., Wang, Y. H., Liu, X. F., and Li, J. Y. (2004). Arabidopsis RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. *Cell Res.* 14 (1), 8–15. doi: 10.1038/sj.cr.7290197

Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res.* 27 (2), 470–478. doi: 10.1093/nar/27.2.470

Kim, S., Soltis, P. S., Wall, K., and Soltis, D. E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Mol. Biol. Evol.* 23 (1), 107–120. doi: 10.1093/molbev/msj014

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., et al. (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.* 19 (16), 1855–1860. doi: 10.1101/gad.1331305

Lapierre, C., Pollet, B., Petit-Conil, M., Toval, G., Romero, J., Pilate, G., et al. (1999). Structural alterations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid O-methyltransferase activity have an opposite impact on the efficiency of industrial kraft pulping. *Plant Physiol.* 119 (1), 153–163. doi: 10.1104/pp.119.1.153

Lee, H. L., and Irish, V. F. (2011). Gene duplication and loss in a MADS box gene transcription factor circuit. *Mol. Biol. Evol.* 28 (12), 3367–3380. doi: 10.1093/molbev/msr169

Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R., and Dean, C. (2002). Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. *Science* 297 (5579), 243–246. doi: 10.1126/science.1072147

Li, H., Chen, G., Pang, H., Wang, Q., and Dai, X. (2021a). Investigation into different wood formation mechanisms between angiosperm and gymnosperm tree species at the transcriptional and post-transcriptional level. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.698602

Li, L., Cheng, X., Lu, S., Nakatsubo, T., Umezawa, T., and Chiang, V. L. (2005). Clarification of cinnamoyl co-enzyme a reductase catalysis in monolignol biosynthesis of aspen. *Plant Cell Physiol.* 46 (7), 1073–1082. doi: 10.1093/pcp/pci120

Li, H., Dai, X., Huang, X., Xu, M., Wang, Q., Yan, X., et al. (2021b). Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in *Populus*. *J. Integr. Plant Biol.* 63 (11), 1906–1921. doi: 10.1111/jipb.13159

Li, X., Duan, X., Jiang, H., Sun, Y., Tang, Y., Yuan, Z., et al. (2006). Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and. *Arabidopsis. Plant Physiol.* 141 (4), 1167–1184. doi: 10.1104/pp.106.080580

Li, H. F., Ran, K., He, P., Wang, H. B., Chang, Y. S., Sun, Q. R., et al. (2015). Genome-wide identification and expression analysis of auxin response factor (ARF) gene family in apple. *J. Plant Physiol.* 51 (7), 1045–1054. doi: 10.13592/j.cnki.ppj.2015.0173

Li, H., Wen, X., Huang, X., Wei, M., Chen, H., Yu, Y., et al. (2022). Genome-wide identification and characterization of TCP gene family members in *Melastoma candidum*. *Molecules* 27 (24), 9036. doi: 10.3390/molecules27249036

Li, L., Zhou, Y., Cheng, X., Sun, J., Marita, J. M., Ralph, J., et al. (2003). Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. *Proc. Natl. Acad. Sci. USA* 100 (8), 4939–4944. doi: 10.1073/pnas.0831166100

Liscum, E., and Briggs, W. R. (1995). Mutations in the NPH1 locus of *Arabidopsis* disrupt the perception of phototropic stimuli. *Plant Cell* 7 (4), 473–485. doi: 10.2307/3870084

Luo, L., and Li, L. (2022). Molecular understanding of wood formation in trees. *Forestry Res.* 2 (1), 0–0. doi: 10.48130/fr-2022-0005

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., et al. (2008). Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol.* 148 (4), 1772–1781. doi: 10.1104/pp.108.124867

Magnani, E., Sjölander, K., and Hake, S. (2004). From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *Plant Cell* 16 (9), 2265–2277. doi: 10.1105/tpc.104.023135

Marella, H. H., Sakata, Y., and Quatrano, R. S. (2006). Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3-like genes from *Physcomitrella patens*. *Plant J.* 46 (6), 1032–1044. doi: 10.1111/j.1365-313x.2006.02764.x

McCarty, D. R., Hattori, T., Carson, C. B., Vasil, V., Lazar, M., and Vasil, I. K. (1991). The viviparous-1 developmental gene of maize encodes a novel transcriptional activator. *Cell* 66 (5), 895–905. doi: 10.1016/0092-8674(91)90436-3

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., et al. (2007). The chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318 (5848), 245–251. doi: 10.1126/science.1143609

Mitsuda, N., Seki, M., Shinozaki, K., and Ohme-Takagi, M. (2005). The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell* 17 (11), 2993–3006. doi: 10.1105/tpc.105.036004

Nicholas, K. B., and Nicholas, H. B. J. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments. *Biology*.

Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., et al. (2005). Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* 17 (2), 444–463. doi: 10.1105/tpc.104.028316

Palafox-Carlos, H., Contreras-Vergara, C. A., Muhlia-Almazán, A., Islas-Osuna, M. A., and González-Aguilar, G. A. (2014). Expression and enzymatic activity of phenylalanine ammonia-lyase and p-coumarate 3-hydroxylase in mango (*Mangifera indica 'Ataulfo'*) during ripening. *Genet. Mol. Res.* 13 (2), 3850–3858. doi: 10.4238/2014.May.16.10

Pekker, I., Alvarez, J. P., and Eshed, Y. (2005). Auxin response factors mediate *Arabidopsis* organ asymmetry *via* modulation of KANADI activity. *Plant Cell* 17 (11), 2899–2910. doi: 10.1105/tpc.105.034876

Peng, Y., Fang, T., Zhang, Y., Zhang, M., and Zeng, L. (2020). Genome-wide identification and expression analysis of auxin response factor (ARF) gene family in longan (*Dimocarpus longan* l.). *Plants* 9 (2), 221. doi: 10.3390/plants9020221

Peng, F. Y., and Weselake, R. J. (2013). Genome-wide identification and analysis of the B3 superfamily of transcription factors in brassicaceae and major crop plants. *Theor. Appl. Genet.* 126 (5), 1305–1319. doi: 10.1007/s00122-013-2054-4

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290 (5499), 2105–2110. doi: 10.1126/science.290.5499.2105

Romanel, E. A. C., Schrago, C. G., Couñago, R. M., Russo, C. A. M., and Alves-Ferreira, M. (2009). Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. *PloS One* 4 (6), e5791. doi: 10.1371/journal.pone.0005791

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., and Katoh, K. (2019). MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Res.* 47 (W1), W5–W10. doi: 10.1093/nar/gkz342

Ruegger, M., Dewey, E., Hobbie, L., Brown, D., Bernasconi, P., Turner, J., et al. (1997). Reduced naphthylphthalamic acid binding in the tir3 mutant of *Arabidopsis* is associated with a reduction in polar auxin transport and diverse morphological defects. *Plant Cell* 9 (5), 745–757. doi: 10.1105/tpc.9.5.745

Sarkanen, K. V., and Hergert, H. L. (1971). *Lignins: occurrence, formation, structure and reactions* (New York: Wiley-Interscience), 916.

Studer, M. H., DeMartini, J. D., Davis, M. F., Sykes, R. W., Davison, B., Keller, M., et al. (2011).). lignin content in natural *populus* variants affects sugar release. *Proc. Natl. Acad. Sci. USA* 108 (15), 6300–6305. doi: 10.1073/pnas.1009252108

Sulis, D. B., and Wang, J. P. (2020). Regulation of lignin biosynthesis by post-translational protein modifications. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00914

Sung, S., and Amasino, R. M. (2004). Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature* 427 (6970), 159–164. doi: 10.1038/nature02195

Suzuki, M., Kao, C. Y., and McCarty, D. R. (1997). The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. *Plant Cell* 9 (5), 799–807. doi: 10.1105/tpc.9.5.799

Swaminathan, K., Peterson, K., and Jack, T. (2008). The plant B3 superfamily. *Trends Plant Sci.* 13 (12), 647–655. doi: 10.1016/j.tplants.2008.09.006

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320 (5875), 486–488. doi: 10.1126/science.1153917

Tsukagoshi, H., Saijo, T., Shibata, D., Morikami, A., and Nakamura, K. (2005). Analysis of a sugar response mutant of *Arabidopsis* identified a novel B3 domain protein that functions as an active transcriptional repressor. *Plant Physiol.* 138 (2), 675–685. doi: 10.1104/pp.104.057752

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313 (5793), 1596–1604. doi: 10.1126/science.1128691

Ulmasov, T., Hagen, G., and Guilfoyle, T. J. (1997). ARF1, a transcription factor that binds to auxin response elements. *Science* 276 (5320), 1865–1868. doi: 10.1126/science.276.5320.1865

Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., et al. (2013). Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* 341 (6150), 1103–1106. doi: 10.1126/science.1241602

Waltner, J. K., Peterson, F. C., Lytle, B. L., and Volkman, B. F. (2005). Structure of the B3 domain from *Arabidopsis thaliana* protein At1g16640. *Protein Sci.* 14 (9), 2478–2483. doi: 10.1110/ps.051606305

Wang, Q., Dai, X., Pang, H., Cheng, Y., Huang, X., Li, H., et al. (2021). BEL1-like homeodomain protein BLH6a is a negative regulator of CAl5H2 in sinapyl alcohol monolignol biosynthesis in poplar. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.695223

Wang, Y., Deng, D., Zhang, R., Wang, S., Bian, Y., and Yin, Z. (2012). Systematic analysis of plant-specific B3 domain-containing proteins based on the genome resources of 11 sequenced species. *Mol. Biol. Rep.* 39 (5), 6267–6282. doi: 10.1007/s11033-012-1448-8

Wang, S., Li, L., Li, H., Sahu, S. K., Wang, H., Xu, Y., et al. (2020). Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nat. Plants* 6 (2), 95–106. doi: 10.1038/s41477-019-0560-3

Wang, J. P., Naik, P. P., Chen, H. C., Shi, R., Lin, C. Y., Liu, J., et al. (2014). Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in *Populus trichocarpa*. *Plant Cell* 26 (3), 894–914. doi: 10.1105/tpc.113.120881

Wang, D., Pei, K., Fu, Y., Sun, Z., Li, S., Liu, H., et al. (2007). Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene* 394 (1-2), 13–24. doi: 10.1016/j.gene.2007.01.006

Wang, H. H., Tang, R. J., Liu, H., Chen, H. Y., Liu, J. Y., Jiang, X. N., et al. (2013). Chimeric repressor of PtSND2 severely affects wood formation in transgenic *populus*. *Tree Physiol.* 33 (8), 878–886. doi: 10.1093/treephys/tpt058

Watahiki, M. K., and Yamamoto, K. T. (1997). The massugu1 mutation of *Arabidopsis* identified with failure of auxin-induced growth curvature of hypocotyl confers auxin insensitivity to hypocotyl and leaf. *Plant Physiol.* 115 (2), 419–426. doi: 10.1104/pp.115.2.419

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25 (9), 1189–1191. doi: 10.1093/bioinformatics/btp033

Wessels, B., Seyfferth, C., Escamez, S., Vain, T., Antos, K., Vahala, J., et al. (2019). An AP2/ERF transcription factor ERF139 coordinates xylem cell expansion and secondary cell wall deposition. *New Phytol.* 224 (4), 1585–1599. doi: 10.1111/nph.15960

Yamaguchi, M., Kubo, M., Fukuda, H., and Demura, T. (2008). Vascular-related NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in *Arabidopsis* roots and shoots. *Plant J.* 55 (4), 652–664. doi: 10.1111/j.1365-313x.2008.03533.x

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., Hattori, T., et al. (2009). *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J.* 58 (5), 843–856. doi: 10.1111/j.1365-313X.2009.03817.x

Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., et al. (2004). Solution structure of the B3 DNA binding domain of the *Arabidopsis* cold-responsive transcription factor RAV1. *Plant Cell* 16 (12), 3448–3459. doi: 10.1105/tpc.104.026112

Yamasaki, K., Kigawa, T., Inoue, M., Watanabe, S., Tateno, M., Seki, M., et al. (2008). Structures and evolutionary origins of plant-specific transcription factor DNA-binding domains. *Plant Physiol. Biochem.* 46 (3), 394–401. doi: 10.1016/j.plaphy.2007.12.015

Yu, Y., Ouyang, Y., and Yao, W. (2018). ShinyCircos: An R/Shiny application for interactive creation of circos plot. *Bioinformatics* 34 (7), 1229–1231. doi: 10.1093/bioinformatics/btx763

Zhang, Y., Gao, M., Singer, S. D., Fei, Z., Wang, H., and Wang, X. (2012). Genome-wide identification and analysis of the TIFY gene family in grape. *PloS One* 7 (9), e44465. doi: 10.1371/journal.pone.0044465

Zhao, Q., and Dixon, R. A. (2011). Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci.* 16 (4), 227–233. doi: 10.1016/j.tplants.2010.12.005

Zhao, L., Luo, Q., Yang, C., Han, Y., and Li, W. (2008). A RAV-like transcription factor controls photosynthesis and senescence in soybean. *Planta* 227 (6), 1389–1399. doi: 10.1007/s00425-008-0711-7

Zheng, S., He, J., Lin, Z., Zhu, Y., Sun, J., and Li, L. (2021). Two MADS-box genes regulate vascular cambium activity and secondary growth by modulating auxin homeostasis in *Populus*. *Plant Commun.* 2 (5), 100134. doi: 10.1016/j.xplc.2020.100134

Zheng-Hua, Y., and Zhong, R. (2015). Molecular control of wood formation in trees. *J. Exp. Bot.* 66 (14), 4119–4131. doi: 10.1093/jxb/erv081

Zhong, R., Mccarthy, R. L., Lee, C., and Ye, Z. H. (2011). Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. *Plant Physiol.* 157 (3), 1452–1468. doi: 10.1104/pp.111.181354

Zhong, R., Richardson, E. A., and Ye, Z. H. (2007). Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. *Planta* 225 (6), 1603–1611. doi: 10.1007/s00425-007-0498-y

Zhong, R., and Ye, D. Z. H. (2006). SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of *Arabidopsis*. *Plant Cell* 18 (11), 3158–3170. doi: 10.1105/tpc.106.047399

Zhou, J., Zhong, R., and Ye, Z.-H. (2014). *Arabidopsis* NAC domain proteins, VND1 to VND5, are transcriptional regulators of secondary wall biosynthesis in vessels. *PloS One* 9 (8), e105726. doi: 10.1371/journal.pone.0105726

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 June 2023
DOI 10.3389/fpls.2023.1196618

OPEN ACCESS

EDITED BY
Jiehua Wang,
Tianjin University, China

REVIEWED BY
Yinglang Wan,
Hainan University, China
Jinshan Gui,
Zhejiang Agriculture and Forestry University, China

*CORRESPONDENCE
Shu-Tang Zhao
shutang@caf.ac.cn

†These authors have contributed equally to this work

RECEIVED 30 March 2023
ACCEPTED 24 May 2023
PUBLISHED 09 June 2023

CITATION
Liu Y-L, Guo Y-H, Song X-Q, Hu M-X and Zhao S-T (2023) A method for analyzing programmed cell death in xylem development by flow cytometry. *Front. Plant Sci.* 14:1196618. doi: 10.3389/fpls.2023.1196618

A method for analyzing programmed cell death in xylem development by flow cytometry

Ying-Li Liu[1†], Ying-Hua Guo[2†], Xue-Qin Song[3], Meng-Xuan Hu[3] and Shu-Tang Zhao[3*]

[1]State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing, China, [2]National Center for Protein Sciences at Peking University, Beijing, China, [3]State Key Laboratory of Tree Genetics and Breeding, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China

Programmed cell death (PCD) is a genetically regulated developmental process leading to the death of specific types of plant cells, which plays important roles in plant development and growth such as wood formation. However, an efficient method needs to be established to study PCD in woody plants. Flow cytometry is widely utilized to evaluate apoptosis in mammalian cells, while it is rarely used to detect PCD in plants, especially in woody plants. Here, we reported that the xylem cell protoplasts from poplar stem were stained with a combination of fluorescein annexin V-FITC and propidium iodide (PI) and then sorted by flow cytometry. As expected, living cells (annexin V-FITC negative/PI negative), early PCD cells (annexin V-FITC positive/PI negative), and late PCD cells (annexin V-FITC positive/PI positive) could be finely distinguished through this method and then subjected for quantitative analysis. The expression of cell-type- and developmental stages-specific marker genes was consistent with the cell morphological observation. Therefore, the newly developed fluorescence-activated cell sorting (FACS) method can be used to study PCD in woody plants, which will be beneficial for studying the molecular mechanisms of wood formation.

KEYWORDS

PCD, xylem development, FACS, gene expression, woody plants

1 Introduction

Programmed cell death (PCD) is a highly regulated program in cell differentiation, and it widely exists in various biological processes. In most multicellular organisms, PCD is part of normal development and growth processes. Wood formation is a typical PCD process (Courtois et al., 2009; Jiang et al., 2021; Cao et al., 2022). The secondary xylem cells are derived from the vascular cambium, and the matured xylem includes xylem fiber cells with thick walls for mechanical support, vessel cells for vertical water transport, and ray cells for radial transport and storage of nutrients. One of the major developmental events of xylem

cell differentiation is the periclinal divisions of cambial meristem cells to produce xylem mother cells (Du et al., 2023). Xylem mother cells then undergo differentiation processes including cell elongation and expansion, secondary wall thickening and lignin deposition, PCD, and final autolysis to form matured secondary xylem cells (Du and Groover, 2010; He et al., 2022). After the concept of PCD was introduced into plant developmental biology, many studies have proved that the differentiation process of xylem tracheary element (TE) in woody plant is a typical PCD process (Pyo et al., 2007; Escamez and Tuominen, 2014).

Previous research had introduced several traditional methods to identify and quantify PCD progress in woody plants, such as microscopic observation, DNA ladder assay, and TUNEL detection. (Majtnerová and Roušar, 2018). However, microscopic examination can only provide the information of morphological changes of PCD cells. Although DNA ladder assay could determine the "DNA ladder" pattern of DNA fragments that occurred during the PCD progress, it can only be used for detecting PCD at a later stage when it is believed to have been ongoing (Majtnerová and Roušar, 2018). As for TUNEL, the *in situ* labeling method for detecting fragmented nuclear DNA to label the PCD process (Moore et al., 2021), there are also several shortcomings that are hard to overcome. First, although TUNEL is mostly used to detect cells with early apoptosis based on the DNA breakage that occurs in the early stage of apoptosis, false positive signals from necrotic cells cannot be totally avoided (Gorczyca et al., 1993). Meanwhile, DNA damage from other sources can also cause false-positive TUNEL signals. In addition, considering that TUNEL detection consists of many steps including tissue fixation, embedding, sectioning, and TUNEL reaction, the incorrect operation of any single step could also lead to false signals. Therefore, new methods need to be developed for the study of the PCD process in woody plants.

Flow cytometry, which allows the analysis of cells in suspension, has been developed in the identification and quantification of PCD (Weir et al., 2005). The cell shape, size, and granularity changes during the PCD process can be inferred from forward scatter (FSC) and side scatter (SSC). The FSC relates to the cell diameter, while the SSC relates to the inner cellular structures (McKinnon, 2018). The advantage of this method is the possibility of combining the scatter signals with fluorescence to allow for the identification of subgroup cells undergoing PCD. In the early stage of PCD, alterations occur at the cell surface: one of the alterations is the translocation of phosphatidylserine (PS) from the inner side of the plasma membrane to the outer layer, by which PS is exposed at the external surface of the cell (Dwyer et al., 2012). At this stage, cells became annexin V positive. In the late stage of PCD, the cell plasma membrane has been compromised, and the cell nuclei could be stained with PI (Nicoletti et al., 1991). Therefore, PI is widely used in conjunction with annexin V to determine if the cells undergoing PCD are in the early or late PCD process (Rieger et al., 2011). Accordingly, it is possible to use flow cytometry to sort cells at different stages during the PCD process based on fluorescence-activated cell sorting (FACS).

FACS is a powerful technique that is widely used for the isolation of cells based on fluorescence. Thousands of cells can be quantified and collected within just a couple of minutes, and different populations can be harvested simultaneously (Clark et al., 2022). FACS has been effectively used to characterize the transcriptome of *Arabidopsis* root cells (Birnbaum et al., 2005) and successfully used to isolate maize root endodermal cells (Ortiz-Ramírez et al., 2018). Recently, self-transcribing active regulatory region sequencing and transcriptomic analysis were performed in sorted cells using FACS in *Arabidopsis* and rice (Rich-Griffin et al., 2020; Tian et al., 2022). However, it is rare to detect gene expression alterations during the PCD process by using FACS in woody plants. The diversity of secondary cell wall thickness increases the difficulty of isolating viable cells in woody plants.

Here, we developed a new, simple, and reproducible method that uses developed and implemented flow cytometric techniques combined with microscopic observations and quantitative reverse transcription polymerase chain reaction (qRT-PCR) to analyze the cells in different stages of PCD in woody plants. The workflow is portrayed in Figure 1.

2 Materials and methods

2.1 Plant materials

The hybrid poplar *Populus alba×P. glandulosa* '84K' was used in this study. The plants were grown in a greenhouse for three months with long-day conditions (16 h light/8 h dark) at 25°C and with a light intensity of 80 mmol m^{-2} s^{-1} and 60% humidity.

2.2 Isolation of stem-differentiating xylem protoplasts

The isolation of protoplasts from stem-differentiating xylem was carried out as previously described (Lin et al., 2014; Liu et al., 2021) with minor modifications. Briefly, the stem below the 5th internode was separated into bark and wood and then cut into about 10 cm-long segments. The segments from wood tissues were submerged in freshly prepared cell wall digestion enzyme solution [1.5% (wt/vol) Cellulase R-10 (Yakult Pharmaceutical Industry, Japan) and 0.4% (wt/vol) pectolyase Y-23 (Yakult Pharmaceutical Industry, Japan) in 20 mM MES pH 5.7, 0.6 M mannitol, and 20 mM KCl solution, as well as 10 mM $CaCl_2$ and 0.1% (wt/vol) bovine serum albumin] for 40 minutes in the dark at room temperature. Next, the enzyme-digested stem segments were transferred into W5 solution (2 mM MES PH 5.7, 125 mM $CaCl_2$, 154 mM NaCl, 0.1 M glucose, and 5 mM KCl) and gently shaken to release protoplasts. Last, 40 μm filcons (BD Biosciences) were used to filter and collect protoplasts in the W5 solution.

2.3 Histological analyses

The stems used for protoplast isolation were sectioned before and after the enzyme was digested by using a vibratome (VT1000S;

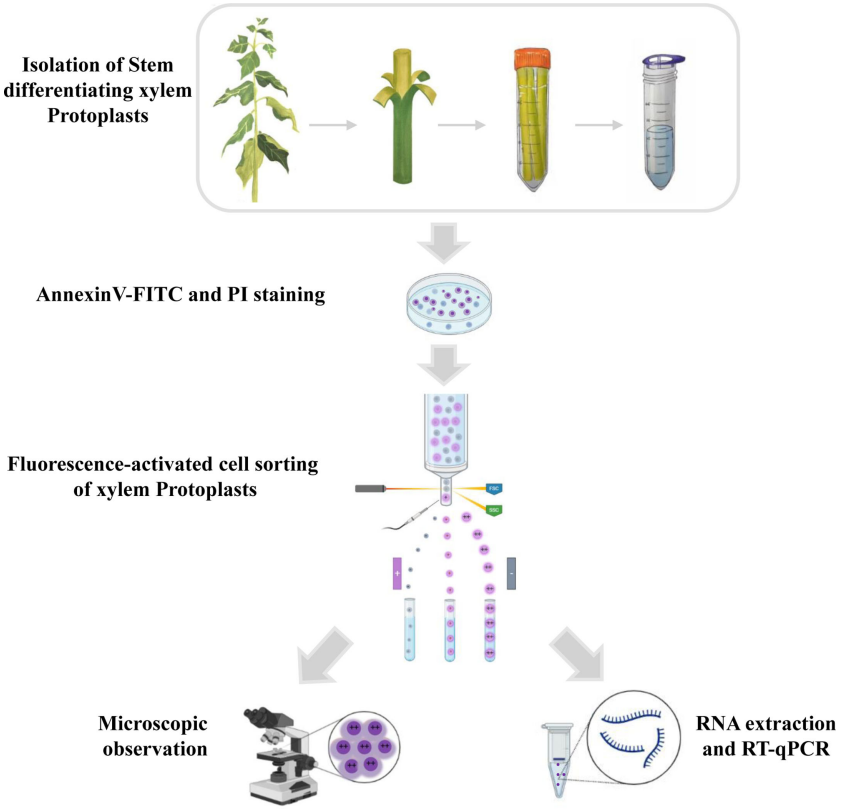


FIGURE 1
Schematic overview of the experiment for programmed cell death in xylem protoplasts using fluorescence-activated cell sorting method. The stems of the poplar were peeled off the bark and digested in an enzymatic solution to remove the cell walls. Protoplasts were then filtered through a cell strainer and collected for downstream analysis. The isolated protoplasts were stained with annexin V-FITC and PI. During FACS, cells can be separated based on different features such as the presence of a fluorophore, cell size, and cell shape. After sorting, different subpopulations of xylem cells were observed by microscope, and the RNA was isolated for transcription research.

Leica, Wetzlar, Germany) with a thickness of 40 μm. Then, the sections were stained with 0.05% toluidine blue for 60 s at room temperature, rinsed three times in water, and photographed with a microscope (IX73; Olympus; Japan).

2.4 Annexin V-FITC/PI staining

The isolated protoplasts were washed twice with W5 solution, centrifuged at 200 g for 5 min, and then resuspended in 200 μL W5 solution. Next, 5 μL annexin V-FITC (solarbio, Beijing, China, Cat.No. CA1020) was added and incubated on ice for 30 min. After incubation, the protoplasts were resuspended in 200 μL WI (0.6 M D-mannitol, 20 mM MES, 20 mM KCl, 0.1% bovine serum albumin) with 1 μL of the 10 mg mL^{-1} PI stock solution for 5 min.

2.5 Imaging flow cytometry analysis

The suspended protoplasts were analyzed with the imaging flow cytometer Image Stream^X Mark II (Amnis). Samples were acquired at 60× magnification. Bright-field (420-480 nm), annexin V-FITC (505-560 nm), and PI (595-642 nm) channels were measured, and at least 10,000 events of single cells per sample were collected. Single-color compensation controls were also acquired. Image analysis was performed with IDEAS version 6.0.

2.6 FACS of xylem protoplasts

The suspended protoplasts were cytometrically analyzed and sorted using a BD Aria SORP cell sorter (BD Biosciences, USA). The procedure and settings used were as in previous reports (Carqueijeiro et al., 2016). A 100 μm size nozzle and 20 psi sheath pressure were used with flow rates of 5,000 events/s. FSC and SSC were used for gating protoplasts from debris. FITC and PI fluorescence were set to separate negative, early PCD, and late PCD xylem cells, and 10,000 events were displayed for each plot.

2.7 Microscopic observation of sorted cells

After sorting, different subgroups of xylem cells were centrifuged at 200 g for 5 min. The supernatant was removed, and 10 μL suspension was evenly coated on the slide. All microscope observations and image acquisitions were performed using the Olympus IX73 (Japan) with 20× magnification. The cells before sorting were also photographed to observe the cell morphology.

2.8 RNA extraction and cDNA synthesis

Sorted cells were collected in microcentrifuge tubes containing RNA extraction buffer RTL with 2-mercaptoethanol. The total RNAs were extracted from cells from the Q2, Q3, and Q4 gates and referred to the method of Birnbaum (Birnbaum et al., 2005). The RNA was isolated using a RNeasy micro extraction kit and RNase-free DNase I set (Qiagen, Hilden, 74004, Germany) with some modifications (adjusting the RTL buffer to 3.5 times of the cell suspension and adjusting 2-mercaptoethanol to 10 µL per 1 mL RLT buffer). First-strand cDNA synthesis was carried out with approximately 300 ng RNA using the SuperScrip III first-strand synthesis system (TaKaRa, Dalian, RR047A, China) according to the manufacturer's instructions.

2.9 qRT-PCR

The genes related to the PCD process and vessel, fiber, ray, and cambium development were selected. Specific qRT-PCR primers were designed with melting temperatures of 58-60°C and amplicon lengths of 80-150 bp using Primer3 software (https://bioinfo.ut.ee/primer3-0.4.0/); these are given in Supplementary Table S1. qRT-PCR was performed in quadruplicate using the TB Green® Premix EX Taq™ II (TaKaRa, Dalian, RR820A, China) on a Roche lightCycler 480 (Roche Applied Science, Penzberg, Upper Bavaria, Germany) according to the manufacturer's instructions. The expression level of the genes was normalized to that of *PagActin* using Roche LightCycler advanced relative quantification analysis. The qPCR program started with a 30 s initial denaturation step at 95°C, followed by 40 cycles of amplification (95°C for 5 s, 60°C for 30 s) with continuous monitoring of the SYBR Green fluorescence. The reaction ended with a melting curve step from 55°C to 95°C at 0.5°C per second. All experiments were repeated at least three times with similar results.

3 Results

3.1 Harvesting protoplasts from differentiating xylem

To isolate protoplasts from hybrid poplar 84K, we used stems harvested from six 3-month-old 84K plantlets (Figure 2A). We then separated the bark and wood tissues of stems below the 5th internode and obtained free protoplasts from the differentiating xylem using a previously described digestion method (Lin et al., 2014; Liu et al., 2021). Stem sectioning showed that about 10 layers of differentiating xylem cells were released by enzyme digestion (Figure 2B). Most cells in the differentiating vascular tissue had been digested and released. The collected protoplasts were derived from differentiating fiber, vessel, and ray cells, which are the three major cell types that form wood (Tung et al., 2023), and they may also contain a small amount of cambium cells. The photograph of protoplasts harvested from differentiating xylem showed that the

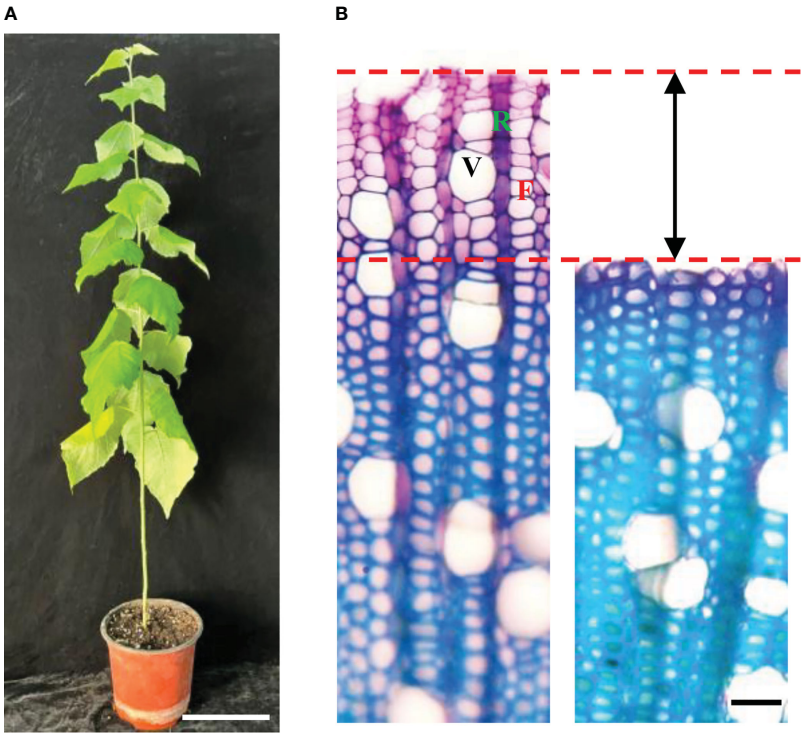


FIGURE 2
Differentiating xylem (DX) protoplast preparation. **(A)** Three-month-old poplar. (bar = 10 cm). **(B)** Cross-sections view of stems showing that about 10 layers of DX cells were released for the protoplast preparation. Left: before enzyme digestion; right: after enzyme digestion (bar = 20 µm). F, fiber; R, ray cell; V, vessel.

cell shape was round, and the cell membrane was complete, indicating that the cell vitality was good. The size of the protoplast cells was diverse, with cell diameters ranging from about 2 μm to 20 μm (Figure 3Ai).

3.2 Imaging flow cytometry-based analysis of xylem PCD cells

In order to determine the PCD of xylem cells based on cell size and fluorescence intensity, we employed the imaging flow cytometer for cell detection. Here, imaging flow cytometry analysis was carried out after annexin V-FITC and PI staining of the xylem cells. On the image stream, each cell had four images, namely, BF (bright field, channel 1), annexin V-FITC (false green, channel 2), PI (false red, channel 3), and annexin V-FITC/PI (channel 4). Cells from the first subgroup had a round morphology observed in the bright-field channel. In addition, they could be categorized as double negative because neither annexin V-FITC nor PI staining could be detected. The cell size of this subgroup was around 1-7 μm, and the cell membrane was complete, indicating that their cell activity was good (Figure 4A). Cells from the second subgroup appeared slightly shrunken with more complex morphologies under bright fields than cells from the first subgroup. Furthermore, annexin V-FITC staining could be detected in cells from this subgroup, indicating phosphatidylserine exposure. However, the PI signals of cells from

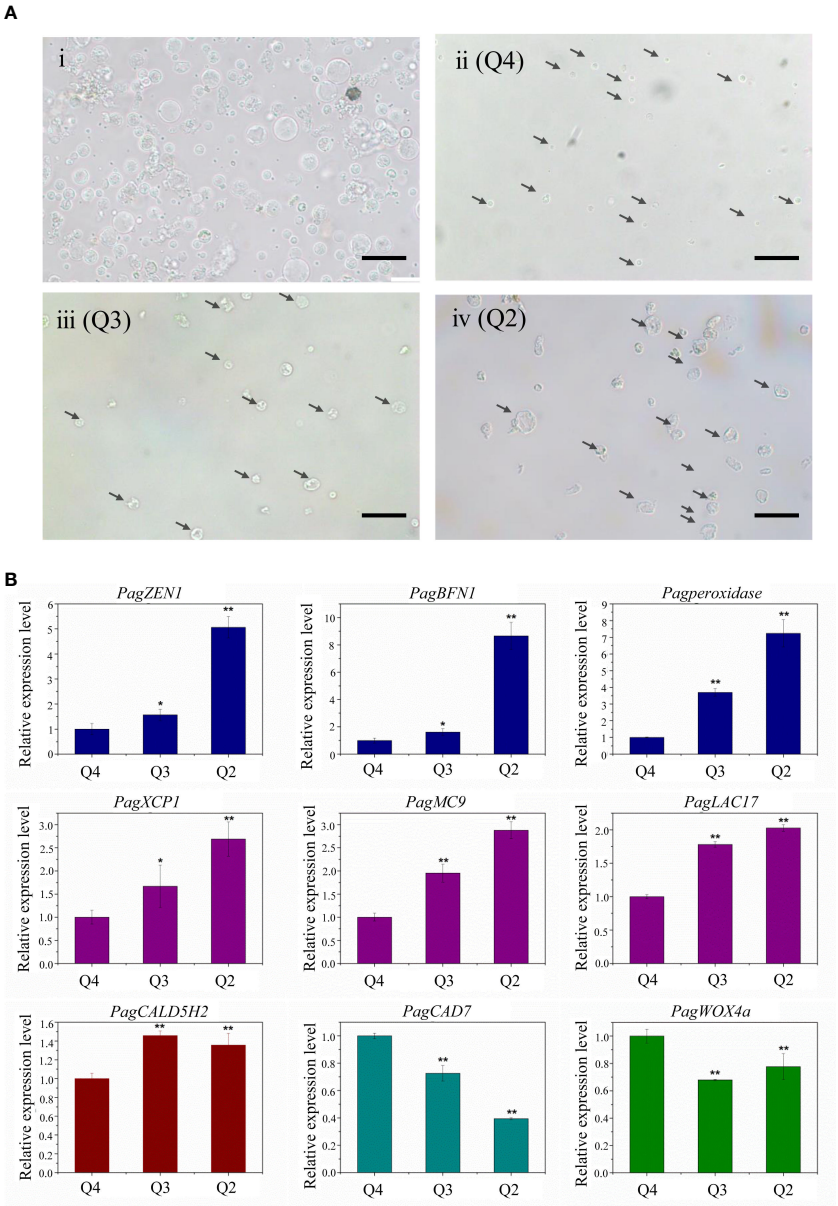


FIGURE 3
(A) Microscopic observation of different stages of xylem protoplast. (Ai) Xylem protoplast before fluorescence-activated cell sorting. (Aii) Annexin V-FITC and PI double negative healthy cells. (Aiii) Annexin V-FITC positive PCD cells. (Aiv) Annexin V-FITC and PI double positive PCD cells. (bar = 20 μm). **(B)** Expression levels of PCD and wood formation-related genes. Student's t-test; *P < 0.05; **P < 0.01.

this subgroup were negative. The cell size of this subgroup was about 8-10 µm. The annexin V-FITC positive and PI negative character of this subgroup suggest the cells were in the early PCD stage since those cells had intact membranes preventing PI staining (Figure 4B). Cells in the third subgroup were the largest, irregularly shaped with condensed, fragmented nuclei. The cell size of this subgroup was about 15 µm. Due to the membrane damage, PI could penetrate into the cells and thereby be detected, resulting in the cell membrane possessing a green FITC fluorescence signal, and the nucleus had a red PI fluorescence signal. Thus, these cells were double positive (Figure 4C). According to the staining results, it is speculated that these group cells were undergoing late PCD.

3.3 FACS and microscopic observation in different PCD stages

By staining isolated protoplasts with annexin V-FITC and PI, fluorescence-activated cell sorting (FACS) was used to distinguish and quantitatively determine the percentage of cells during different

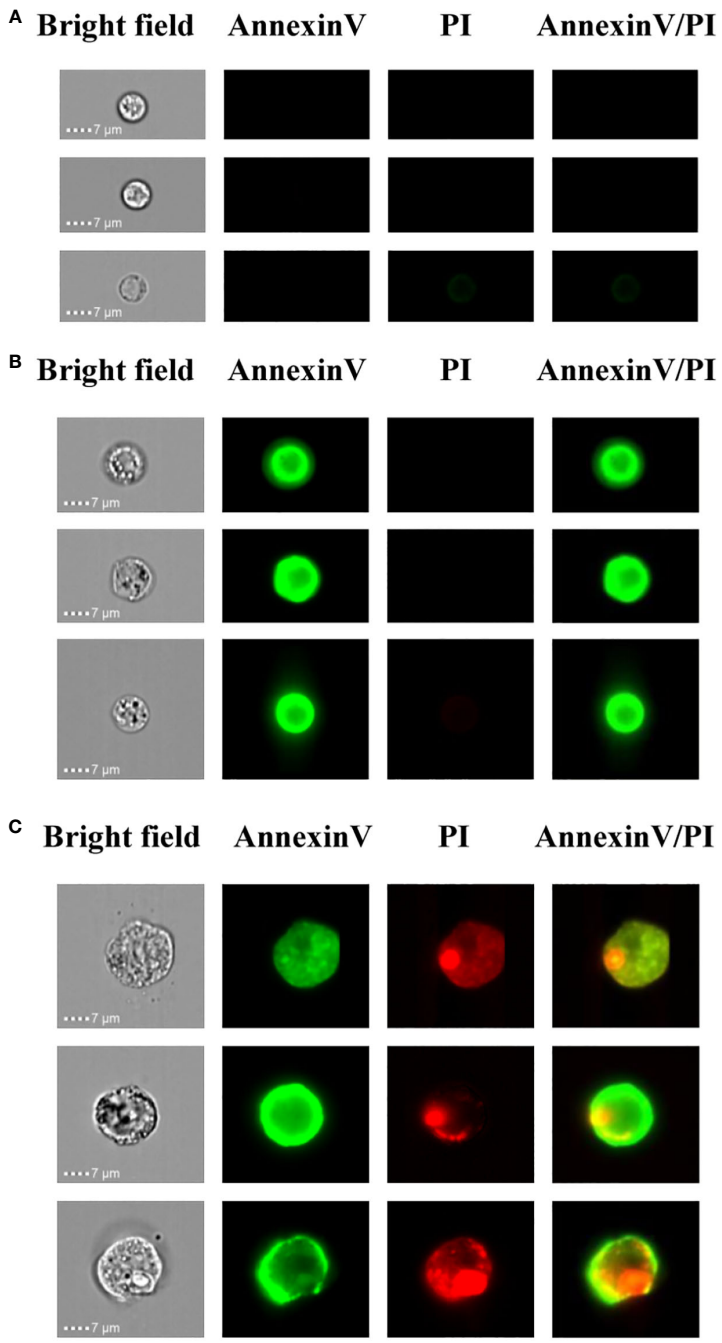


FIGURE 4
Imaging flow cytometry-based analysis of xylem PCD cells. **(A)** Annexin V-FITC and PI double negative healthy cells. **(B)** Annexin V-FITC positive and PI negative PCD cells. **(C)** Annexin V-FITC and PI double positive PCD cells.

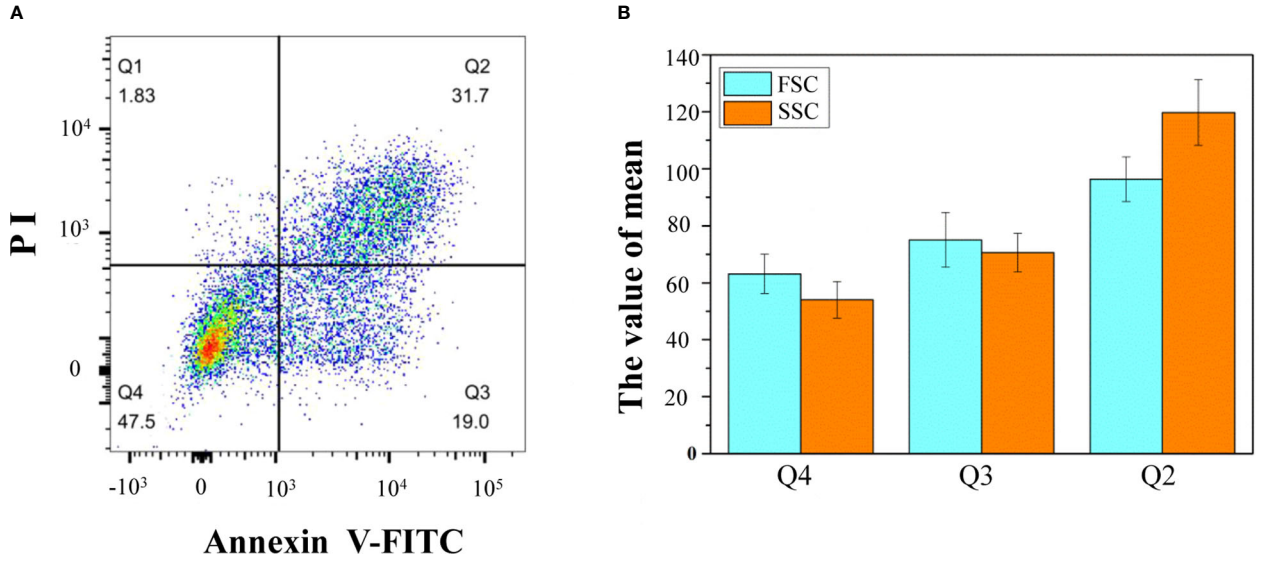


FIGURE 5
Flow cytometric analysis of annexin V-FITC/PI staining of differentiating xylem protoplast. **(A)** The upper-left quadrants (Q1) contain the annexin V-FITC negative and PI positive cells; the upper-right quadrants (Q2) contain the annexin V-FITC positive and PI positive cells; the lower-right quadrants (Q3) contain the annexin V-FITC positive and PI negative cells; the lower-left quadrants (Q4) contain the annexin V-FITC negative and PI negative cells; **(B)** the mean values of different types of cells. One representative experiment out of three is shown.

stages of the PCD process (Figure 5A). The whole-cell group was displayed in a dot plot with annexin V-FITC as abscissa and PI as ordinate. The Q1 gate (annexin V-FITC negative and PI positive, representing necrosis cells not undergoing PCD) accounted for 1.83%, which can be ignored. The Q2 gate annexin V-FITC positive and PI positive cells, accounting for 31.7% of the total cells, were judged as late PCD cells. The Q3 gate annexin V-FITC positive and PI negative cells, accounting for 19% of the total cells, were judged as early PCD cells. The Q4 gate annexin V-FITC negative and PI negative cells, accounting for 47.5% of the total cells, were judged as healthy cells. The FSC and SSC signals were used to measure cell size and complexity, respectively. A higher FSC signal represents apparently larger cells. A higher SSC signal means higher complexity as a consequence of cellular changes. In this study, according to the mean values of FSC and SSC, the healthy cells were the smallest, and the late apoptotic cells were the largest (Figure 5B).

Microscopic observation allows the visualization of biochemical and molecular changes associated with the PCD process and distinguishing morphologic changes of cells in different PCD stages. The microscopic observation of xylem protoplast cells after FACS showed that the morphological differences among different subgroups were distinct. The cells from Q4 were small and round, and its plasma membrane was integrity, which confirmed that these cells are healthy (Figure 3Aii). Accordingly, the cells of this subgroup corresponded to the annexin V-FITC negative/PI negative subgroup, as determined by imaging flow cytometry, which could be referred to as healthy cells. The cells from Q3, which correlated with the early PCD stage, were slightly larger. Although the cells shrank, their membrane remained intact, which is consistent with the characteristics of cells during the initial stages of PCD (Figure 3Aiii). These cells corresponded to the cells in the annexin V-FITC positive/PI negative subgroup detected in the imaging flow cytometry analysis. As for Q2, the cells were the largest, and an irregular shape with shrinkage, plasma membrane blebbing, cell detachment, and nuclear condensation were observed, indicating these cells were in the late PCD stage (Figure 3Aiv). These cells had the character of cells in the late PCD process and were consistent with the cells in the annexin V-FITC positive/PI positive subgroup in the imaging flow cytometry analysis. The fine consistency of the data from FACS and the image stream fluorescence analysis indicates that using cell imagery or the morphology of gated populations can clearly distinguish xylem cells before or in the early and late stage of the PCD process in woody plants.

3.4 Analysis of the expression of marker genes in different PCD stages

To test the feasibility of using the sorted cells for transcription studies, the sorted xylem cells were collected directly for RNA extraction. RNA from the Q2, Q3, and Q4 subgroups was isolated separately, with a yield of 300 ng of RNA from about 50,000 cells. qRT-PCR analysis of the genes involved in the PCD process and xylem development was carried out (Figure 3B). The expression of PCD-related genes (*PagZEN1*, *PagBFN1*, *PagPeroxidase*) showed the lowest expression in the Q4 subgroup and the highest in Q2 subgroup, corresponding well with the living and intact cells and cells in the late PCD process. During wood formation, the expression of genes related to vessel development (*PagXCP1*, *PagMC9*, *PagLAC17*) also increased with the occurrence of PCD. The fiber development-related gene *PagCAld5H2* was highly expressed in the Q3 subgroup, indicating early PCD. The expression of the ray parenchyma development-related *PagCAD7* gene was higher in the Q4 subgroup. The highest expression of the cambium development-related gene *PagWOX4a* was also found in the Q4 subgroup, which involved most living cells, indicating that

there was a small number of cambium cells in the stripped stem segments.

4 Discussion

Secondary xylem is the most abundant biomass produced by land plants and is important for global carbon sinks (Ye and Zhong, 2015). After the deposition of secondary walls, TEs and fibers undergo PCD, resulting in the degradation of their protoplasts (Escamez and Tuominen, 2014). The lack of efficient and accurate methods to detect PCD in woody plants has seriously hindered the progress of basic research in this field. Flow cytometry is a technique that enables the analysis of the population dynamics of cell development, the measurement of cell size and shape, and even the fluorescence characteristics of suitable labeled single cells (Weir et al., 2005). Here, we used flow cytometry to investigate each developmental stage in the developmental process during the PCD of xylem cell in woody plants.

The isolation of protoplasts by tissue digestion and cell wall removal without damaging the cell membrane was the crucial step for FACS analysis in the woody plant. In order to obtain a protoplast with good quality for xylem cell sorting, an optimized protocol was developed to digest cells for the production of a high yield of pure and stable xylem protoplasts from wood-forming tissues. After digestion, the isolated xylem protoplasts are generally larger in diameter (Figure 4, Figure 5A) than the mammalian cells (Antoniadi et al., 2022), which are osmosensitive and fragile, so the concentration of mannitol in the collected liquid was adjusted to keep the integrity of protoplasts. The obtained protoplasts involved fiber, vessel, and ray cells, three major cell types that form wood. They also contained a small number of cambium cells (Figure 2B). These different kinds of cells were successfully separated by FACS in our experiment (Figure 5).

To obtain xylem cells at specific developmental stages from wood-forming tissues for transcriptional studies, stained cells were isolated using FACS, and the total RNA was isolated for qRT-PCR analysis. Vessel elements are a specific type of TE that die after secondary cell wall thickening. ZEN1 is a key enzyme in the degradation of nuclear DNA during the PCD of TEs (Ito and Fukuda, 2002). *BEN1* encodes S1-type DNases that are associated with PCD in plants (Farage et al., 2011). In our results, the expression of *PagZEN1* and *PagBFN1* was the highest in the Q2 subgroup, indicating these cells were in the late PCD process. *Arabidopsis* xylem cysteine proteases XCPs were reported to accumulate in the vacuole and work in PCD-associated cell clearance during TE differentiation. *XCP1* was frequently used as the marker for xylogenesis and TE PCD and was specifically expressed in the vessel element (Avci et al., 2008; Liu et al., 2020). *PagXCP1* and *PagLAC17* were shown to be expressed at the later stage PCD of vessel formation in cluster 7 (Li et al., 2021), which was consistent with our results in this study. *AtMC9* is a xylem-specific metacaspase located in the apoplast and vacuole, and it has been shown to participate in regulating TE post-mortem autolysis (Bollhöner et al., 2013) and xylem cell death (Courtois et al., 2009). The expression of *PagMC9* was also increased with the occurrence of PCD. *CAld5H2* was identified as the fiber-specific marker in *P. trichocarpa* by laser microdissection (Shi et al., 2017), which was significantly upregulated in cluster 2 and involved in the S monolignol biosynthesis in fiber cells (Li et al., 2021). Our results showed that the expression of *PagCAld5H2* was more upregulated in Q3 than in Q2. *PagCAD7* was shown to be localized in the xylem parenchyma cells, providing lignin precursors to the adjacent vessels and fibers (Chen et al., 2021; Li et al., 2021). In our study, we detected that the expression of the *PagCAD7* gene was the highest in the healthy cell subgroups. qRT-PCR analysis of the genes involved in the PCD and different types of xylem cells further confirmed the suitability of the sorted cells for transcriptomic studies, which will benefit the study of transcriptional profiles from specific cell populations in woody plants.

We used imaging flow cytometry for the PCD research of woody plants for the first time. Imaging flow cytometry provides unique opportunities in the detecting and preparing images of individual cells that are being analyzed. The combination of microscopy and flow cytometry makes it possible to take images of xylem cells in bright fields and different fluorescence channels. It can obtain the images of several thousand cells with 20× magnification in a short period of time (Pietkiewicz et al., 2015). This study demonstrated the ability of the image stream flow cytometry to discriminate healthy cells and the early and late stages of PCD xylem cells. The developing xylem cells contain different types of cells, such as vessels, fibers, rays, which have distinct cell sizes. Combined with fluorescent staining at different stages of apoptosis, imaging flow cytometry can effectively distinguish them, which will be helpful for observations and statistics analyses of the PCD of different types of xylem cells in woody plants.

According to the analysis of the expression level of xylem developing related cell-type-specific marker genes, the cell morphology, and the imaging flow cytometry observation of the sorted cells, it could be speculated that the healthy cell subpopulations (Q4) may have included cambium cells, ray cells, and some fiber cells that had not undergone PCD. The subpopulations of the early PCD cells (Q3) were composed mostly of fiber cells. The subpopulations of late PCD cells (Q2) were composed mostly of vessel cells. This combination of classical flow cytometry with the microscopy method makes it possible to analyze the biochemical and morphological features of dying cells directly in one measurement, thereby allowing for the quantification of several thousand cells in a short period of time, together with an objective analysis of pictures by image-based features (George et al., 2004). Additionally, the staining procedure of this method is simple, and the results could be rapidly obtained and reproducible. Notably, this method facilitates the identification of gene expression levels for specific cells at different stages of PCD.

Flow cytometer in combination with the classical annexin V/PI staining allowed us to establish the method for distinguishing the early and late stage of xylem cells during PCD at the single-cell level. Compared to previous PCD determination by TUNEL at the organ level, our method exerts a better resolution to explore the xylem cells in the specific developmental process of PCD. Meanwhile, this method is not limited to poplar trees but is applicable to other woody plant materials. In addition, this method can also be used for the function characterization of PCD-related genes in transgenic materials.

5 Conclusion

We have developed a method using flow cytometry to rapidly analyze cells undergoing the PCD process and using FACS to sort different kinds of PCD cells in woody plants. This method was carried out using isolated xylem cells with annexin V-FITC and PI staining to reveal the PCD process. The cell morphology observed by microscopy and the expression level of cell-type-specific marker genes determined by qRT-PCR further confirmed the reliability of this method. This new method will facilitate the analysis of the PCD process of woody plants.

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

Y-LL: performing—the experiments and data curation; writing—original draft and editing. Y-HG: performing—the experiments and editing. X-QS: writing—reviewing and editing. M-XH: performing—the experiments. S-TZ: designing the study, writing—reviewing and editing, and funding acquisition. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2019YFE0119100).

Acknowledgments

We thank the Flow Cytometry Core at the National Center for Protein Sciences at Peking University for the help with the fluorescence-activated cell sorting analysis. We thank the Center for Quantitative Biology at Peking University for the assistance with the use of imaging flow cytometry, particularly Fei Wang for the technical help. We also thank for Mu-Yao Li for drawing the workflow of the whole protocol.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1196618/full#supplementary-material

SUPPLEMENTARY TABLE 1
List of primers used in this work.

References

Antoniadi, I., Skalický, V., Sun, G., Ma, W., Galbraith, D. W., Novák, O., et al. (2022). Fluorescence activated cell sorting-a selective tool for plant cell isolation and analysis. *Cytom. A.* 101, 725–736. doi: 10.1002/cyto.a.24461

Avci, U., Earl, H., Ismail, I. O., Beers, E. P., and Haigler, C. H. (2008). Cysteine proteases XCP1 and XCP2 aid micro-autolysis within the intact central vacuole during xylogenesis in arabidopsis roots. *Plant J.* 56, 303–315. doi: 10.1111/j.1365-313X.2008.03592.x

Birnbaum, K., Jung, J. W., Wang, J. Y., Lambert, G. M., Hirst, J. A., Galbraith, D. W., et al. (2005). Cell type-specific expression profiling in plants *via* cell sorting of protoplasts from fluorescent reporter lines. *Nat. Methods* 2, 615–619. doi: 10.1038/nmeth0805-615

Bollhöner, B., Zhang, B., Stael, S., Denancé, N., Overmyer, K., Goffner, D., et al. (2013). Post mortem function of AtMC9 in xylem vessel elements. *New Phytol.* 200, 498–510. doi: 10.1111/nph.12387

Cao, S., Guo, M., Cheng, J., Cheng, H., Liu, X., Ji, H., et al. (2022). Aspartic proteases modulate programmed cell death and secondary cell wall synthesis during wood formation in poplar. *J. Exp. Bot.* 73, 6876–6890. doi: 10.1093/jxb/erac347

Carqueijeiro, I., Guimarães, A. L., Bettencourt, S., Martínez, T., Guedes, J. G., Gardner, R., et al. (2016). Isolation of cells specialized in anticancer alkaloid metabolism by fluorescence-activated cell sorting. *Plant Physiol.* 171, 2371–2378. doi: 10.1104/pp.16.01028

Chen, Y., Tong, S., Jiang, Y., Ai, F., Feng, Y., Zhang, J., et al. (2021). Transcriptional landscape of highly lignified poplar stems at single-cell resolution. *Genome Biol.* 22, 319. doi: 10.1186/s13059-021-02537-2

Clark, N. M., Elmore, J. M., and Walley, J. W. (2022). To the proteome and beyond: advances in single-cell omics profiling for plant systems. *Plant Physiol.* 188, 726–737. doi: 10.1093/plphys/kiab429

Courtois, C. L., Pesquet, E., Sjödin, A., Muñiz, L., Bollhöner, B., Kaneda, M., et al. (2009). A unique program for cell death in xylem fibers of populus stem. *Plant J.* 58, 260–274. doi: 10.1111/j.1365-313X.2008.03777.x

Du, J., and Groover, A. (2010). Transcriptional regulation of secondary growth and wood formation. *J. Integr. Plant Biol.* 52, 17–27. doi: 10.1111/j.1744-7909.2010.00901.x

Du, J., Wang, Y. C., Chen, W., Xu, M., Zhou, R., Shou, H., et al. (2023). High-resolution anatomical and spatial transcriptome analyses reveal two types of meristematic cell pools within the secondary vascular tissue of poplar stem. *Mol. Plant* 9, S1674–2052(23)00069-2. doi: 10.1016/j.molp.2023.03.005

Dwyer, D. J., Camacho, D. M., Kohanski, M. A., Callura, J. M., and Collins, J. J. (2012). Antibiotic-induced bacterial cell death exhibits physiological and biochemical hallmarks of apoptosis. *Mol. Cell.* 46, 561–572. doi: 10.1016/j.molcel.2012.04.027

Escamez, S., and Tuominen, H. (2014). Programmes of cell death and autolysis in tracheary elements: when a suicidal cell arranges its own corpse removal. *J. Exp. Bot.* 65, 1313–1321. doi: 10.1093/jxb/eru057

Farage, S., Burd, S., Sonego, L., Mett, A., Belausov, E., Gidoni, D., et al. (2011). Localization of the arabidopsis senescence- and cell death-associated BFN1 nuclease: from the ER to fragmented nuclei. *Mol. Plant* 4, 1062–1073. doi: 10.1093/mp/ssr045

George, T. C., Basiji, D. A., Hall, B. E., Lynch, D. H., Ortyn, W. E., Perry, D. J., et al. (2004). Distinguishing modes of cell death using the ImageStream multispectral imaging flow cytometer. *Cytom. A.* 59, 237–245. doi: 10.1002/cyto.a.20048

Gorczyca, W., Gong, J., and Darzynkiewicz, Z. (1993). Detection of DNA strand breaks in individual apoptotic cells by the *in situ* terminal deoxynucleotidyl transferase and nick translation assays. *Cancer Res.* 53, 1945–1951. doi: 10.1016/0304-3835(93)90034-7

He, H., Song, X. Q., Jiang, C., Liu, Y. L., Wang, D., Wen, S. S., et al. (2022). The role of senescence-associated gene101 (PagSAG101a) in the regulation of secondary xylem formation in poplar. *Joural Integr. Plant Biol.* 64, 73–86. doi: 10.1111/jipb.13195

Ito, J., and Fukuda, H. (2002). ZEN1 is a key enzyme in the degradation of nuclear DNA during programmed cell death of tracheary elements. *Plant Cell.* 14, 3201–3211. doi: 10.1105/tpc.006411

Jiang, C., Wang, J., Leng, H. N., Wang, X., Liu, Y., Lu, H., et al. (2021). Transcriptional regulation and signaling of developmental programmed cell death in plants. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.702928

Li, H., Dai, X. R., Huang, X., Xu, M. X., Wang, Q., Yan, X. J., et al. (2021). Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in populus. *J. Integr. Plant Biol.* 63, 1906–1921. doi: 10.1111/jipb.13159

Lin, Y. C., Li, W., Chen, H., Li, Q., Sun, Y. H., Shi, R., et al. (2014). A simple improved-throughput xylem protoplast system for studying wood formation. *Nat. Protoc.* 9, 2194–2205. doi: 10.1038/nprot.2014.147

Liu, Y. L., Wang, L. J., Li, Y., Guo, Y. H., Cao, Y., and Zhao, S. T. (2021). A small guanosine triphosphate binding protein PagRabE1b promotes xylem development in poplar. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.686024

Liu, P. L., Zhang, X., Mao, J. F., Hong, Y. M., Zhang, R. G., Yilan, E., et al. (2020). The tetracentron genome provides insight into the early evolution of eudicots and the formation of vessel elements. *Genome Biol.* 21, 291. doi: 10.1186/s13059-020-02198-7

Majtnerová, P., and Roušar, T. (2018). An overview of apoptosis assays detecting DNA fragmentation. *Mol. Biol. Rep.* 45, 1469–1478. doi: 10.1007/s11033-018-4258-9

McKinnon, K. M. (2018). Flow cytometry: an overview. *Curr. Protoc. Immunol.* 120, 5.1.1–5.1.11. doi: 10.1002/cpim.40

Moore, C. L., Savenka, A. V., and Basnakian, A. G. (2021). TUNEL assay: a powerful tool for kidney injury evaluation. *Int. J. Mol. Sci.* 22, 412. doi: 10.3390/ijms22010412

Nicoletti, I., Migliorati, G., Pagliacci, M. C., Grignani, F., and Riccardi, C. (1991). A rapid and simple method for measuring thymocyte apoptosis by propidium iodide staining and flow cytometry. *J. Immunol. Methods* 139, 271–279. doi: 10.1016/0022-1759(91)90198-o

Ortiz-Ramírez, C., Arevalo, E. D., Xu, X., Jackson, D. P., and Birnbaum, K. D. (2018). An efficient cell sorting protocol for maize protoplasts. *Curr. Protoc. Plant Biol.* 3, e20072. doi: 10.1002/cppb.20072

Pietkiewicz, S., Schmidt, J. H., and Lavrik, I. N. (2015). Quantification of apoptosis and necroptosis at the single cell level by a combination of imaging flow cytometry with classical annexin v/propidium iodide staining. *J. Immunol. Methods* 423, 99–103. doi: 10.1016/j.jim.2015.04.025

Pyo, H., Demura, T., and Fukuda, H. (2007). TERE; a novel cis-element responsible for a coordinated expression of genes related to programmed cell death and secondary wall formation during differentiation of tracheary elements. *Plant J.* 51, 955–965. doi: 10.1111/j.1365-313X.2007.03180.x

Rich-Griffin, C., Eichmann, R., Reitz, M. U., Hermann, S., Woolley-Allen, K., Brown, P. E., et al. (2020). Regulation of cell type-specific immunity networks in arabidopsis roots. *Plant Cell.* 32, 2742–2762. doi: 10.1105/tpc.20.00154

Rieger, A. M., Nelson, K. L., Konowalchuk, J. D., and Barreda, D. R. (2011). Modified annexin v/propidium iodide apoptosis assay for accurate assessment of cell death. *J. Vis. Exp.* 24, 2597. doi: 10.3791/2597

Shi, R., Wang, J. P., Lin, Y. C., Li, Q., Sun, Y. H., Chen, H., et al. (2017). Tissue and cell-type co-expression networks of transcription factors and wood component genes in populus trichocarpa. *Planta* 245, 927–938. doi: 10.1007/s00425-016-2640-1

Tian, W., Huang, X., and Ouyang, X. (2022). Genome-wide prediction of activating regulatory elements in rice by combining STARR-seq with FACS. *Plant Biotechnol. J.* 20, 2284–2297. doi: 10.1111/pbi.13907

Tung, C. C., Kuo, S. C., Yang, C. L., Yu, J. H., Huang, C. E., Liou, P. C., et al. (2023). Single-cell transcriptomics unveils xylem cell development and evolution. *Genome Biol.* 24, 3. doi: 10.1186/s13059-022-02845-1

Weir, I. E., Maddumage, R., Allan, A. C., and Ferguson, I. B. (2005). Flow cytometric analysis of tracheary element differentiation in zinnia elegans cells. *Cytom. A.* 68 (2), 81–91. doi: 10.1002/cyto.a.20194

Ye, Z. H., and Zhong, R. (2015). Molecular control of wood formation in trees. *J. Exp. Bot.* 66, 4119–4131. doi: 10.1093/jxb/erv081

frontiers | Frontiers in Plant Science

TYPE Data Report
PUBLISHED 17 August 2023
DOI 10.3389/fpls.2023.1218515

OPEN ACCESS

EDITED BY
Quanzi Li,
Chinese Academy of Forestry, China

REVIEWED BY
Yong Li,
Henan Agricultural University, China
Su Chen,
Northeast Forestry University, China

*CORRESPONDENCE
Huan Liu
liuhuan@genomics.cn
Chengzhong He
hecz@swfu.edu.cn

†These authors have contributed equally to this work

RECEIVED 07 May 2023
ACCEPTED 27 July 2023
PUBLISHED 17 August 2023

CITATION
Sahu SK, Liu M, Li R, Chen Y, Wang G, Fang D, Sahu DN, Wei J, Wang S, Liu H and He C (2023) Chromosome-scale genome of Indian rosewood (*Dalbergia sissoo*). *Front. Plant Sci.* 14:1218515. doi: 10.3389/fpls.2023.1218515

Chromosome-scale genome of Indian rosewood (*Dalbergia sissoo*)

Sunil Kumar Sahu[1†], Min Liu[1,2†], Ruirui Li[1,3†], Yewen Chen[1], Guanlong Wang[1,4], Dongming Fang[1], Durgesh Nandini Sahu[1], Jinpu Wei[1], Sibo Wang[1], Huan Liu[1,2*] and Chengzhong He[5*]

[1]State Key Laboratory of Agricultural Genomics, Key Laboratory of Genomics, Ministry of Agriculture, BGI Research, Shenzhen, China, [2]BGI Life Science Joint Research Center, Northeast Forestry University, Harbin, China, [3]College of Life Sciences, Chongqing Normal University, Chongqing, China, [4]College of Science, South China Agricultural University, Guangzhou, China, [5]Key Laboratory for Forest Genetic & Tree Improvement and Propagation in Universities of Yunnan Province, Southwest Forestry University, Kunming, China

KEYWORDS

genome, chromosome, *Dalbergia*, timber tree, comparative genomics, rosewood, phylogeny

Introduction

Dalbergia sissoo (Leguminosae) is a large, deciduous tree native to the Indian subcontinent and southern Iran, and is widely distributed in the tropics (Bhagwat et al., 2015), typically reaching heights of 15-25 meters, with a maximum height of up to 30 meters (Al-Snafi, 2017). It is also known as North Indian rosewood or sheesham. It has a long, straight trunk and a dense, rounded crown. The leaves are compound, with 5-9 leaflets. The flowers are small and white or pink. The fruit is a pod that contains 1-2 seeds. *D. sissoo* is an economically important timber plant and a very useful multipurpose tree, with high-value wood that is hard, heavy, strong, durable, elastic, weather-resistant, and rot-resistant. It is also a valuable medicinal herb and high-grade spice (Son and Manh Ha, 2022). Many *Dalbergia* species are currently receiving international attention for conservation due to the overexploitation of their valuable heartwood (Hung et al., 2020). It is listed as a Least Concern species on the International Union for Conservation of Nature (IUCN) Red List and is protected by the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). A metabolic pathway enrichment analysis of *D. sissoo* stem extracts has shown that the differential metabolites are mainly enriched in three metabolic pathways: flavonoid biosynthesis, isoflavonoid biosynthesis, and flavonol and flavone biosynthesis. These pathways are involved in the production of flavonoids, which are a group of plant compounds with a wide range of biological activities, including antioxidant, anti-inflammatory, and antimicrobial properties. The formation and accumulation of heartwood components have always been a hot topic in the study of heartwood formation. Most of these heartwood components are extracts or secretions produced from metabolism and degradation of certain tissues. They give the heartwood a unique appearance and exhibit resistance to pathogenic fungi as potential biological

protectants (Li et al., 2022). However, there is little overall understanding of the molecular basis of *D. sissoo* secondary metabolite biosynthesis pathways, which hinders progress in molecular breeding and revealing its heartwood formation mechanism. It's also worth mentioning that research in forest tree genomics has made significant progress over the past two decades, which has provided more information to understand the genetic basis of traits such as wood formation, growth and the role of secondary metabolites (Isabel et al., 2020; Sahu et al., 2023).

Genome sequences of many woody plants have been published, such as *Juglans regia* (Marrano et al., 2020), *Liriodendron chinense* (Chen et al., 2019), *Eucalyptus grandis* (Myburg et al., 2014), *Dalbergia odorifera* (Hong et al., 2020), *Dipterocarpus turbinatus* (Wang et al., 2022), *Tectona grandis* (Yasodha et al., 2018), *Populus trichocarpa* (Djerbi et al., 2005) and *Picea abies* (Nystedt et al., 2013), providing the basis for further research into the molecular mechanisms regulating wood development and quality (Sahu and Liu, 2023). In this study, we successfully assembled a 661.00 Mb *D. sissoo* genome based on 214.42 Gb of 10x Genomics next-generation sequencing data, with a scaffold N50 of 7.17 Mb. Subsequently, by combining 139.93 Gb of Hi-C data, 99.9% of the Scaffolds were anchored to 10 pseudochromosomes. Chromosome-scale genome assembly will promote our understanding of the fast-growing characteristics and evolution of *D. sissoo* and facilitate the revelation of its molecular breeding and wood formation mechanisms.

Results

Genome assembly

To assemble the chromosome-level genome of *D. sissoo*, we generated 214.42 Gb of 10x Genomics sequencing data and 139.93 Gb of high-throughput chromatin conformation capture (Hi-C) sequencing data (Supplementary Table S1). The estimated genome size based on GCE (genomic charactor estimator) analysis was 756.56 Mb. The heterozygosity was 0.68%, indicating a moderately high level of heterozygosity, and the repeat content was estimated to be 48.06% (Supplementary Table S2), suggesting that the *D. sissoo* genome may contain a large number of repeat sequences. Finally, Supernova (v2.1.1) successfully assembled the *D. sissoo* genome into a 661.00 Mb draft, with scaffold N50 size of 7.17 Mb. Subsequently, by combining 139.93 Gb of Hi-C data, 96.6% of scaffolds were anchored to 10 pseudochromosomes (Figure 1A), resulting in a chromosome-level genome size of 660.37 Mb, with a scaffold N50 of 56.15 Mb (Supplementary Table S3). BUSCO evaluation showed 96.6% completeness, with 1329 genes aligned out of 1375 core genes, including 1282 complete single-copy BUSCOs and 47 complete duplicated BUSCOs. The BUSCO evaluation for the chromosome-level genome was 96.2% (Supplementary Table S4-1). DNA data alignment to the genome showed an alignment rate of 98.54%, and the GC content of the genome was 33.59% (Table 1), within the normal range of plant genome GC content, ruling out the possibility of bacterial contamination. Therefore, the assembled *D. sissoo* genome has high BUSCO completeness, high data alignment rate, and a normal GC distribution range, indicating a high-quality and contiguity (Figure 1B).

Genome annotation

Based on *de novo* prediction and comparison with known repeat databases, repeat sequences accounted for 52.8% of the genome, with LTR types being the most common (Supplementary Table S5). Structural annotation identified 29,737 protein-coding genes in the genome, with a BUSCO evaluation result of 95.40% (Supplementary Table S6), indicating high completeness of the predicted gene set. The functional annotation revealed that 28,689 genes had functional information, which represents 96.48% (Supplementary Table S7). In addition, noncoding RNA in the *D. sissoo* genome was predicted, and a total of 3,848 ncRNAs were identified, with a sequence length of 0.48 Mb, accounting for 0.07% of the genome. Among them, 123 miRNAs, 1,141 tRNAs, 1,812 rRNAs, and 770 snRNAs were identified (Supplementary Table S8, Table 2).

Phylogenetic analysis and species divergence time estimation

To determine the evolutionary relationships between *D. sissoo* and other species, we identified 108 single-copy orthologs using OrthoFinder (v2.3.1) from 11 representative plant species (Supplementary Table S9). The protein sequence alignment of these single-copy orthologs was clustered using the MCL algorithm (v14-137) included in OrthoFinder and used to generate a phylogenetic tree with *Vitis vinifera* as the outgroup. The phylogenetic tree showed that *D. sissoo* first diverged from the common ancestors of *Arabidopsis thaliana* and *P. trichocarpa*, and then later from *Malus domestica*. The closest evolutionary relationship of *D. sissoo* was with *D. odorifera*, estimated to have diverged about 14.3 million years ago (MYA) (Figure 2A).

Gene family expansion and contraction

Gene family evolution analysis using CAFE on the 11 plant species showed that *D. sissoo* had 749 significantly expanded and 1470 significantly contracted gene families. KEGG enrichment analysis of the expanded gene families revealed that most of the genes were related to secondary metabolites such as flavonoids, terpenoids, and ABC transporters. In addition, genes related to photosynthesis were also expanded significantly (Supplementary Figure 3A). The GO enrichment analysis of the expansion gene family of *D. sissoo* also enriched many photosynthesis-related terms such as photosynthesis, light reaction, photosystem I, photosystem II reaction center, and chlorophyll II binding, and so on (Supplementary Figure 3B), suggesting that the fast-growing characteristics of *D. sissoo* are the result of the expansion of a large number of photosynthesis-related genes. The contracted gene

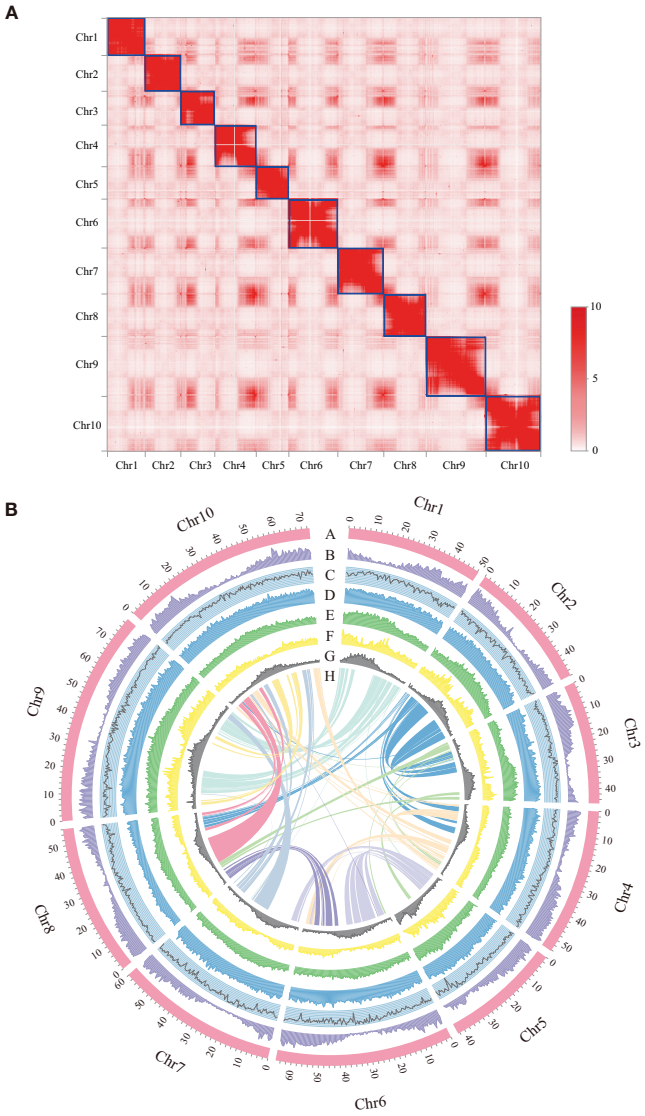


FIGURE 1
Features of *Dalbergia sissoo* genome. **(A)** Hi-C map of the *D sissoo* genome showing genome-wide all-by-all interactions. The map shows a high resolution of individual chromosomes that are scaffolded and assembled independently. The heat map colors ranging from light pink to dark red indicate the frequency of Hi-C interaction links from low to high (0–10). **(B)** Circos plot of *D sissoo* genome. Concentric circles from outermost to innermost, show (A) chromosomes and megabase values, (B) gene density, (C) GC content, (D) repeat density, (E) LTR density, (F) LTR *Copia* density, (G) LTR *Gypsy* density and (H) inter-chromosomal synteny (feature B-G are calculated in non-overlapping 500 Kb sliding windows).

families were significantly enriched in 10 KEGG pathways, including vitamin B6 metabolism, phenylpropanoid biosynthesis, cutin, suberin and wax biosynthesis, and the citric acid cycle. (Supplementary Figure 3C).

Whole-genome duplication analysis

Here, we employed wgd software and compared the relative genomic CDS sequences of *D. odorifera*, *Glycine Max* (Schmutz et al., 2010) and *Medicago ruthenica* (Yin et al., 2021) to calculate Ks values, and the occurrence of whole genome duplication (WGD) events was judged by the distribution of Ks values (Figure 2B). It was found that *D. sissoo*, *D. odorifera*, *G. Max* and *M. ruthenica* experienced a WGD event common to legumes, from which we inferred that the WGD event of *D. sissoo* occurred close to the first WGD event of *G. Max* (~59 Mya) (Schmutz et al., 2010), followed by another independent WGD event of the genus *Glycine* in which soybean is located ~13 Mya. No independent WGD event was observed for the genus *Dalbergia*. The distribution of four gene types - single-copy, multiple-copy, unique, and other genes - across various tree species are displayed in Figure 2C.

Identification of gene duplication types

The number of genes in the *D. sissoo* genome for whole genome duplication (WGD), tandem duplication (TD), proximal

TABLE 1 Statistics of *D. sissoo* genome assembly and assessment.

Assembly		*Dalbergia sissoo*
Genome-sequencing depth(X)	10X genomics sequencing (Gb)	214.42
	Hi-C (Gb)	139.93
Estimated genome size (Mb)		756.56
Estimated heterozygosity (%)		0.68
Assembly size (Mb)		661.00
GC content (%)		33.59
Scaffold N50 (Kb)		7165.92
BUSCO completeness of assembly (%)		96.6
Complete single-copy BUSCO (%)		93.2
Complete duplicated BUSCO (%)		3.4
Total length of pseudochromosome assembly (Mb)		660.37
Pseudochromosome number		10
Scaffold N50 of pseudochromosome assembly (Kb)		56151.84
BUSCO completeness of pseudochromosome assembly (%)		96.2
The rate of pseudochromosome anchored genome (%)		99.9

duplication (PD), and transposition duplication (TRD), and dispersed duplication (DD) types were 6522, 800, 607, 2895 and 4500, respectively. The predicted duplication types were compared with the expanded gene family, and five replication types had 578, 223, 183, 311, and 734 genes in the expanded gene family, totaling 2029, respectively (Supplementary Figure 2A). Among all the WGD and DD type genes contributed the most to the expansion of the *D. sissoo* gene family.

TABLE 2 Genome annotation of *D. sissoo*.

Annotation	*Dalbergia sissoo*
Number of predicted protein-coding genes	29740
Average gene length (bp)	4135.19
Average exon length (bp)	222.06
Average exon number per gene	5.47
Average intron length (bp)	654.18
miRNAs	123
rRNAs	906
tRNAs	1141
Percentage of repeat sequencing (%)	52.8
LTR *Copia* (%)	8.92
LTR *Gypsy* (%)	27.63
LINE (%)	1.29
SINE (%)	0.02
DNA transposons (%)	7.84
Percentage of functional annotation genes (%)	96.48

Discussion

Dalbergia sissoo, commonly known as Indian rosewood, is an ecologically and economically vital tropical timber species (Bhagwat et al., 2015). However, a lack of genomic resources has constrained genetic research in *D. sissoo*, including studies of heartwood formation, a key trait for timber quality (Hong et al., 2020; Hung et al., 2020) (Li et al., 2022). The availability of high-quality chromosome-level reference genomes can greatly advance genetic research in economically valuable plant species (Sahu and Liu, 2023). In this study, the successful assembly of a chromosome-scale *D. sissoo* genome represents a major advance for investigating heartwood formation and highlights the potential impact on conserving and sustainably utilizing this important tree. The *D. sissoo* genome revealed typical features of woody plants, with relatively few genes (29,737) compared to herbaceous species but abundant repeats (52%) and secondary metabolites (Hong et al., 2020) (Neale and Kremer, 2011; Yasodha et al., 2018). Comparative genomics showed expansions in several gene families related to heartwood synthesis and defense compared to other trees (Sahu et al., 2019; Fan et al., 2020). These expanded genes likely contribute to *D. sissoo*'s production of medicinal chemicals and durable, insect-resistant timber. These findings mirror observations in other tropical genera like Eucalyptus and Mango (Myburg et al., 2014). The high-quality reference genome will facilitate identifying genes and networks involved in heartwood formation and valuable wood traits, providing useful markers for breeding and conservation

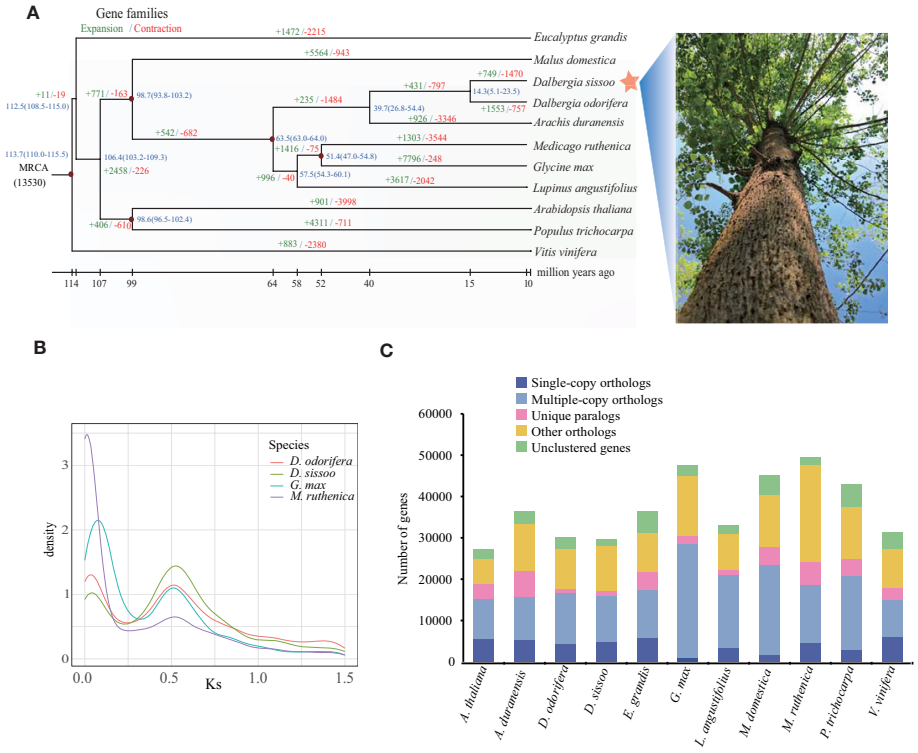


FIGURE 2
Evolution of the *D sissoo* genome. **(A)** Phylogenetic tree showing the taxonomic position of *D. sissoo*. The blue numbers denote divergence time of each node (MYA: million years ago). The numbers in green indicate the number of gene families that expanded in the species during evolution, and the numbers in red indicate the number of gene families that contracted. **(B)** Ks distribution plot. **(C)** The distribution of single-copy, multiple-copy, unique, and other genes in the 11 plant species.

(Yasodha et al., 2018; Hung et al., 2020; Guo et al., 2021). It also enables population genomics and association studies to elucidate adaptations in these ecologically and economically vital Neotropical trees (Naidoo et al., 2019; Hong et al., 2020). Furthermore, the genome can serve as a reference for resequencing other threatened *Dalbergia* species exploited for timber. The *D. sissoo* draft genome thus represents a vital contribution for genetics-based improvement and sustainable forestry of rosewoods and tropical trees.

Materials and methods

Plant materials, library construction, and sequencing

The fresh material of *D. sissoo* was obtained from Ruili Botanical Garden in Yunnan Province, China. The total DNA of young leaves was extracted by the standard (cetyltrimethylammonium bromide) CTAB method (Sahu et al., 2012), and the library was constructed according to the standard procedure of 10X genomics library kit (Chromium Genome Chip Kit v1, 10X Genomics, Pleasanton, USA), and the obtained target DNA library was sequenced using BGISEQ-500 sequencer to obtain read length data. The RNA was extracted from young leaves, xylem and phloem tissues using the PureLink RNA Mini Kit (Thermo Fisher Scientific, Carlsbad, CA, USA), and the library was built according to the standard procedure of the TruSeq RNA Sample Preparation Kit manual (Illumina, San Diego, CA, USA), and the obtained library was sequenced using the BGISEQ-500 platform to obtain transcriptome data at 100bp pair-end. Furthermore, according to the standard procedure of the Hi-C experiment, young leaves of *D. sissoo* were selected for cross-linked DNA extraction and library construction, and the constructed Hi-C library was sequenced on the BGISEQ-500 platform for 100bp pair-end sequencing to obtain Hi-C data. DNA and RNA data were subjected to FastQC (v 0.11.9) analysis to assess the quality of the raw downstream data, followed by filtering using Trimmomatic (v 0.39) (Bolger et al., 2014) based on the QC results, and the statistical information of each filtered data is shown in Supplementary Table 1.

Genome assembly and assessment

We used DNA data from 10X genomics an input for the GCE (v 1.0.0) (Liu et al., 2013) and Kmerfreq_16bit (Wang et al., 2020) for genome size prediction. The genome was then preliminarily assembled using the supernova software for the 10X genomics data. To further improve the assembly quality and construct the genome at the chromosome level, the anchoring of the chromosome was performed using Juicer (v 1.5) (Durand et al., 2016) and 3D-DNA (v180419) (Dudchenko et al., 2017).

After obtaining the results of the genome assembly, the quality of the genome assembly was assessed in five main ways: (1)

Counting Scaffold N50 and N90 length. (2) Using BUSCO (v3.0.1) (Simão et al., 2015) to assess the number and proportion of results of chromosome assembly occupying the angiosperm core gene set database (embryophyta_odb10). (3) All filtered DNA data were compared back to the genome using Bwa-mem (v0.7.12)(Li, 2013), and the read comparison rate was counted. (4) All filtered RNA data were compared back to the genome using Tophat2 (v2.1.1) (Kim et al., 2013), and the read ratio was counted. (5) In order to verify the gene contamination of the assembly, the above comparison results were used as input in soap.coverage 2.27 (http://soap.genomics.org.cn/) to calculate the relationship between GC content and coverage, and to display the distribution between the coverage of reads and GC content.

Genome annotation

Genome annotation mainly includes repetitive sequence annotation, gene annotation and non-coding RNA annotation. Firstly, repeat annotation was performed using a method based on *ab initio* prediction and comparison with a database of known repeats. For tandem repeat sequences, tandem structure identification was performed using TRF (v 4.10.0) (Benson, 1999). For dispersed repeat sequences, LTR_Finder (v 1.0.7) (Xu and Wang, 2007) was first used for *de novo* prediction of LTR sequences based on repetitive features, and RepeatModeler (v 2.0.1) (Flynn et al., 2020) was used for *ab initio* prediction of other repetitive elements based on the comparison of the sequences themselves, and a local repeat database was constructed. Finally, by combining the local repeat database and the downloaded repeat sequence library, Repbase (v 21.12) (Jurka, 2000), using RepeatMasker (v 4.0.7) (Chen, 2004) and RepeatProteinMask (v 4.0.7) (Chen, 2004) to perform the prediction at the nucleic acid and protein sequencing levels, respectively.

For gene structure annotation, we adopted a combination of *de novo*, homology, and transcriptome-based annotations. For *de novo* annotation, the transcriptome reads were first compared back to the reference genome using Hisat2 (v 2.2.1)(Kim et al., 2015), and the correctly compared reads were assembled using Stringtie (v 1.3.3b) (Pertea et al., 2015) and then the obtained transcripts were compared back to the reference genome by PASA (v 2.1.0) (Haas et al., 2008) for gene model prediction. For homology annotations, gene sets of the leguminous relatives of *Arachis duranensis*, *G. max*, and *M. truncatula*), which are also in the same legume family as *D. sissoo*, were downloaded from the NCBI or Phytozome websites as evidence of homology annotation. For transcriptome annotation, transcriptome data were assembled using Trinity (v 2.0.6) (Haas et al., 2013), and the transcripts obtained were used as transcriptome annotation evidence. Finally, the above evidence obtained after *de novo* annotation, homology annotation, and transcriptome annotation were jointly integrated as input files for MAKER (v 2.31.11) (Cantarel et al., 2008), which was run to output the final gene structure annotation results. The protein sequences of 29737 genes obtained from structural annotation were compared with Interpro, Swissprot, KEGG, NR, Trembl, and COG functional databases using BLASTP (v2.2.31) (Camacho et al., 2009) with the threshold set to E-value <le-05. For ncRNA annotation, BLASTN (v 2.2.26)(Camacho et al., 2009) was used in combination with the highly conserved rRNA database of closely related species to search for rRNAs; tRNAscan-SE (v 1.3.1) (Lowe and Eddy, 1997) was used to search for genomic tRNAs based on secondary structure features; infernal (v 1.1.2) (Nawrocki and Eddy, 2013) was used in combination with the Rfam database (v 12.0) (Nawrocki et al., 2015) for secondary structure detection of snRNAs and microRNAs.

Phylogenomic analysis

Gene family clustering analysis was performed using OrthoFinder (v2.3.1) (Emms and Kelly, 2015) combining protein sequences of genes encoding a total of 10 plants: *A. thaliana*, *A. duranensis*, *P. trichocarpa*, *D. odorifera*, *Lupinus angustifolius*, *G. max*, *E. grandis*, *M. ruthenica*, *Malus domestica*, and *V. vinifera*. Based on OrthoFinder clustering results, single-copy orthologous gene families were extracted using Linux commands, and then all protein sequences of single-copy gene families were compared in multiple sequences using MAFFT (v7.310) (Katoh and Standley, 2013). After filtering, a sequence supermatrix was generated by concatenating all genes in the single-copy gene family in order of species origin. Finally, this matrix file was used as the input of IQTREE (v1.6.1) (Nguyen et al., 2015) for phylogenetic tree construction. The obtained tree construction results were used to root and visualize the phylogenetic tree using FigTree (v1.4) (http://tree.bio.ed.ac.uk/software/figtree/). Subsequently, the phylogenetic tree was further combined with the MCMCTREE module in the PAML (v4.5) (Yang, 2007) and the TimeTree divergence time database (http://www.timetree.org) information to perform divergence time estimation of target species in the tree.

Gene family analysis

The expansion and contraction analysis of the gene family was performed using CAFÉ (v4.2.1) (De Bie et al., 2006). Firstly, the species gene family number statistics file obtained from OrthoFinder clustering and the MCMCTREE predicted species divergence time tree file were used as input to CAFÉ. CAFÉ employs the self-contained birth-and-death evolution model to compare the number of genes of different species in each gene family with each other to derive the expansion and contraction of gene families, followed by enrichment analysis based on GO and KEGG annotations to determine the functional meaning of the expanded and contracted genes.

Whole-genome duplication analysis

We used the wgd package's BLASTP module (Zwaenepoel and Van de Peer, 2019) to conduct self-comparison and inter-comparison of CDS sequences of *D. sissoo* and its close relatives, and MCL clustering was used to find direct homologous and paralogous gene families based on BLASTP results; then MAFFT module was used to perform multiple sequence concatenation for

each gene family, the Codeml module was used to calculate Ks, and finally the R scripts were used to derive Ks distribution maps of Ks of different species.

Identification of gene duplication events

DupGen_finder (Qiao et al., 2019) was used to identify the gene duplication type of *D. sissoo* genome. First, BLASTP (v2.2.26) was used to compare the gene and protein sequences of *D. sissoo* (e<10-5). Then gene position information was extracted from the gene structure annotation GFF file, including the sequence ID where the gene is located, the gene ID, and the start and end position information of the gene on the sequence. Finally, the alignment result file and gene location information were used as an input in DupGen_finder.unique to predict gene duplication types.

Data availability statement

The raw sequencing data and genome assembly data reported in this article have been deposited into CNGB Sequence Archive (CNSA) of the China National GeneBank DataBase (CNGBdb) with accession number CNP0004041 (https://db.cngb.org/search/project/CNP0004041/). Annotation files are available via figshare (https://figshare.com/articles/dataset/22662127).

Author contributions

HL, SS, and CH led and designed this project. HL and SS conceived the study. SS, and JW collected the leaf and tissue samples. SS, ML, RL, GW, and YC contributed to the sample preparation and performed the genome and chromosome-scale assembly. SS, ML, SW, YC, DF, GW, DS, RL, and SW performed annotation and comparative genomic analyses. SS, RL, and ML wrote the original draft manuscript. SW, ML, CH, DS, DF, and HL, revised and edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Major Science and Technology Projects of Yunnan Province (Digitalization, Development and Application of Biotic Resource, 202002AA100007), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27020104).

Acknowledgments

This work is part of the 10KP project (https://db.cngb.org/10kp/) (Cheng et al., 2018) and is also supported by China National GeneBank (CNGB; https://www.cngb.org/).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1218515/full#supplementary-material

References

Al-Snafi, A. E. (2017). Chemical constituents and pharmacological effects of Dalbergia sissoo-A review. *IOSR J. Pharm.* 7 (2), 59–71. doi: 10.9790/3013-0702015971

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573

Bhagwat, R. M., Dholakia, B. B., Kadoo, N. Y., Balasundaran, M., and Gupta, V. S. (2015). Two new potential barcodes to discriminate Dalbergia species. *PloS One* 10 (11), e0142965. doi: 10.1371/journal.pone.0142965

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 1–9. doi: 10.1186/1471-2105-10-421

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18 (1), 188–196. doi: 10.1101/gr.6743907

Chen, N. (2004). Using Repeat Masker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* 5 (1), 4.10. doi: 10.1002/0471250953.bi0410s05

Chen, J., Hao, Z., Guang, X., Zhao, C., Wang, P., Xue, L., et al. (2019). Liriodendron genome sheds light on angiosperm phylogeny and species–pair differentiation. *Nat. Plants* 5 (1), 18–25. doi: 10.1038/s41477-018-0323-6

Cheng, S., Melkonian, M., Smith, S. A., Brockington, S., Archibald, J. M., Delaux, P.-M., et al. (2018). 10KP: A phylodiverse genome sequencing plan. *GigaScience* 7 (3), 1–9. doi: 10.1093/gigascience/giy013

De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22 (10), 1269–1271. doi: 10.1093/bioinformatics/btl097

Djerbi, S., Lindskog, M., Arvestad, L., Sterky, F., and Teeri, T. T. (2005). The genome sequence of black cottonwood (Populus trichocarpa) reveals 18 conserved cellulose synthase (CesA) genes. *Planta* 221, 739–746. doi: 10.1007/s00425-005-1498-4

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). *De novo* assembly of the Aedes aEgypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356 (6333), 92–95. doi: 10.1126/science.aal3327

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3 (1), 95–98. doi: 10.1016/j.cels.2016.07.002

Emms, D. M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16 (1), 1–14. doi: 10.1186/s13059-015-0721-2

Fan, Y., Sahu, S. K., Yang, T., Mu, W., Wei, J., Cheng, L., et al. (2020). Dissecting the genome of star fruit (Averrhoa carambola L.). *Hortic. Res.* 7 (1), 94. doi: 10.1038/s41438-020-0306-4

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* 117 (17), 9451–9457. doi: 10.1073/pnas.1921046117

Guo, X., Fang, D., Sahu, S. K., Yang, S., Guang, X., Folk, R., et al. (2021). Chloranthus genome provides insights into the early diversification of angiosperms. *Nat. Commun.* 12 (1), 6930. doi: 10.1038/s41467-021-26922-4

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8 (8), 1494–1512. doi: 10.1038/nprot.2013.084

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, 1–22. doi: 10.1186/gb-2008-9-1-r7

Hong, Z., Li, J., Liu, X., Lian, J., Zhang, N., Yang, Z., et al. (2020). The chromosome-level draft genome of Dalbergia odorifera. *Gigascience* 9 (8), giaa084. doi: 10.1093/gigascience/giaa084

Hung, T. H., So, T., Sreng, S., Thammavong, B., Boounithiphonh, C., Boshier, D. H., et al. (2020). Reference transcriptomes and comparative analyses of six species in the threatened rosewood genus Dalbergia. *Sci. Rep.* 10 (1), 1–14. doi: 10.1038/s41598-020-74814-2

Isabel, N., Holliday, J. A., and Aitken, S. N. (2020). Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation. *Evolutionary Appl.* 13 (1), 3–10. doi: 10.1111/eva.12902

Jurka, J. (2000). Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.* 16 (9), 418–420. doi: 10.1016/s0168-9525(00)02093-x

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30 (4), 772–780. doi: 10.1093/molbev/mst010

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14 (4), 1–13. doi: 10.1186/gb-2013-14-4-r36

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint*. doi: 10.48550/arXiv.1303.3997

Li, M., Liu, M., Wang, B., and Shi, L. (2022). Metabonomics analysis of stem extracts from dalbergia sissoo. *Molecules* 27 (6), 1982. doi: 10.3390/molecules27061982

Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., et al. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *arXiv preprint*. doi: 10.48550/arXiv.1308.2012

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25 (5), 955–964. doi: 10.1093/nar/25.5.955

Marrano, A., Britton, M., Zaini, P. A., Zimin, A. V., Workman, R. E., Puiu, D., et al. (2020). High-quality chromosome-scale assembly of the walnut (Juglans regia L.) reference genome. *Gigascience* 9 (5), giaa050. doi: 10.1093/gigascience/giaa050

Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., et al. (2014). The genome of Eucalyptus grandis. *Nature* 510 (7505), 356–362. doi: 10.1038/nature13308

Naidoo, S., Slippers, B., Plett, J. M., Coles, D., and Oates, C. N. (2019). The road to resistance in forest trees. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00273

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., et al. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 43 (D1), D130–D137. doi: 10.1093/nar/gku1063

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29 (22), 2933–2935. doi: 10.1093/bioinformatics/btt509

Neale, D. B., and Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nat. Rev. Genet.* 12 (2), 111–122. doi: 10.1038/nrg2931

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32 (1), 268–274. doi: 10.1093/molbev/msu300

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497 (7451), 579–584. doi: 10.1038/nature12211

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20 (1), 1–23. doi: 10.1186/s13059-019-1650-2

Sahu, S. K., and Liu, H. (2023). Long-read sequencing (method of the year 2022): the way forward for plant omics research. *Mol. Plant* 16 (5), 791–793. doi: 10.1016/j.molp.2023.04.007

Sahu, S. K., Liu, M., Chen, Y., Gui, J., Fang, D., Chen, X., et al. (2023). Chromosome-scale genomes of commercial timber trees (Ochroma pyramidale, Mesua ferrea, and Tectona grandis). *Scientific Data* 10 (1), 512. doi: 10.1038/s41597-023-02420-8

Sahu, S. K., Liu, M., Yssel, A., Kariba, R., Muthemba, S., Jiang, S., et al. (2019). Draft Genomes of Two Artocarpus Plants, Jackfruit (A. heterophyllus) and Breadfruit (A. altilis). *Genes (Basel)* 11 (1), 1–17. doi: 10.3390/genes11010027

Sahu, S. K., Thangaraj, M., and Kathiresan, K. (2012). DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. *Int. Scholarly Res. Notices* 2012, 1–16. doi: 10.5402/2012/205049

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *nature* 463 (7278), 178–183. doi: 10.1038/nature08670

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19), 3210–3212. doi: 10.1093/bioinformatics/btv351

Son, N. T., and Manh Ha, N. (2022). Siamese, Indian, and Brazilian rosewoods: A review on phytochemistry, applications, and pharmacology. *Natural Product Commun.* 17 (4), 1934578X221096962. doi: 10.1177/1934578X221096962

Wang, S., Liang, H., Wang, H., Li, L., Xu, Y., Liu, Y., et al. (2022). The chromosome-scale genomes of Dipterocarpus turbinatus and Hopea hainanensis (Dipterocarpaceae) provide insights into fragrant oleoresin biosynthesis and hardwood formation. *Plant Biotechnol. J.* 20 (3), 538–553. doi: 10.1111/pbi.13735

Wang, H., Liu, B., Zhang, Y., Jiang, F., Ren, Y., Yin, L., et al. (2020). Estimation of genome size using k-mer frequencies from corrected long reads. *arXiv preprint*. doi: 10.48550/arXiv.2003.11817

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35 (suppl_2), W265–W268. doi: 10.1093/nar/gkm286

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24 (8), 1586–1591. doi: 10.1093/molbev/msm088

Yasodha, R., Vasudeva, R., Balakrishnan, S., Sakthi, A. R., Abel, N., Binai, N., et al. (2018). Draft genome of a high value tropical timber tree, Teak (Tectona grandis L. f): insights into SSR diversity, phylogeny and conservation. *DNA Res.* 25 (4), 409–419. doi: 10.1093/dnares/dsy013

Yin, M., Zhang, S., Du, X., Mateo, R. G., Guo, W., Li, A., et al. (2021). Genomic analysis of Medicago ruthenica provides insights into its tolerance to abiotic stress and demographic history. *Mol. Ecol. Resour.* 21 (5), 1641–1657. doi: 10.1111/1755-0998.13363

Zwaenepoel, A., and Van de Peer, Y. (2019). wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35 (12), 2153–2155. doi: 10.1093/bioinformatics/bty915

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 October 2023
DOI 10.3389/fpls.2023.1241908

OPEN ACCESS

EDITED BY
Jiehua Wang,
Tianjin University, China

REVIEWED BY
Jaroslav Klápště,
Tree Breeding Australia, Australia
Kyu-Suk Kang,
Seoul National University,
Republic of Korea

*CORRESPONDENCE
Naoki Tani
ntani@affrc.go.jp
Mohammad Na'iem
mnaiem@ugm.ac.id

RECEIVED 17 June 2023
ACCEPTED 13 September 2023
PUBLISHED 31 October 2023

CITATION
Akutsu H, Na'iem M, Widiyatno, Indrioko S, Sawitri, Purnomo S, Uchiyama K, Tsumura Y and Tani N (2023) Comparing modeling methods of genomic prediction for growth traits of a tropical timber species, *Shorea macrophylla*. *Front. Plant Sci.* 14:1241908. doi: 10.3389/fpls.2023.1241908

Comparing modeling methods of genomic prediction for growth traits of a tropical timber species, *Shorea macrophylla*

Haruto Akutsu[1], Mohammad Na'iem[2]*, Widiyatno[2], Sapto Indrioko[2], Sawitri[2], Susilo Purnomo[3], Kentaro Uchiyama[4], Yoshihiko Tsumura[5] and Naoki Tani[5,6]*

[1]Graduate School of Science and Technology, University of Tsukuba, Tsukuba, Ibaraki, Japan, [2]Faculty of Forestry, Gadjah Mada University, Yogyakarta, Indonesia, [3]PT. Sari Bumi Kusuma, Pontianak, West Kalimantan, Indonesia, [4]Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Tsukuba, Ibaraki, Japan, [5]Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan, [6]Forestry Division, Japan International Research Center for Agricultural Sciences, Tsukuba, Ibaraki, Japan

Introduction: *Shorea macrophylla* is a commercially important tropical tree species grown for timber and oil. It is amenable to plantation forestry due to its fast initial growth. Genomic selection (GS) has been used in tree breeding studies to shorten long breeding cycles but has not previously been applied to *S. macrophylla*.

Methods: To build genomic prediction models for GS, leaves and growth trait data were collected from a half-sib progeny population of *S. macrophylla* in Sari Bumi Kusuma forest concession, central Kalimantan, Indonesia. 18037 SNP markers were identified in two ddRAD-seq libraries. Genomic prediction models based on these SNPs were then generated for diameter at breast height and total height in the 7th year from planting (D7 and H7).

Results and discussion: These traits were chosen because of their relatively high narrow-sense genomic heritability and because seven years was considered long enough to assess initial growth. Genomic prediction models were built using 6 methods and their derivatives with the full set of identified SNPs and subsets of 48, 96, and 192 SNPs selected based on the results of a genome-wide association study (GWAS). The GBLUP and RKHS methods gave the highest predictive ability for D7 and H7 with the sets of selected SNPs and showed that D7 has an additive genetic architecture while H7 has an epistatic genetic architecture. LightGBM and CNN1D also achieved high predictive abilities for D7 with 48 and 96 selected SNPs, and for H7 with 96 and 192 selected SNPs, showing that gradient boosting decision trees and deep learning can be useful in genomic prediction. Predictive abilities were higher in H7 when smaller number of SNP subsets selected by GWAS *p*-value was used, However, D7 showed the contrary tendency, which might have originated from the difference in genetic architecture between primary and secondary growth of the species. This study suggests that GS with GWAS-based SNP selection can be used in breeding for non-cultivated tree species to improve initial growth and reduce genotyping costs for next-generation seedlings.

KEYWORDS

dipterocarpaceae, timber species, tree breeding, GWAS, genomic prediction, machine learning

1 Introduction

Shorea macrophylla belongs to the Dipterocarpaceae, the dominant tree family in the tropical rainforests in Southeast Asia (Ghazoul, 2016). It is commercially important in Borneo as a source of timber and fruit that is harvested to produce oil (Randi et al., 2019). It can also be used in plantation forestry because of its rapid initial growth (Lee et al., 1997). Breeding resources such as progeny and provenance trials are rarely available for tropical tree species with the exception of species such as eucalyptus and teak (Grattapaglia & Kirst, 2008; Corriyanti & Muharyani, 2020). Even when breeding resources exist, their scales and the associated facilities are often limited due to a lack of breeding history. Genomic selection (GS; Meuwissen et al., 2001), which has been put forward as a way to shorten tree breeding cycles (Grattapaglia et al., 2018; Lebedev et al., 2020), is therefore an attractive option for expediting the breeding of tropical species. Another advantage of GS in tree breeding is that a training population can be established without mating families or *a priori* information about the family. GS could therefore greatly accelerate improvement in the focal traits of tropical tree species. Further, the genetic resources of tree species with limited breeding history are those of an almost wild population, so their genetic diversity typically exceeds that of a breeding population (Jones et al., 2006). This study investigates the potential for using GS to accelerate tree breeding in species with limited breeding history and little established collections of breeding resources.

GS has a faster cycle time than pedigree-based breeding because it allows individuals to be evaluated early in their lives by calculating their genomic estimated breeding values (GEBVs) using their genomic information. Many studies have compared the performance (i.e., accuracy) of different methods for predicting GEBVs, which is important because the genetic gains from GS depend on the accuracy of the GEBVs (Lebedev et al., 2020). However, these studies have mainly used parametric methods, which can only account for additive genetic effects, in order to capture inheritable genetic effects for parent selection. However, non-additive effects (i.e., dominant and epistatic effects) are also of interest because their inclusion can increase the accuracy of GEBVs and the prediction of genetic responses (Varona et al., 2018). Accordingly, when using simulated data, the predictive ability achieved with non-parametric methods exceeds that for parametric methods when epistatic genetic effects are present (Ober et al., 2011; González-Camacho et al., 2012; Howard et al., 2014). Some GS studies have used non-parametric methods to account for both additive and non-additive genetic effects (Chen et al., 2019; de Almeida Filho et al., 2019; Abdollahi-Arpanahi et al., 2020; Yan et al., 2021). However, there has been no detailed comparison of parametric and non-parametric methods to identify those providing the most accurate GEBVs.

Machine learning using deep learning (DL) methods has recently been remarkably successful in diverse applications involving the processing and analysis of complex datasets including image recognition and natural language processing. Many traits important in forestry, such as growth, are considered to be complex traits because they are regulated by groups of many polygenes. DL appears to be well-suited to the analysis of such complex traits and is therefore increasingly being used for this purpose (Bellot et al., 2018; Ma et al., 2018; Abdollahi-Arpanahi et al., 2020). Among DL models, convolutional neural networks (CNNs) seem to provide more accurate genomic predictions than multilayer perceptrons (MLPs) (Pérez-Enciso & Zingaretti, 2019), possibly because CNNs apply filters to small parts of the input data and can thus model patterns within each part of the data. This may be advantageous when analyzing genotype data covering most of an entire genome because of the genome's continuous nature. Additionally, CNNs can accept two-dimensional input data such as images, while other methods can accept only one-dimensional data. This means that information on the nuclear phase can be incorporated into the model. Gradient boosting decision trees (GBDTs) are another class of machine learning methods that perform well in classification, regression, and ranking tasks and are starting to be used in GS studies (Li et al., 2018; Azodi et al., 2019; Abdollahi-Arpanahi et al., 2020). GBDTs rely on gradient boosting, in which learning is achieved through iterative minimization of a loss function using base-learners. GBDTs generate decision trees for the base-learners to make predictions, which construct tree-structed predictive models that split samples using their feature values as criteria. Some GBDT implementations have high computational efficiency and achieve accurate predictions in diverse practical applications.

Genomic prediction models are typically built by genotyping an enormous number of genetic markers covering the whole genome of individuals from a training population, which are then regressed against phenotypic traits. However, to apply such models, it then becomes necessary to genotype thousands of progeny individuals derived from the training population. Because progeny populations are generally much larger than training populations, the cost of such genotyping can be very high. Additionally, this approach suffers from the "large *P*, small *N* problem" because the number of explanatory variables (genetic markers in this case) greatly exceeds the number of observations (individuals in the training population). To overcome these problems, some studies have focused on subsets of genetic markers selected using ranking values. The predictive accuracy of such models can rival or exceed that of conventional models using every available marker (Hiraoka et al., 2018; Li et al., 2018; e Sousa et al., 2019).

This is the second empirical study on GS for Dipterocarpaceae and builds on an earlier work (Sawitri et al., 2020) showing that GS could improve the productivity and quality of *S. platyclados*. To investigate the potential of genomic breeding for *S. macrophylla* and other tropical forestry species with limited breeding resources, we aimed to: (1) identify the methods that yield the highest GEBVs predictability, (2) evaluate the effectiveness of DLs and GBDTs in genomic prediction, and (3) determine the impact of genetic marker selection on predictive accuracy.

2 Materials and methods

2.1 Collecting plant material and phenotyping

A progeny trial of *S. macrophylla* at the PT. Sari Bumi Kusuma forest concession, Central Kalimantan, Indonesia was targeted for genomic prediction modeling. The progeny trial was established in 2006 using open-pollinated seeds from 94 mother trees in natural populations in the forest concession. The progeny was cultivated in a nursery for ten months and then planted with a 6 x 3 meter spacing. Five trees derived from each mother tree were planted using a Randomized Complete Block Design (RCBD) in a single block and eight block replicates were prepared, giving 3760 progeny trees in total. The phenotypic variance of the progeny population was deemed adequate because significant between-family differences in diameter at breast height (DBH) and total height (HT) were observed 4 years after planting (Widiyatno et al., 2014). Two blocks on relatively gentle terrain containing 940 trees were selected to build genomic prediction models. Three trees from each mother tree were removed for thinning in 2015, 9 years after planting. Leaf samples were collected in 2018 from 361 surviving trees, packed in plastic bags with silica gel, and then transferred to a laboratory at JIRCAS and stored in -30°C freezer until DNA extraction. The HT and DBH of these trees were measured at 1, 2, 3, 4, 7, 9, 11.5, and 12.5 years after planting. HT was measured using a measuring rod up to 11.5 years after planting, then using a Vertex instrument at 12.5 years old. DBH was measured by measurement tape.

2.2 DNA sequencing and genotyping

Leaf samples (n=361) were individually pulverized in liquid nitrogen, then total genomic DNA was extracted from 60 mg of the resulting powder using a modified Cetyltrimethylammonium bromide (CTAB) method (Murray & Thompson, 1980). The crude DNA was purified using the NucleoSpin® gDNA Clean-up kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany) following the manufacturer's instructions. The DNA content of each cleaned-up sample was quantified using a Qubit® dsDNA BR Assay Kit and Qubit® 3.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, U.S.A.), and the samples were adjusted to a final DNA concentration of 25 ng/mL. Two double digest Restriction-site Associated DNA sequencing (ddRAD-seq) libraries were then generated by digesting 500 ng of the DNA using *Mse*I with *Pst*I or *Mlu*CI with *Bgl*II (New England Biolabs Inc., Ipswich, MA, U.S.A.), as described by Peterson et al. (2012). *Mse*I and *Mlu*CI are 4-base cutters, while *Pst*I and *Bgl*II are 6-base cutters. The resulting DNA fragments were then ligated with adapter cassettes at restriction enzyme cleavage sites. Since the 4-base cutters had greater numbers of cleavage sites, the corresponding adapters were designed to form Y-shaped structures that suppressed fragment amplification in cases where *Mse*I or *Mlu*CI sites were on both sides. Indexed PCR fragments were amplified using KAPA HiFi DNA Polymerase (Kapa Biosystems Inc., Wilmington, MA, U.S.A.) and the Access Array Barcode Library for Illumina® Seqencers——384, Single Direction (Standard BioTools Inc., South San Francisco, CA, U.S.A.) following the manufacturer's instructions. We then selected 430-470 bp fragments using a 2% agarose gel with the BluePippin system (Sage Science Inc., Beverly, MA, U.S.A.). These size-selected PCR fragments were adjusted to a concentration of 10 nM by assuming an average fragment size of 450 bp. Finally, each library solution was sequenced using 2 lanes on the Illumina Hiseq X platform (illumine, San Diego, CA, USA) to obtain 151 bp paired-end sequences.

dDocent version 2.8.13 (Puritz et al., 2014) was used with default settings for quality filtering, read mapping, and SNP calling on raw reads from Hiseq X sequencing of the *Mse*I/*Pst*I library. Version 2.9.4 of the same software was used to perform the same tasks for the *Mlu*CI/*Bgl*II library. During the read mapping step, all reads were mapped to the draft genome of *S. leprosula* (Ng et al., 2021) which belongs to the same genus as *S. macrophylla*. The output file in variant call format (VCF), which was named "TotalRawSNPs.vcf", was then filtered using VCFtools version 0.1.16 (Danecek et al., 2011). SNP filtering was performed using a 10-step procedure. In the first step, samples were filtered by applying the following conditions: missing data ≤ 50%, quality value ≥ 30, minor allele ≥ 3 and depth ≥ 3 (–max-missing 0.5, –minQ 30, –mac 3 and –minDP 3). Second, samples were filtered by the proportion of missing data ≤ 0.65 (F_MISS on out.imiss > 0.65). Third, SNPs were filtered by applying the criteria proportion of non-missing data ≥ 0.95, minor allele frequency ≥ 0.05 and mean depth ≥ 10 (–max-missing 0.95, –maf 0.05 and –min-meanDP 10). Fourth, the vcf file was filtered using dDcent_filters v2.9.4 (Puritz et al., 2018) by applying filters "based on allele balance at heterozygous loci, locus quality, and mapping quality/Depth"; filters "based on overlapping forward and reverse reads"; filters "based on properly paired status" assuming paired-end library; filters "based on high depth and lower than 2*DEPTH quality score"; and filters "based on maximum mean depth" with maximum mean depth cutoff 134. Fifth, we removed SNPs with depths below 60% of the average of groups comprising SNPs within distances of 151 base pairs by applying the filter "based on within locus depth mismatch" from dDocent_filters. This was implemented in a modified part of dDocent_filters because it supports the use of vcf files based on a reference genome. Sixth, we applied a filter based on the significance of deviation from Hardy-Weinberg Equilibrium being ≥ 0.001 (–hwe 0.001). Seventh, the two vcf files for the two libraries were combined using "vcfcombine" in vcflib version 1.0.3 (Garrison et al., 2022) with priority given to the library using *Pst*I and *Bgl*II. Eighth, SNPs were filtered by applying the "Filter Genotype Table Taxa" and "Filter Genotype Table Sites" in TASSEL 5.2 (Bradbury et al., 2007) with the following criteria "Min Proportion of Sites Present" 0.9, "Site Min Count" 60% of number of taxa and "Site Min Allele Freq" 0.05. Ninth, solitary SNPs within each scaffold were removed using R version 4.2.1 (R Core Team, 2020) and the vcfR package version 1.13.0 (Knaus & Grünwald, 2017) for beagle imputation. Then, we used beagle 5.4 (Browning et al., 2018; Browning et al., 2021) with

the default effective population size: ne value (100,000) to estimate an appropriate N_e value. The log2-transformed values of the estimated ne obtained from the log file were divided into classes of 0.1 width, and the most frequent class unless the default value class was considered the appropriate class. Finally, missing data were imputed by beagle 5.4 using a N_e value of 38,968, which is between the values for the two most frequent classes.

2.3 Genetic structure and linkage disequilibrium

We analyzed genetic structure using principal component analysis (PCA). First, SNPs within 1000 bp of one-another were thinned using VCFtools (–thin 1000) to prevent detection of biased structure due to neighboring SNPs. The vcf data were converted to matrices with the following coding: -1 for homozygous reference alleles, 0 for heterozygous alleles, and 1 for homozygous alternative alleles: 1. PCA was performed in R using the "prcomp" function. We also analyzed the linkage disequilibrium of each SNP using TASSEL 5.2 with a window size of 870 to complete all combinations within each scaffold. To obtain a trendline for linkage disequilibrium decay within the inter-SNP distance, the r^2 values were regressed against the inter-SNP distance using the "loess" function with the following parameter settings: degree = 2, control = loess. control(surface = "interpolate"). The "span" parameter of the function was optimized using four-fold cross-validation, minimizing the root mean squared error of the regression line and the observed data.

2.4 Spatial structure analysis for micro-environmental effects

The observed values of the phenotypic data in the progeny trial were corrected by performing a spatial structure analysis to reduce the impact of micro-environment effects, which could otherwise introduce noise that would reduce the accuracy of the genomic prediction models (Chen et al., 2018). This analysis was implemented in R using the "remlf90" function of the breedR package version 0.12.5 (Munoz & Sanchez, 2020) with a splines model to fit spatial structure along the coordinates of the trees using 12 and 13 knots (default setting). We then subtracted the estimated values due to spatial effects from the observed phenotypic values to obtain the corrected phenotypic values. Outliers among the adjusted phenotypic values were removed using the 5% two-tailed Smirnov-Grubbs test, implemented in R.

2.5 Genomic heritability

We analyzed genotype and phenotype data to check heritability and SNP effects on each phenotype. Samples were removed if at least one of them was NA within the same year in these analyses. To identify well-predicted phenotypes, narrow-sense genomic heritability was calculated using the equation: $h^2 = \sigma_a^2/\sigma_y^2$ where σ_a^2 is additive genetic variance and σ_y^2 is total phenotypic variance, which were averaged over ten iterations for each trait. Variances were obtained from variance components estimated using the Bayesian Ridge Regression (BRR) model, which was implemented in R using "BGLR" function of the BGLR package version 1.1.0 (Pérez & de los Campos, 2014).

2.6 Genome-wide association study using all individuals

A genome-wide association study (GWAS) was conducted using DBH and HT in the 7th year (D7 and H7) as the traits of interest. These traits were selected (1) because they have relatively high heritability, (2) to avoid maternal effects expressed strongly after germination and during initial growth, (3) because the observation values of HT in the 11th year were discrete, and (4) since measurements using measurement rods are more accurate than those obtained using the Vertex system.

The aim of the GWAS was to identify SNPs significantly associated with the traits of interest and to verify that both traits are complex, therefore, all individuals were used in the first GWAS. This was done in R using the "mrMLM" function of the mrMLM package version 5.0.1 (Wen et al., 2018) with FASTmrEMMA (Tamba & Zhang, 2018). In this method, a single-locus method is first used to scan the whole genome and calculate *P*-values for each SNP. Second, LOD scores are calculated for strongly trait-associated SNPs; these scores indicate the probability of association between the SNP and quantitative trait loci (QTL) of the trait. Population and kinship structures were flattened using principal components 1 to 4 and a covariance matrix (K) inferred by the genetic markers as default setting of the "mrMLM" function (Wen et al., 2018). *P*-values were corrected by the false discovery rate (FDR) to reduce the incidence of false positive SNPs.

2.7 Building genomic prediction models and SNP marker selection

Genomic prediction models were generated using each of 12 methods described in the section 2.8. All models were built using 10 training/validation sets and 10 model construct replicates were generated to assess split effects in a half-sib population. After the split of training/validation sets, we confirmed that all SNP markers maintain their polymorphism in each of training populations (Table S4). No model construct replicates were generated for GBLUP because the nature of this method means that all replicates yield identical results. Model building was done by first splitting the full dataset into one training and one validation set with a size ratio of 3:1. Genotype data was coded as specified previously (homozygous reference allele: -1, heterozygous allele: 0, homozygous alternative allele: 1). For CNN2D, nuclear phase data were coded as reference allele: 0 and alternative allele: 1. The phenotype and genotype values in the training data were then normalized using "StandardScaler" in the scikit-learn library. Validation data were also normalized using parameters from

training data normalization. These scaling steps were omitted for decision tree methods (RF, LGB, XGB), and for genotype values in GBLUP and CNN2D. Next, each model was built using the training data and validated using the validation data. For hyper-parameter optimization, 3-fold inner cross-validation was performed using "TPESampler" to evaluate sampled parameters, maximizing the Pearson correlation coefficient between observed and predicted values. For comparative purposes, predictive ability was defined as the Pearson correlation coefficient between the observed values and the predicted values for the validation data set.

The second GWAS was performed on the training population to attempt to reduce the number of markers used in the genome prediction model by selecting markers with relatively strong additive genetic effects (Figure S2). As the result of the marker selection, the average of percentage sharing SNP markers between the pair of two training populations was 23% for the 48 and 29% for the 192 selected SNP markers for D7 and 25% for the 48 and 28% for the 192 selected SNP markers for H7 (Tables S5, S6) The Additional models were generated by selecting the 48, 96, and 192 SNPs with the lowest *P*-values for each training population using FASTmrEMMA. Genomic prediction models were then built for each SNP subset in the same manner as the models for all SNPs. It should be noted that models using all SNPs were not generated using CNN1D and CNN2D because the structures of these methods are unsuitable for models with many SNPs.

2.8 Genomic prediction modeling methods used in this study

Genomic best linear unbiased prediction: Genomic best linear unbiased prediction (GBLUP; VanRaden, 2008) was implemented in R using the "mmer" function in the sommer package version 4.1.8 (Covarrubias-Pazaran, 2016; Covarrubias-Pazaran, 2018). An additive relationship matrix was generated using the "A.mat" function in sommer package for use in GBLUP models.

Bayesian linear methods: BayesA and BayesB (Meuwissen et al., 2001), Bayesian LASSO (BL; Park & Casella, 2008), BayesC (Habier et al., 2011), and Bayesian ridge regression (BRR; Pérez & de los Campos, 2014) were implemented in R using the "BGLR" function of the BGLR package. These models all have the form $y = 1\mu + \sum_{j=1}^{J} \boldsymbol{X}_j \boldsymbol{\beta}_j + \epsilon$ where y is the continuous response, μ is the intercept, $\boldsymbol{X}_j$ are predictors, $\boldsymbol{\beta}_j$ are vectors of effects, and ϵ is residuals. They were mainly characterized by their own prior distribution (Gianola, 2013).

Reproducing kernel Hilbert space regression: Reproducing kernel Hilbert space regression (RKHS; Wahba, 1978) was implemented in R using the "BGLR" function of the BGLR package. The resulting model has the form $y = 1\mu + \sum_{l=1}^{L} u_l + \epsilon$ where y is the continuous response, μ is the intercept, $\boldsymbol{u}_l$ is a kernel matrix of random effects, and ϵ is residuals. The distance matrix for observations was calculated using the "dist" function with the optional method = "euclidean". The bandwidth parameter was set to 1 when generating the kernel matrix.

Random forest: Random forest (RF; Breiman, 2001) was implemented in Python version 3.9.5 (Van Rossum & Drake, 2009) using "RandomForestRegressor" in scikit-learn library version 1.1.2 (Fabian et al., 2011). RF employs decision trees in which the objective variable is calculated by dividing the samples based on their features. These trees are built repeatedly using different sample sets sampled by bootstrap sampling. Final predictions are then obtained by averaging.

Gradient boosting decision trees: Two GBDT methods, eXtreme Gradient Boosting (XGB; Chen & Guestrin, 2016) and Light Gradient Boosting Machine (LGB; Ke et al., 2017) were implemented in Python using "XGBRegressor" in xgboost library version 1.6.2 and "LGBMRegressor" in lightgbm library version 3.3.2. Because both methods have many (8) parameters to optimize (Table S1), hyper-parameter optimization was performed to set their values. "TPESampler" in Optuna library version 3.0.2 (Akiba et al., 2019) was used to identify optimal parameters to avoid overfitting and obtain better predictions. XGB and LGB had several common parameters because both algorithms are decision trees, but each method also had some unique parameters chosen based on their specifications or importance. In particular, "max_depth" and "num_leaves" were considered important and their values were chosen to control tree complexity.

Deep learning: Two deep neural networks, CNN1D and CNN2D, were implemented in Python using the Keras library version 2.8.0 (Chollet, 2015). Their architectures are detailed in Table S3 and were inspired by the model of Simonyan and Zisserman (2015) that achieved remarkable success in image recognition. CNN2D differed from CNN1D with respect to input dimensions, convolution, and pooling layers because they have two dimensions in their nuclear phase. The nuclear phase dimension was summarized in the final pooling layer. CNN1D and CNN2D also had 7 common parameters that were optimized by Optuna (Table S2) using "TPESampler" due to their common architecture.

3 Results

3.1 Genotyping and genetic analysis

278,898,826,194 bases on 1,847,012,094 reads and 279,470,755,304 bases on 1,850,799,704 reads were obtained from the *Mse*I/*Pst*I and *Mlu*CI/*Bgl*II libraries, respectively, using the Illumina HiSeq X platform. After performing separate assembly and filtering steps for each library using the dDocent pipeline, the *Mse*I/*Pst*I and *Mlu*CI/*Bgl*II libraries produced 6,516,667 sites from 352 samples and 10,217,939 sites from 368 samples, respectively, which were stored in vcf file format. In the filtering step, some individuals were removed due to excessive missing data or mislabeling. The final vcf files for the *Mse*I/*Pst*I and *Mlu*CI/*Bgl*II libraries contained data on 11,425 and 10,605 polymorphic sites from 290 samples and were merged into a single vcf file that was then filtered using TASSEL 5.2 and imputed by beagle 5.4 to generate a vcf file representing 18,037 polymorphic sites from 290 samples that was used in all subsequent analyses.

Genetic structure was assessed by PCA using 1000 bp thinned genotype data containing 8709 SNPs. The first two principal components (PC1 and PC2) explained 4.58% and 2.60% of the genetic variance, respectively. Additionally, clusters of samples were observed in the left (PC1) and central (PC2) regions of the space (Figure 1). Linkage disequilibrium decay was assessed based on pairwise r^2 values between all SNPs located on the same scaffold. Figure 2 plots these r^2 values against the corresponding inter-SNP distances in base pairs and shows the associated trendline. An r^2 intercept line at 0.1 is also shown to estimate the LD decay of the *S. macrophylla* population. The regression line intersects this intercept line at a distance of 2336 bp.

3.2 Spatial analysis of phenotypic traits

A spatial structure analysis using a two-dimensional spline model was performed to detect and remove spatial bias in both focal traits (DBH and HT). Trees with lower values of both traits were concentrated on the left side of the progeny trial, which was captured by the analysis and the estimated spatial effect values were also smaller in this area. On the other hand, trees located around the center of the trial had higher phenotypic and spatial effect values. On the right side of the trial, there were some trees whose phenotypic values were lower, accentuating the relatively high phenotypic values of both traits in trees at the center of the trial area. Since the planting locations of progenies from each mother tree were arranged according to a RCBD, the spatial effects observed in the trial can be largely attributed to environmental heterogeneity. We therefore adjusted the raw phenotypic values to minimize the impact of this heterogeneity. The raw phenotypes, spatial effects, and adjusted phenotypes for D7 and H7 were visualized using heatmaps (Figure 3). The adjusted seventh-year phenotypes of trees with genotyping data were also visualized using histograms (Figure 4).

3.3 Genomic heritability

Narrow-sense genomic heritability for DBH and HT from year 1 to year 12.5 was calculated based on variance components obtained from Bayesian ridge regression for sample sets without missing data (Table 1). The highest genomic heritability values were 0.406 for DBH in year 1 and 0.395 for HT in year 11.5. The lowest values were 0.288 for DBH and 0.271 for HT, both in year 12.5. The narrow-sense genomic heritability of DBH was always higher than that of HT except in years 2 and 11.5.

3.4 Genome-wide association study using all individuals

P-values and LOD scores of SNPs for D7 and H7 were calculated using FASTmrEMMA for the sample sets considered in the genomic heritability analysis. The results were visualized using a Q-Q plot and a Manhattan plot showing all *P*-values and LOD scores above 3, which was taken as the significance threshold (Figure 5). For D7, three SNPs had significant LOD scores but non-significant $-\log_{10}(P)$ values (4.0052, 3.8663 and 6.4149) at the 5% statistical significance level after FDR correction. For H7, five SNPs had significant LOD scores but non-significant $-\log_{10}(P)$ values (5.8854, 4.7025, 5.1203, 3.7573, and 3.88) at the 5% statistical significance level after FDR correction (Figure 5).

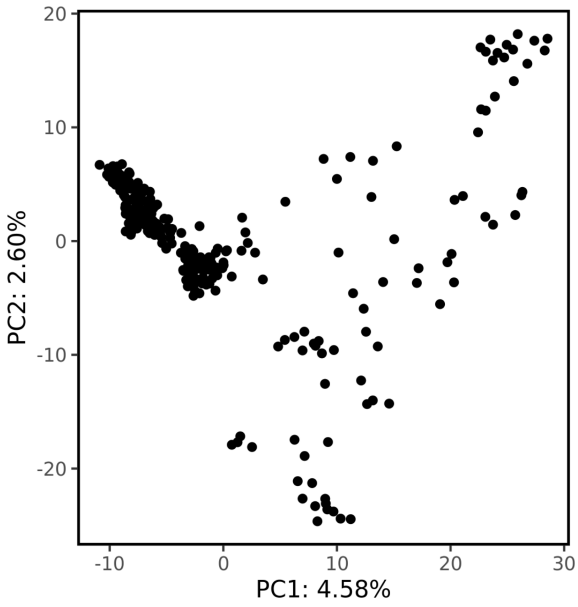


FIGURE 1
The population structure revealed by principal component analysis. The proportion of the total variance explained by each principal component is shown on the axes.

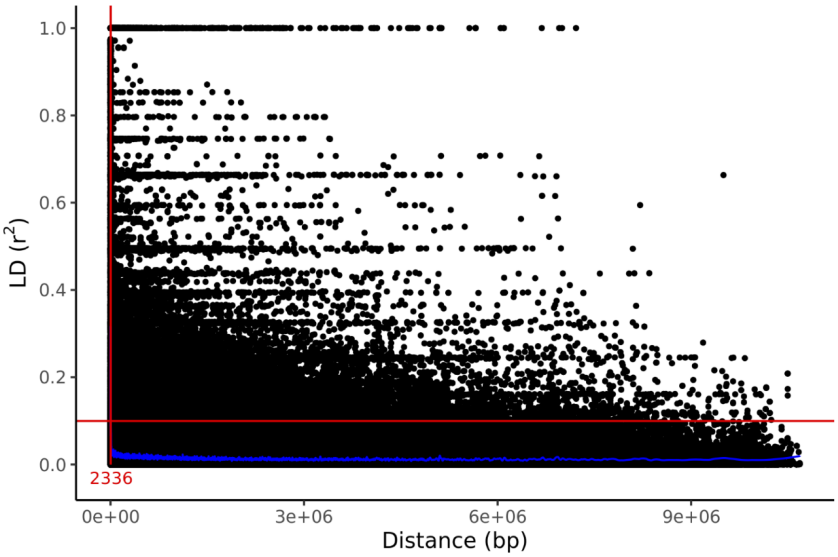


FIGURE 2
LD decay in the studied population indicated by a scatter plot of r^2 against the inter-SNP distance. The figure also shows the LOESS regression line (in blue), an intercept at r^2 = 0.1, and the distance at their intersection (red).

3.5 Comparison of genomic prediction modeling methods

8480 genomic prediction models were built using 12 models (with 10 model building replicates, 10 sample split replicates, 4 different numbers of selected SNPs based on the *p* value of the second GWAS using each of the training populations, and 2 focal traits per method). The median of predictive abilities within model building replicates and sample split replicates were positive and majority of the predictive abilities was significantly deviated from zero except CNN1D of D7 (96, 192, all SNPs), GBLUP of H7 (all SNPs), LGB of D7 (all SNPs), RKHS of D7 (96 SNPs), XGB of D7 (all SNPs), XGB of H7 (all SNPs) (Table 2, Table S7). The highest median predictive ability for D7 was 0.188, which was achieved using RF with the full set of SNPs. For H7, the highest median predictive ability was 0.231, which was achieved using RKHS with 48 selected SNPs. Predictive accuracies could not be calculated for H7 using three of the GBLUP models using the full set of SNPs because their predictions were identical to those for the test set (shown in parentheses in Table 2). When the median of predictive accuracy was compared among the selected number of markers, the

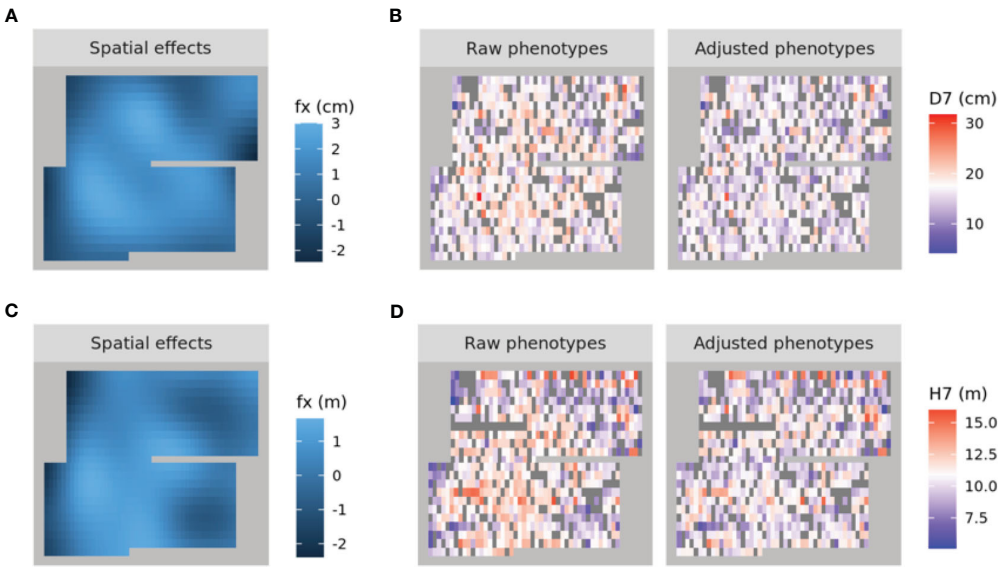


FIGURE 3
Heatmaps of the spatial effects for D7 **(A)** and H7 **(C)** as well as the corresponding raw phenotypes and adjusted phenotypes (**B, D**, respectively). In **(A)** and **(C)**, higher to lower values are represented by shades ranging from pale to dark blue. In **(B)** and **(D)**, higher to lower values are represented by shades ranging from red to blue, with white denoting intermediate values. NA values are shown in dark grey.

10.3389/fpls.2023.1241908

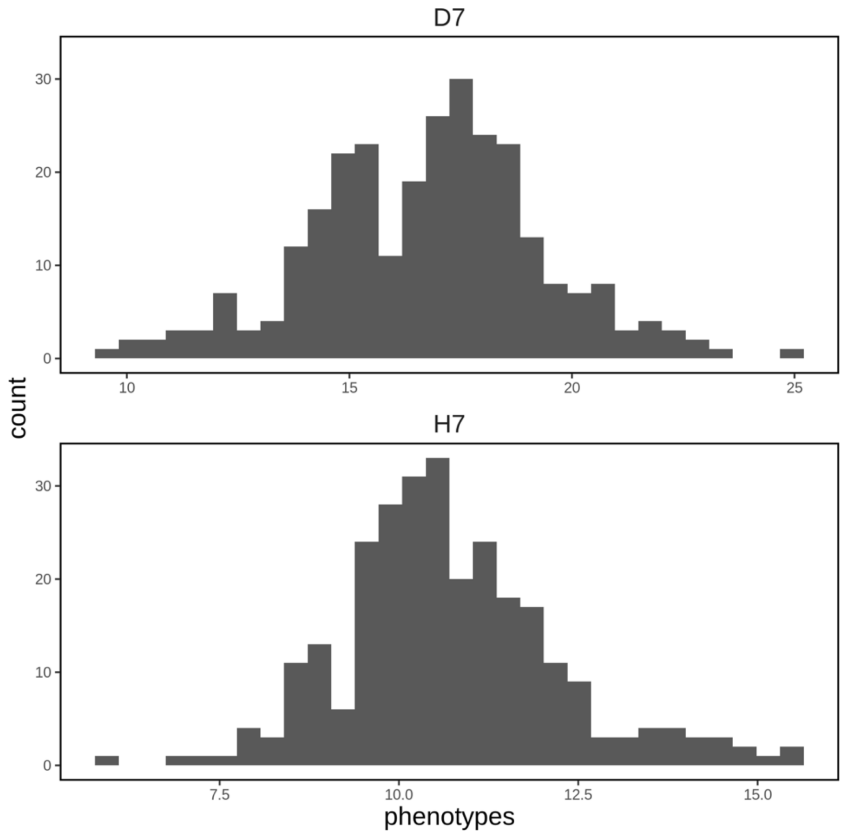


FIGURE 4
Histograms of adjusted phenotypes for D7 and H7 that were used to construct the genomic prediction model. The units of the horizontal axis are cm for D7 and m for H7.

higher number of SNPs tended to yield higher estimation accuracy for D7 for all methods except LGB and CNN2D. On the other hand, for H7, a lower number of SNPs tended to give higher estimation accuracy for all methods.

4 Discussion

4.1 Linkage disequilibrium analysis and genome-wide association study

After accounting for linkage disequilibrium decay, the approximate genomic coverage was estimated to be $2236 \times 8709 = 19,473,324$ bp based on the intersection of the regression line and the $r^2 = 0.1$ line (2336 bp) and the number of SNPs (8709 SNPs) after thinning the genotype data at every 1000 bp. The *S. macrophylla* genome size was estimated to be 363.816 Mbp using the formula of Doležel et al. (2003) because the *C*-value of *S. macrophylla* is 0.372 (Ng et al., 2016). Therefore, the genotype data in this study represents around 5.35% of the total genome. The GWAS for D7 and H7 detected 3 and 5 SNPs with LOD scores above 3, respectively, but no significant SNPs were detected in terms of *P*-values (Figure 5). The inability to identify QTLs for D7 and H7 may be due to the sparsity of marker coverage within the genome and the fact that both traits are probably regulated by complex genetic backgrounds that are not readily characterized. In order to obtain ideal results in GWAS, it is necessary to further increase the marker density and the power of LD detection by increasing the number of individuals.

4.2 Comparison of methods for building genomic prediction models and population subdivision

GBLUP achieved maximum predictive abilities of 0.176 for D7 and 0.206 for H7. Notably, when considering only results with

TABLE 1 Narrow-sense genomic heritability for diameter at breast height (DBH) and tree height (HT) from year 1 to year 12.5.

year	1	2	3	4	7	9	11.5	12.5
DBH	**0.406**	0.304	0.312	0.373	0.366	0.350	0.391	0.288
HT	0.303	0.324	0.284	0.304	0.358	0.312	**0.395**	0.271

The highest genomic heritability values on DBH and HT were shown in bold.

10.3389/fpls.2023.1241908

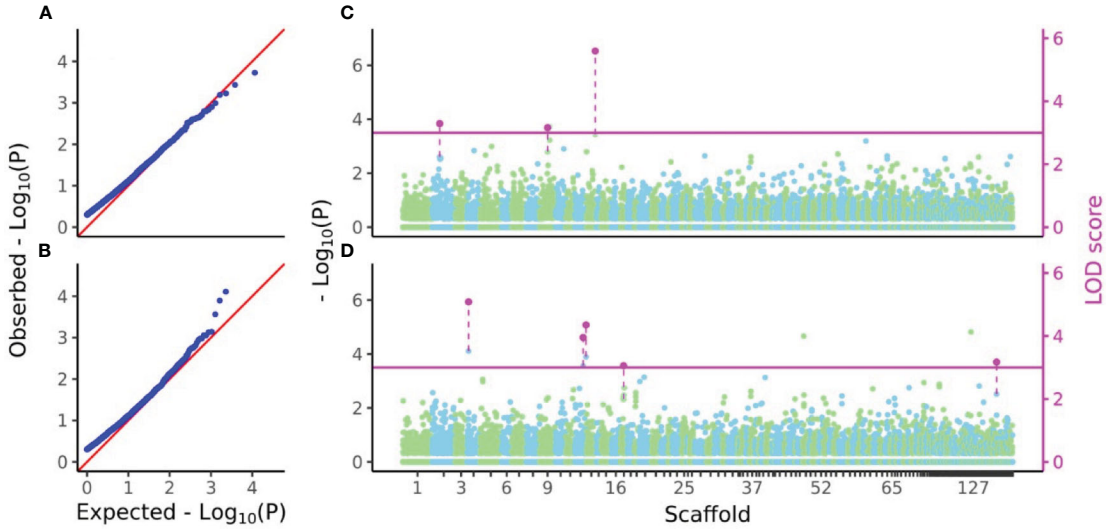


FIGURE 5
Q-Q and Manhattan plots of D7 and H7. **(A, B)** Q-Q plots of $-\log_{10}(P)$ values from GWAS of D7 and H7. **(C, D)** Manhattan plots of $-\log_{10}(P)$ values and LOD scores from FASTmrEMMA of D7 and H7. In **(A)** and **(B)**, zero values were removed from all $-\log_{10}(P)$ values before calculation of expected values. In **(C)** and **(D)**, all $-\log_{10}(P)$ values of SNPs were separately plotted in light green and sky blue in color which represents localization of SNPs to each scaffold. Only LOD scores above the threshold 3 were shown in magenta and connected by dashed line with the plots of $-\log_{10}(P)$ values of the same SNPs.

TABLE 2 Median predictive ability values of genomic prediction models when used with 48, 96, and 192 or with the full set of 18037 SNPs.

Traits	Number of SNPs	Predictive ability (median)					
		GBLUP	BayesA	BayesB	BayesC	BRR	BL
D7	48	0.100	0.087	0.086	0.085	0.087	0.088
	96	0.125	0.116	0.122	0.121	0.113	0.119
	192	**0.176**	0.154	0.152	0.156	0.157	0.152
	18037	0.130	0.134	0.133	0.134	0.132	0.136
H7	48	0.206	0.211	0.200	0.200	**0.212**	0.208
	96	0.201	0.209	0.206	0.204	0.207	0.209
	192	0.160	0.157	0.174	0.167	0.146	0.167
	18037	0.056	0.110	0.108	0.109	0.108	0.107
Traits	**Number of SNPs**	**Accuracy (median)**					
		RKHS	RF	LGB	XGB	CNN1D	CNN2D
D7	48	0.086	0.121	**0.167**	0.119	0.062	0.108
	96	0.106	0.125	**0.142**	0.093	0.039	0.073
	192	0.136	0.164	0.155	0.120	0.082	0.098
	18037	0.140	**0.188**	0.119	0.090	–	–
H7	48	**0.231**	0.204	0.162	0.181	0.179	0.184
	96	0.209	0.195	0.157	0.191	**0.218**	0.199
	192	0.159	0.165	0.142	0.148	**0.200**	0.173
	18037	0.125	**0.151**	0.092	0.075	–	–

Abbreviations of methods are as follows: GBLUP for genomic best linear unbiased prediction, BRR for Bayesian ridge regression, BL for Bayesian LASSO, RF for random forest, LGB for light gradient boosting machine, XGB for extreme gradient boosting, CNN1D for convolutional neural network for one-dimensional input and CNN2D for convolutional neural network for two-dimensional input. Predictive ability was defined as the correlation coefficient between the observed and predicted values for each model. The tabulated values are medians of 100 replicates based on 10 pairs of training and validation sets and 10 model construction replicates. The highest accuracy values for each trait and number of SNPs are shown in bold.

selected SNPs, the value for D7 was the highest achieved with any method in this work (Table 2). Similarly, in a GS study on Japanese cedar, GBLUP achieved greater accuracy than BayesB for most traits including growth, wood properties, and male fertility. BayesB is considered to be effective only in cases involving a small number of strongly linked QTLs (Jannink et al., 2010), so the superior performance of GBLUP suggests that the focal traits in this work are governed by large numbers of QTLs (Hiraoka et al., 2018). Therefore, the selected SNPs (192) are possibly linked to many QTLs in D7, which contributes to the higher predictive ability of GBLUP than BayesB. Among the linear Bayesian methods (BayesA, BayesB, BayesC, BRR, and BL), BRR had the best predictive ability for both D7 (0.157) and H7 (0.212) but only outperformed the next best models (BayesC and BayesA, respectively) by 0.001 for both traits, and there were only minor differences in accuracy between all of these methods with the exception of BayesB (Table 2). An earlier study on cattle similarly found no difference in predictive ability between the Bayesian methods, which was consistent with a simulated scenario involving multiple QTLs (Foroutaifar, 2020). BayesB achieved a higher predictive ability for H7 when using a relatively large number of SNPs (192 SNPs) and a lower accuracy when using a relatively small number of SNPs (48 SNPs) when compared to other Bayesian linear methods. BayesB has also proved to be an effective method for traits dominated by QTL with large effects in several studies (Daetwyler et al., 2010; Wang et al., 2019). We therefore suggest that the difference between BayesB and other Bayesian linear methods was due to its reliance on the assumption that there were many SNPs with no genetic variance (Meuwissen et al., 2001). RKHS achieved a maximum predictive ability of 0.140 for D7 and 0.231 for H7 and offered the best performance for H7 of all the tested methods (Table 2). This is consistent with an earlier GS study on pine, where RKHS was slightly more accurate than BayesA for DBH, HT, and simulated traits including non-additive genetic effects. This outcome is also consistent with the argument that RKHS non-explicitly accounts for non-additive genetic effects (de Almeida Filho et al., 2019) and suggests that such effects have an important influence on H7. The predictive ability achieved for H7 when using RKHS was only slightly sensitive to the number of selected SNPs but was highest when using only 48. This may be because the relative influence of non-additive genetic effects was increased when restricting the analysis to the SNPs with the largest effects. RF achieved a maximum predictive ability of 0.188 for D7 when using the full set of SNPs, and 0.204 for H7 when using 48 selected SNPs. RF achieved the highest predictive ability for D7 out of all the methods tested in this work. The predictive ability for H7 was always highest when using the full set of SNPs, irrespective of the choice of method (Table 2). RF also achieved better accuracy than BayesB and GBLUP in genomic predictions for cedar growth and woody strength (Hiraoka et al., 2018). Nonlinear methods such as RF should be more effective when the relationship between the feature and objective variables is nonlinear, which is expected to be the case when epistatic effects account for the majority of genetic variance (Jannink et al., 2010). D7 and H7 may thus have epistatic genetic structures.

No studies have yet investigated whether observed GP model performance can reliably be extrapolated to next generation of a training population. However, since the test population in this work consisted of just 71 individuals, it is very likely that predictive ability was influenced by the relationship between the training and test populations. In an earlier study on maize, the highest predictive ability was achieved with models based on a dataset in which both parents were shared by the training and test populations, followed by a dataset where only the mothers were shared. Performance was significantly worse for datasets that shared only fathers or with no shared parents (Yan et al., 2021). The progeny trial examined in this work was produced from 94 mother trees from a single natural forest with up to 5 siblings each, and test population consisted of 71 randomly selected trees. It is therefore very likely that the bias of the shared mother trees affected the accuracy of the predictions. More accurate predictions would probably be obtained if siblings from the same mother trees were distributed evenly between the test and training populations.

4.3 Performance of deep learning and gradient boosting decision trees for genomic prediction

CNN1D achieved maximum predictive abilities of 0.082 for D7 and 0.218 for H7. It also achieved the highest predictive ability of any method when using the 96 and 192 SNP datasets for H7 and the lowest predictive ability of any method when using the 48, 96, and 192 SNP datasets for D7 (Table 2). Similar results were obtained in a GS study on strawberries and blueberries, in which CNN exhibited better predictive ability than linear models for traits with relatively strong epistatic effects (Zingaretti et al., 2020). In addition, the predictive ability achieved for H7 with RKHS was highest for the 48 SNP dataset and declined as the number of SNPs increased. This implies that the epistatic effects in question are particularly strong among the small number of genes close to the selected 48 SNPs. The predictive accuracy achieved for H7 with CNN1D was highest when using the 96 and 192 SNP datasets, indicating that more genes in the vicinity of these SNPs are involved in regulating this trait. Maximum predictive abilities of 0.108 for D7 and 0.199 for H7 were achieved with CNN2D, in which the haplotype patterns were modeled using the CNN filter (Table 2). Although the basic structure of the CNN2D model is similar to that of the CNN1D model, the inclusion of nuclear phase information estimated by beagle seems to weaken the trend observed with the latter model due to the greater complexity of CNN2D. LGB achieved a maximum predictive ability of 0.167 for D7 and 0.162 for H7. It also had the highest predictive ability of any method for D7 when using the 48 and 96 SNP datasets (Table 2). An earlier GS study on maize suggested that LGB provides better accuracy than other GBDTs and rrBLUP, and that genome-wide epistatic interactions can be cumulatively learned by LGB given a sufficiently large population (Yan et al., 2021). However, the number of samples in this work (281) was much smaller than in the maize study and may have been insufficient for LGB to cumulatively learn the genome-wide epistasis interactions. The high accuracy achieved for D7 with the 96 and 192 SNP datasets may also be partly due to the bimodal distribution of the phenotypic values for this trait (Figure 4). Most

models considered in this work assume that the values of the objective variable are normally distributed, and their accuracy may suffer if this assumption is violated. However, LGB models can predict non-unimodal traits without issue. XGB achieved a maximum predictive ability of 0.120 for D7 and 0.191 for H7; both of these values are lower than those achieved with linear models in most cases (Table 2). However, in genomic predictions using simulated data, the predictive ability of XGB exceeded those of BayesB, GBLUP, RF, CNN, and MLP when dominant and epistasis effects were present in addition to additive genetic effects (Abdollahi-Arpanahi et al., 2020). However, XGB achieved poor predictive ability when applied to real data in this work even though the results obtained with RKHS and RF suggest that the selected SNPs (for H7) and all SNPs (for both D7 and H7) have nonadditive genetic structures. Given the very large sample size of the simulation study (over 10,000 samples), this discrepancy may be due to the low number of samples included in this work.

4.4 Influence of marker selection on genomic prediction and detection of genetic effects

Predictive ability varied with the number of included SNPs in the order 18037 (all) > 192 > 96 > 48 SNP for D7 and 48 > 96 > 192 > 18037 (all) for H7, and its difference were statistically significant in a few modeling methods (Table 2). The Q-Q plots of $-\log_{10}(P)$ values suggest that false positives did not unduly affect the GWAS results for any training population (Figure S2). For H7, the higher predictive ability of the models with SNP selection compared to those using the full SNP dataset may be due to the elimination of noise introduced by unlinked SNPs. ddRAD-seq randomly selects DNA fragments from the whole genome with a certain density and therefore includes some SNPs that are not linked to the target trait. The $-\log_{10}(P)$ values of such SNPs should be very small in GWAS, and removing such SNPs should increase model accuracy. The improvement in prediction accuracy due to the decrease in the number of markers was particularly noticeable in H7, and tree height might be relatively less complex than tree diameter D7. Although the complexity of gene regulation in primary and secondary growth should vary depending on tree species, secondary growth is likely to involve a large number of genes compared to primary growth due to the greater diversity of cell types and processes associated with the secondary growth in general (Oh et al., 2003; Demura and Fukuda, 2007). The difference of complexity of gene regulation might be main reason of the opposite tendencies between H7 and D7.

Other genomic prediction studies on disease resistance in aquaculture species have similarly found that SNP marker selection improves the accuracy of genomic prediction models (Luo et al., 2021). The predictive ability for D7 exceeded that for H7 irrespective of the chosen method when using the full SNP dataset (Table 2), possibly because the genetic heritability of D7 exceeds that of H7. However, the highest accuracy for H7 was achieved when using 48 selected SNPs whereas the highest accuracy for D7 was achieved with the full set of SNPs (Table 2). This may be related to the fact that the phenotypic values for H7 were normally distributed whereas those for D7 were bimodal (Figure 4). Because the first stage in the SNP analysis was performed using a linear model with FASTmrEMMA, it is possible that the pipeline was best suited for pseudo-normally distributed focal variables like H7 and was thus more effective at detecting significant SNPs for H7 than for D7.

The differences in predictive accuracy between semiparametric and nonparametric methods were greater than those between parametric methods (Table 2). Parametric methods only consider additive genetic effects, whereas nonparametric methods do not exclude other genetic effects such as dominance and epistasis and thus produce models that can account for a wider range of genetic effects. This is consistent with the findings of Howard et al. (2014), whose studies using simulated data showed that four semiparametric/nonparametric methods achieved markedly differing accuracies that were related to their ability to explain both additive and epistatic genetic structure. The differences in predictive ability between the tested methods thus suggest that the SNPs selected for H7 have a relatively strong epistatic genetic structure but those selected for D7 may have a more additive genetic structure. These results show that to maximize predictive ability when constructing a genomic prediction model based on GWAS, it is important to select a method appropriate for the genetic structure of the selected SNPs.

5 Conclusion

Genomic prediction models were constructed for two growth traits in the tropical timber tree species *S. macrophylla*: DBH (D7) and HT (H7) in the seventh year from transplantation at the PT. Sari Bumi Kusuma forest concession in central Kalimantan, Indonesia. Unlike species with established lineages and pedigrees such as crops and livestock, the studied progeny trial retains high genetic diversity due to the large effective population size of mother trees and the reliance on open-pollinated mating to regenerate the reproductive source of the trial. Despite this high genetic diversity, the measured values correlated positively with the outputs of genomic prediction models for D7 and H7. When GWAS data were used to select SNP subsets, different sets of SNPs were selected for each split of the training/test population and the highest predictive accuracies were achieved with the full set of SNPs for D7 and 48 selected SNPs for H7, these opposite tendencies might be originated from difference of genetic architecture between primary and secondary growth of the species. These results show that the SNP subset with the highest predictive ability in H7 can be used for genotyping next-generation populations in breeding programs in order to reduce costs while maximizing genetic gains. Although it is necessary to further increase the marker density and the power of LD detection by increasing the number of individuals, the genomic prediction models and subsequent selection at seedling stage using the models show potential to accelerate breeding cycles for non-cultivated tree species.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA988453/; https://doi.org/10.5281/zenodo.8051012; https://doi.org/10.5061/dryad.kkwh70s8d.

Author contributions

NT, KU, and YT conceived the ideas of the project. MN and SP established progeny trials and scientific data from the progeny trials were obtained by MN, SP, W, and SI, and NT, MN, W, SI, S, and SP carried out field work. NT, MN, W, SI, S, and YT contributed to searching for funds. HA, NT and KU performed the experiment and data analysis and HA and NT wrote the first draft of the manuscript and contributed to the writing and revision of the manuscript. All authors have approved the manuscript for publication. All authors contributed to the article.

Funding

This research was partly supported by a JIRCAS-UGM joint project entitled "Enhancement of productivity using genetic resources in tropical rainforest and development of carbohydrate usage from unutilized biomass in Indonesia (grant number a1A401b; JIRCAS) and the Science and Technology Research Partnership for Sustainable Development (SATREPS; JPMJSA2101).

Acknowledgments

The authors would like to thank the technical staff of the progeny trial in PT-Sari Bumi Kusuma for their invaluable contributions during sample collection and phenotype measurement, Drs. Y. Yamamoto, H. Oka, K. Hayashi, S. Noguchi for project management, Ms. D. Y. Li for DNA extraction, Ms. R. Yamada for project logistics, Prof. Dr. Kamijo for his advice on tropical forestry and Prof. Dr. H. Iwata for his advice on genomic prediction models.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1241908/full#supplementary-material

References

Abdollahi-Arpanahi, R., Gianola, D., and Peñagaricano, F. (2020). Deep learning versus parametric and ensemble methods for genomic prediction of complex phenotypes. *Genet. Selection Evol.* 52 (1), 1–15. doi: 10.1186/s12711-020-00531-z

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2623–2631. doi: 10.1145/3292500.3330701

Azodi, C. B., Bolger, E., McCarren, A., Roantree, M., de los Campos, G., and Shiu, S. H. (2019). Benchmarking parametric and machine learning models for genomic prediction of complex traits. *G3: Genes Genomes Genet.* 9 (11), 3691–3702. doi: 10.1534/g3.119.400498

Bellot, P., de los Campos, G., and Pérez-Enciso, M. (2018). Can deep learning improve genomic prediction of complex human traits? *Genetics* 210 (3), 809–819. doi: 10.1534/genetics.118.301298

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23 (19), 2633–2635. doi: 10.1093/bioinformatics/btm308

Breiman, L. (2001). "Random Forest," in *Machine Learning*. Springer, New York. doi: 10.1007/978-3-030-62008-0_35

Browning, B. L., Tian, X., Zhou, Y., and Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* 108 (10), 1880–1890. doi: 10.1016/j.ajhg.2021.08.005

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103 (3), 338–348. doi: 10.1016/j.ajhg.2018.07.015

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *J. Assoc. Physicians India*. 785–794. doi: 10.1145/2939672.2939785

Chen, Z. Q., Baison, J., Pan, J., Westin, J., Gil, M. R. G., and Wu, H. X. (2019). Increased prediction ability in Norway spruce trials using a marker X environment interaction and non-additive genomic selection model. *J. Heredity* 110 (7), 830–843. doi: 10.1093/jhered/esz061

Chen, Z., Helmersson, A., Westin, J., Karlsson, B., and Wu, H. X. (2018). Efficiency of using spatial analysis for Norway spruce progeny tests in Sweden. *Ann. For. Sci.* 75 (1). doi: 10.1007/s13595-017-0680-8

Chollet, F. (2015) *Keras*. Available at: https://github.com/fchollet/keras.

Corriyanti, C., and Muharyani, N. (2020). The opportunities and challenges of jati plus perhutani. *Wood Res. J.* 9 (1), 1–3. doi: 10.51850/wrj.2018.9.1.1-3

Covarrubias-Pazaran, G. (2016). Genome-Assisted prediction of quantitative traits using the r package sommer. *PloS One* 11 (6), e0156744. doi: 10.1371/journal.pone.0156744

Covarrubias-Pazaran, G. (2018). Software update: Moving the R package sommer to multivariate mixed models for genome-assisted prediction. *BioRxiv*. doi: 10.1101/354639

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., and Woolliams, J. A. (2010). The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185 (3), 1021–1031. doi: 10.1534/genetics.110.116855

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi: 10.1093/bioinformatics/btr330

de Almeida Filho, J. E., Guimarães, J. F. R., Fonsceca E Silva, F., De Resende, M. D. V., Muñoz, P., Kirst, M., et al. (2019). Genomic prediction of additive and non-additive effects using genetic markers and pedigrees. *G3: Genes Genomes Genet.* 9 (8), 2739–2748. doi: 10.1534/g3.119.201004

Demura, T., and Fukuda, H. (2007). Transcriptional regulation in wood formation. *Trends Plant Sci.* 12 (2), 64–70. doi: 10.1016/j.tplants.2006.12.006

Doležel, J., Bartoš, J., Voglmayr, H., and Greilhuber, J. (2003). Letter to the editor. *Cytometry* 51A (2), 127–128. doi: 10.1002/cyto.a.10013

e Sousa, M. B., Galli, G., Lyra, D. H., Granato, Í.S.C., Matias, F. I., Alves, F. C., et al. (2019). Increasing accuracy and reducing costs of genomic prediction by marker selection. *Euphytica* 215 (2), 18. doi: 10.1007/s10681-019-2339-z

Fabian, P., Gaël, V., Alexandre, G., Vincent, M., Bertrand, T., Olivier, G., et al. (2011). Scikit-learn: machine learning in python. *J. OfMachine Learn. Res.* 12 (85), 2825–2830. doi: 10.1289/EHP4713

Foroutaifar, S. (2020). Accuracy and sensitivity of different Bayesian methods for genomic prediction using simulation and real data. *Stat. Appl. Genet. Mol. Biol.* 19 (3), 20190007. doi: 10.1515/sagmb-2019-0007

Garrison, E., Kronenberg, Z. N., Dawson, E. T., Pedersen, B. S., and Prins, P. (2022). A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PloS Comput. Biol.* 18 (5), 1–14. doi: 10.1371/journal.pcbi.1009123

Ghazoul, J. (2016). *Dipterocarp Biology, Ecology, and Conservation* (Oxford, UK: Oxford University Press). doi: 10.1093/acprof:oso/9780199639656.001.0001

Gianola, D. (2013). Priors in whole-genome regression: The Bayesian alphabet returns. *Genetics* 194 (3), 573–596. doi: 10.1534/genetics.113.151753

González-Camacho, J. M., de los Campos, G., Pérez, P., Gianola, D., Cairns, J. E., Mahuku, G., et al. (2012). Genome-enabled prediction of genetic values using radial basis function neural networks. *Theor. Appl. Genet.* 125 (4), 759–771. doi: 10.1007/s00122-012-1868-9

Grattapaglia, D., and Kirst, M. (2008). Eucalyptus applied genomics: From gene sequences to breeding tools. *New Phytol.* 179 (4), 911–929. doi: 10.1111/j.1469-8137.2008.02503.x

Grattapaglia, D., Silva-Junior, O. B., Resende, R. T., Cappa, E. P., Müller, B. S. F., Tan, B., et al. (2018). Quantitative genetics and genomics converge to accelerate forest tree breeding. *Front. Plant Sci.* 871. doi: 10.3389/fpls.2018.01693

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

Hiraoka, Y., Fukatsu, E., Mishima, K., Hirao, T., Teshima, K. M., Tamura, M., et al. (2018). Potential of genome-wide studies in unrelated plus trees of a coniferous species, cryptomeria japonica (Japanese cedar). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01322

Howard, R., Carriquiry, A. L., and Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes Genomes Genet.* 4 (6), 1027–1046. doi: 10.1534/g3.114.010298

Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: From theory to practice. *Briefings Funct. Genomics Proteomics* 9 (2), 166–177. doi: 10.1093/bfgp/elq001

Jones, T. H., Steane, D. A., Jones, R. C., Pilbeam, D., Vaillancourt, R. E., and Potts, B. M. (2006). Effects of domestication on genetic diversity in Eucalyptus globulus. *For. Ecol. Manage.* 234 (1–3), 78–84. doi: 10.1016/j.foreco.2006.06.021

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* 2017, 3147–3155.

Knaus, B. J., and Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17 (1), 44–53. doi: 10.1111/1755-0998.12549

Lebedev, V. G., Lebedeva, T. N., Chernodubov, A. I., and Shestibratov, K. A. (2020). Genomic selection for forest tree improvement: Methods, achievements and perspectives. *Forests* 11 (11), 1190. doi: 10.3390/f11111190

Lee, H. S., Itoh, A., Kanzaki, M., and Yamakura, T. (1997). Height growth of engkabang jantong, shorea macrophylla (De vr.) ashton, in a plantation forest in sarawak. *Tropics* 7 (1/2), 67–80. doi: 10.3759/tropics.7.67

Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., and Li, Y. (2018). Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Front. Genet.* 9. doi: 10.3389/fgene.2018.00237

Luo, Z., Yu, Y., Xiang, J., and Li, F. (2021). Genomic selection using a subset of SNPs identified by genome-wide association analysis for disease resistance traits in aquaculture species. *Aquaculture* 539, 736620. doi: 10.1016/j.aquaculture.2021.736620

Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J., et al. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta* 248 (5), 1307–1318. doi: 10.1007/s00425-018-2976-9

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4), 1819–1829. doi: 10.1093/genetics/157.4.1819

Munoz, F., and Sanchez, L. (2020) *breedR: Statistical Methods for Forest Genetic Resources Analysts*. Available at: https://github.com/famuvie/breedR.

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8 (19), 4321–4326. doi: 10.1093/nar/8.19.4321

Ng, C. H., Lee, S. L., Tnah, L. H., Ng, K. K. S., Lee, C. T., and Madon, M. (2016). Genome size variation and evolution in Dipterocarpaceae. *Plant Ecol. Diversity* 9 (5–6), 437–446. doi: 10.1080/17550874.2016.1267274

Ng, K. K. S., Kobayashi, M. J., Fawcett, J. A., et al (2021). The genome of Shorea leprosula (Dipterocarpaceae) highlights the ecological relevance of drought in aseasonal tropical rainforests. *Commun Biol* 4, 1166. doi: 10.1038/s42003-021-02682-1

Ober, U., Erbe, M., Long, N., Porcu, E., Schlather, M., and Simianer, H. (2011). Predicting genetic values: A kernel-based best linear unbiased prediction with genomic data. *Genetics* 188 (3), 695–708. doi: 10.1534/genetics.111.128694

Oh, S., Park, S., and Han, K. H. (2003). Transcriptional regulation of secondary growth in Arabidopsis thaliana. *J. Exp. Bot.* 54 (393), 2709–2722. doi: 10.1093/jxb/erg304

Park, T., and Casella, G. (2008). The bayesian lasso. *J. Am. Stat. Assoc.* 103 (482), 681–686. doi: 10.1198/016214508000000337

Pérez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198 (2), 483–495. doi: 10.1534/genetics.114.164442

Pérez-Enciso, M., and Zingaretti, L. M. (2019). A guide for using deep learning for complex trait genomic prediction. *Genes* 10 (7), 19. doi: 10.3390/genes10070553

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PloS One* 7 (5), e37135. doi: 10.1371/journal.pone.0037135

Puritz, J. B., Hollenbeck, C. M., and Gold, J. R. (2014). *dDocent* : a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* 2, e431. doi: 10.7717/peerj.431

Puritz, J. B., Suchan, T., Dimens, P. V., and Chollenbeck, (2018) *dDocent_filters (2.5.2)*. Available at: https://github.com/jpuritz/dDocent/blob/v2.5.2/scripts/dDocent_filters.

Randi, A., Julia, S., Kusumadewi, Y., Robiansyah, I., Shomat, F., Tanggaraju, S., et al. (2019). *Shorea macrophylla, Light Red Meranti THE IUCN RED LIST OF THREATENED SPECIES*^{TM}. doi: 10.13140/RG.2.2.33146.03529

R Core Team (2020). *R: A language and environment for statistical computing. (4.0.2)* (R Foundation for Statistical Computing). Available at: https://www.r-project.org/.

Sawitri, Tani, N., Na'iem, M., Widiyatno, Indrioko, S., Uchiyama, K., et al. (2020). Potential of Genome-Wide association studies and Genomic Selection to improve productivity and quality of commercial timber species in tropical rainforest, a case study of Shorea platyclados. *Forests* 11 (2), 239. doi: 10.3390/f11020239

Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. 1–14.

Tamba, C. L., and Zhang, Y. M. (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. *BioRxiv*. doi: 10.1101/341784

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91 (11), 4414–4423. doi: 10.3168/jds.2007-0980

Van Rossum, G., and Drake, F. L. (2009). *Python 3 Reference Manual* (Scotts Valley, CA: CreateSpace).

Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Front. Genet.* 9. doi: 10.3389/fgene.2018.00078

Wahba, G. (1978). Improper Priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Stat. Society. Ser. B (Methodological)* 40 (3), 364–372. doi: 10.1111/j.2517-6161.1978.tb01050.x

Wang, X., Miao, J., Chang, T., Xia, J., An, B., Li, Y., et al. (2019). Evaluation of GBLUP, BayesB and elastic net for genomic prediction in Chinese simmental beef cattle. *PloS One* 14 (2), 1–14. doi: 10.1371/journal.pone.0210442

Wen, Y. J., Zhang, H., Ni, Y. L., Huang, B., Zhang, J., Feng, J. Y., et al. (2018). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings Bioinf.* 19 (4), 700–712. doi: 10.1093/bib/bbw145

Widiyatno, Naiem, M., Purnomo, S., and Jatmoko, (2014). Evaluation of four years old progeny test of shoreamacrophylla in PT sari bumi kusuma, central kalimantan. *Proc. Environ. Sci.* 20, 809–815. doi: 10.1016/j.proenv.2014.03.098

Yan, J., Xu, Y., Cheng, Q., Jiang, S., Wang, Q., Xiao, Y., et al. (2021). LightGBM: accelerated genomically designed crop breeding through ensemble learning. *Genome Biol.* 22 (1), 1–24. doi: 10.1186/s13059-021-02492-y

Zingaretti, L. M., Gezan, S. A., Ferrão, L. F. V., Osorio, L. F., Monfort, A., Muñoz, P. R., et al. (2020). Exploring deep learning for complex trait genomic prediction in polyploid outcrossing species. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00025

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 14 November 2023
DOI 10.3389/fpls.2023.1196561

OPEN ACCESS

EDITED BY
Quanzi Li,
Chinese Academy of Forestry, China

REVIEWED BY
Jose Luis Cabrera Ponce,
Unidad Irapuato (CINVESTAV), Mexico
Jens Stougaard,
Aarhus University, Denmark

*CORRESPONDENCE
Bingsong Zheng
bszheng@zafu.edu.cn
Xiaofei Wang
xfwang@zafu.edu.cn

†These authors have contributed equally to this work

RECEIVED 30 March 2023
ACCEPTED 20 October 2023
PUBLISHED 14 November 2023

CITATION
Ying W, Wen G, Xu W, Liu H, Ding W, Zheng L, He Y, Yuan H, Yan D, Cui F, Huang J, Zheng B and Wang X (2023) *Agrobacterium rhizogenes*: paving the road to research and breeding for woody plants. *Front. Plant Sci.* 14:1196561. doi: 10.3389/fpls.2023.1196561

Agrobacterium rhizogenes: paving the road to research and breeding for woody plants

Wei Ying[1,2†], Guangchao Wen[1,2†], Wenyuan Xu[1,2], Haixia Liu[1,2], Wona Ding[3], Luqing Zheng[4], Yi He[1,2], Huwei Yuan[1,2], Daoliang Yan[1,2], Fuqiang Cui[1,2], Jianqin Huang[1], Bingsong Zheng[1,2*] and Xiaofei Wang[1,2*]

[1]State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, Zhejiang, China, [2]Zhejiang Provincial Key Laboratory of Forest Aromatic Plants-based Healthcare Functions, Zhejiang A&F University, Hangzhou, Zhejiang, China, [3]College of Science and Technology, Ningbo University, Ningbo, Zhejiang, China, [4]College of Life Sciences, Nanjing Agricultural University, Nanjing, Jiangsu, China

Woody plants play a vital role in global ecosystems and serve as valuable resources for various industries and human needs. While many woody plant genomes have been fully sequenced, gene function research and biotechnological breeding advances have lagged behind. As a result, only a limited number of genes have been elucidated, making it difficult to use newer tools such as CRISPR-Cas9 for biotechnological breeding purposes. The use of *Agrobacterium rhizogenes* as a transformative tool in plant biotechnology has received considerable attention in recent years, particularly in the research field on woody plants. Over the past three decades, numerous woody plants have been effectively transformed using *A. rhizogenes*-mediated techniques. Some of these transformed plants have successfully regenerated. Recent research on *A. rhizogenes*-mediated transformation of woody plants has demonstrated its potential for various applications, including gene function analysis, gene expression profiling, gene interaction studies, and gene regulation analysis. The introduction of the Ri plasmid has resulted in the emergence of several Ri phenotypes, such as compact plant types, which can be exploited for Ri breeding purposes. This review paper presents recent advances in A. rhizogenes-mediated basic research and Ri breeding in woody plants. This study highlights various aspects of A. rhizogenes-mediated transformation, its multiple applications in gene function analysis, and the potential of Ri lines as valuable breeding materials

KEYWORDS

gene function analysis, hairy root, regeneration, Ri plasmid, transformation

1 Introduction

The woodiness proportion among the global vascular plant population ranges from 45% to 48% (FitzJohn et al., 2014). Woody plants play a vital role in meeting human needs by serving as a source of energy, construction materials, and sustenance, while also offering essential ecosystem services such as carbon sequestration, biodiversity support, and climate regulation (Trumbore et al., 2015). Woody plants, encompassing both trees and shrubs, exhibit comparatively lengthier generation times than their herbaceous counterparts, which typically display shorter generation times. Understanding the fundamental biology of woody plants is of utmost importance to enhance their environmental resilience, productivity, and other desirable traits through technological advancements, thus facilitating the cultivation of novel varieties. Among the various methods available, *Agrobacterium*-mediated transformation is the most commonly employed approach. Both *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* possess the ability to infect plant cells and facilitate the transfer of a DNA segment, referred to as T-DNA, which carries oncogenes from the pathogen to the plant cells. Subsequently, this T-DNA integrates into the genome of the host plant. In the case of *A. tumefaciens*-mediated transformation, there are two obstacles for woody plants: *in vitro* (tissue culture-dependent) manipulation and regeneration from explants. Conversely, *A. rhizogenes*-mediated transformation allows for *ex vitro* (tissue culture-independent) manipulation, resulting in the production of composite plants within a shorter time frame (Meng et al., 2019; Cao et al., 2023). *A. rhizogenes* is a Gram-negative bacterium, which contains the root-inducing (Ri) plasmid that induces plants to produce hairy roots from wounds (Moore et al., 1979; White and Nester, 1980a; Chilton et al., 1982; Lothar et al., 1982). Currently, a wide range of plant species, including angiosperms (both dicotyledonous and monocotyledonous plants), gymnosperms, and even moss, have been found to be susceptible to successful infection by *A. rhizogenes* (De Cleene and De Ley, 1981; Porter and Flores, 1991; Spiess et al, 1977; Masako et al., 2004). The T-DNA fragment on the Ri plasmid integrates into the host plant genome, causing hairy root formation. Plants regenerate from hairy roots, while Ri T-DNA transmits through meiosis (Tepfer, 1984). Therefore, *A. rhizogenes* has become a useful tool for plant biotechnology. The principal steps and factors involved in *A. rhizogenes*-mediated plant transformation are similar to those of *A. tumefaciens* and have been comprehensively examined in previous studies (Gelvin, 2009; Gelvin, 2010; Pitzschke, 2013). It can be suggested that due to similarity in the linear organization of genetic loci that perform equivalent functions during T-DNA transfer, Ti and Ri plasmids appear to be very similar in structure and function in regard to mobilization and transfer of T-DNA (White and Nester, 1980b; Risuleo et al., 1982; Hooykaas et al., 1984; Huffman et al., 1984; Jouanin, 1984). This is confirmed by the genomic sequence of *A. rhizogenes* strain LBA9402 (Hooykaas and Hooykaas, 2021). Four *rol* (for root locus) genes, *rolA*, *rolB*, *rolC*, and *rolD*, have been the object of intense study and are mainly responsible for inducing hairy root formation (Mauro et al., 2017). Analyses have been largely performed in dicotyledonous plants such as tobacco, carrot, tomato, and kiwi, transformed with single genes or in combination (Capone et al., 1989; Christophe et al., 1991; Rugini et al., 1991; Van Altvorst et al., 1992). Auxin synthesis genes, such as *aux1* and *aux2*, collaborate with *rol* genes to facilitate the induction of hairy root formation by supplying auxin. However, the absence of the auxin biosynthetic genes *tms1* and *tms2* in mannopine-type Ri plasmids does not affect root induction, suggesting that Ri plasmids may also alter the development of transformed explants through signaling other than auxin synthesis (Hansen et al., 1991). Opine synthesis genes including *ags* and *mas* can synthesize different opines. The classification of *A. rhizogenes* strains is primarily determined by the specific type of opine they produce. These strains can be divided into four groups: agropine (Petit et al., 1983), mannopine (Petit et al., 1983), mikimopine (Akira et al., 1990), and cucumopine (Elisabeth et al., 1988). According to the opine genes present in Ri plasmids, more than 20 strains of *A. rhizogenes* used for genetic transformation have been assigned to these aforementioned categories (Bahramnejad et al., 2019).

Almost three decades after the initial successful transformation of *A. rhizogenes* in grapevine (Guellec et al., 1990), numerous successful transformations have been performed in woody plants, typically in the form of hairy roots or composite plants. In addition, the regeneration of transgenic hairy roots in various plants has further demonstrated the applicability of the *A. rhizogenes* transformation system in molecular plant breeding. Hairy roots cultivated *in vitro* serve as a primary means for the synthesis of secondary metabolites, particularly within medicinal plants (Sharma et al., 2013). Furthermore, the utilization of hairy roots and composite plants in scientific investigations to explore gene functionality is extensive, encompassing various aspects such as root development, wood formation (Plasencia et al., 2016), interactions between roots and soil microbes (Plett et al., 2014), and plant allelopathy (Stanisic et al., 2019). This review aims to delve into the process of transformation of woody plants through the application of *A. rhizogenes*, while also discussing the potential implications of such transformations for basic research and biotechnological advancement.

2 Agrobacterium rhizogenes as a root agent

The propagation of plants through stem cuttings is commonly used in the commercial production of ornamental plants, medicinal plants, and timber trees. However, some commercially important tree species exhibit a low root formation rate. As far back as 1930, the induction of hairy roots was observed in nursery apple trees (Riker et al., 1930), and it was subsequently determined that the root induction was caused by *A. rhizogenes* (Hildebrand, 1934). Since then, *A. rhizogenes* has been used to enhance root formation in plants that are difficult to propagate through stem cuttings (Zavattieri et al., 2016).

The use of *A. rhizogenes* as a rooting inducer for difficult-to-root plants has been found to be remarkably effective and suitable for a

wide range of plants. The determination of the root-promoting ability of *A. rhizogenes* is contingent upon the combination of bacterial strains and plant genotypes. Within woody plant species, there exists variation in rooting efficiency. Additionally, the specific strain of *A. rhizogenes* employed is a critical factor. For instance, in the case of the challenging-to-root woody species, "Golden Delicious" apple, the induction of rooting rates ranges from 0% to 20%, depending on the strain of *A. rhizogenes* utilized. Notably, strains A4 and 232 successfully induce adventitious root formation, whereas strains 178 X A4T and R1000 are unable to do so (Pateña et al., 1988). Cuttings from mature jujube trees (*Ziziphus jujuba* Mill.) exhibit a significant challenge in rooting. However, the application of *A. rhizogenes*, specifically strain TR105, resulted in the highest root formation percentage (65%), which was twice as high as that of the uninoculated cuttings (32.5%). Conversely, strain A4 did not show any significant difference in root formation (Mochammad et al., 1996). Interestingly, in the case of *Corylus avellana*, the rooting rate reached 100% when inoculated with a combination of *A. rhizogenes* (A7 + 22) (Bassil et al., 1991). In addition to bacterial strains, the success of inoculation and production of hairy roots is significantly influenced by plant genotypes and states, as evidenced by multiple studies. For instance, in hazelnut, the stimulation of rooting of cuttings was found to be influenced by both the cultivar and the date of cutting collection (Bassil et al., 1991). Similar results have also been observed in other research groups (Magnussen et al., 1994; Mihaljevic et al., 1996; Sarmast et al., 2019). The infectivity and adventitious root production of *A. rhizogenes* in host plant tissue are contingent upon the compatibility between *A. rhizogenes* and host plants, the responsiveness of the plant tissues to the T-DNA, the production of phytohormones, and the juvenile state of the host tissues.

Auxin is a well-known root inducer, and exogenously applied auxins have been shown to accelerate the rooting process in cuttings of a wide variety of plant species. When auxin and *A. rhizogenes* are combined, they exhibit varying effects on branch rooting, including synergy, antagonism, or no effect. Previous studies have observed a synergistic action between IBA and *A. rhizogenes* in inducing rooting in radiata pine (Li and Leung, 2003) and walnut (Caboni et al., 1996). Conversely, an antagonistic action between IAA and *A. rhizogenes* has been observed in inducing rooting in *Pinus monticola* (McAfee et al., 1993). The combined effect of *A. rhizogenes* to stimulate rooting is contingent upon the specific species and genotypes of micro propagated fruit trees. In all tested genotypes, root formation was observed following infection with *A. rhizogenes*. Three distinct responses were observed: genotypes that rooted without the presence of auxins showed a decrease in rooting percentage when auxin and infection were combined; genotypes that rooted only with auxin exhibited either no effect or a synergistic effect between auxins and infection; genotypes that rooted solely with *A. rhizogenes* displayed either no effect or an antagonistic effect between auxins and infection (Carmine et al., 1998). According to Zarei et al. (2020), the induction of rooting in *Picea abies*, a species known for its reluctance in rooting, cannot be achieved solely through the use of *A. rhizogenes*. However, when *A. rhizogenes* is combined with auxin, successful rooting can be achieved (Zarei et al., 2020).

A histological investigation revealed that the development of auxin-induced roots differs from that of *A. rhizogenes*. In the presence of NAA, adventitious roots are generated endogenously, originating within the vascular tissues of the stem. Conversely, adventitious roots formed in response to *A. rhizogenes* infection exhibit both endogenous and exogenous growth patterns. In the process of endogenous root formation, calli are generated within the cortex, leading to the subsequent formation of tracheid nests, which results in a bulge in the stem. Additionally, exogenous callus, formed at the base of shoot, also gives rise to tracheid nests. Consequently, roots form from both of these callus structures (Ellen and Juvenal, 1993). The rooting process in walnut was found to be influenced by the combined action of IBA and *A. rhizogenes*, as well as the antagonistic effect of IAA and *A. rhizogenes*. Notably, a significant reduction in the roots containing bacteria was observed when *A. rhizogenes* was combined with either IAA or IBA (Falasca et al., 2000).

It is imperative to conduct strain screening to optimize the rooting rate for each specific plant species. The utilization of multiple strains of *A. rhizogenes* offers a captivating approach to enhance rooting. The modulation of plant sensitivity to auxin by *A. rhizogenes* is believed to contribute to the promotion of rooting (Shen et al., 1988; Spanò et al., 1988; Petersen et al., 1989; Christophe et al., 1991). This modulation is thought to occur through variations in endogenous auxin levels and auxin sensitivities across different plant species (Carmine et al., 1998).

3 Agrobacterium rhizogenes as genetic engineers

3.1 Transformation

A. rhizogenes not only facilitates the rooting of difficult-to-propagate plants, but also allows for the integration of foreign genes into the plant genome through binary vectors. Researchers have successfully transferred foreign genes into various woody plant species, including *Larix decidua* (Huang et al., 1991), *Alhagi pseudoalhagi*, *Eucalyptus camaldulensis* (Balasubramanian et al., 2011), *Prunus* (Bosselut et al., 2011; Xu et al., 2020), *Parasponia* (Cao et al., 2012), *Trema* (Cao et al., 2012), *Poncirus trifoliata* (Xiao et al., 2014), *Solanum erianthum* (Sarkar et al., 2020), *Populus* (Neb et al., 2017), *Salix purpurea* (Gomes et al., 2019), *Malus prunifolia* (Yamashita et al., 2004), and *Litchi chinensis* (Qin et al., 2021), by infecting plant organs such as cotyledons, hypocotyls, stem segments, root segments, leaves, petioles, callus, and *in vitro* shoots with *A. rhizogenes* (Table 1). The use of *in vitro* shoots as explants has enabled the development of composite plants, which have numerous biological applications, such as nutrient absorption, biotic and abiotic stress tolerance, and signal exchange between aboveground and underground plant parts. However, the tissue culture of woody plants presents several technical challenges, such as browning and the need for aseptic conditions, which require

TABLE 1 Woody plants transformed by *A. rhizogenes*.

Plant species	Infected explants	*A. rhizogenes* strain(s)	Infected condition	Transformants	Regeneration pathway	References
Actinidia deliciosa	Cuttings	NIAES 1724	*In vitro*	Transgenic plants	Organogenesis*	Yazawa et al., 1995
Aesculus hippocastanum	Androgenic embryos	A4GUS	*In vitro*	Transgenic plants	Organogenesis	Zdravkovic-Korac et al., 2004
Ailanthus altissima, *Aralia elata*, *Clerodendrum chinense*	Seedlings	K599	*Ex vitro*	Transgenic plants	Organogenesis	Cao et al., 2023
Alhagi pseudoalhagi	Cotyledon and hypocotyl	A4	*In vitro*	Transgenic plants	Organogenesis	Wang et al., 2001
Allocasuarina verticilkita	Epicotyl, cotyledon, and hypocotyl	A4, 2659	*In vitro*	Transgenic plants	Organogenesis*	Phelep et al., 1991
Aralia elata	Petiole, roots, leaves	ATCC 15834	*In vitro*	Transgenic plants	Somatic embryogenesis	Kang et al., 2006
Cajanus cajan	Root, hypocotyl, stem, cotyledon, leaves, and petiole	ATCC43057, R1601, LBA9402, A4, ATCC15834	*In vitro*	Hairy root		Jiao et al., 2020
Camellia assamica	Leaves	LBA9402	*In vitro*	Hairy root		Tisserant et al., 2016
Camellia sinensis	Callus	ATCC15834	*In vitro*	Hairy root		Rana et al., 2016
Camellia sinensis	Seedlings	A4	*Ex vitro*	Composite plants		Alagarsamy et al., 2018
Caragana sinica, *Aquilaria sinensis*, *Malus domestica*, *Malus hupehensis*, *Malus pallasiana*	Seedings	K599	*In vitro* and *ex vitro*	Transgenic plants	Organogenesis	Wu et al., 2012
Carica papaya	Leaves	LBA9402	*In vitro*	Transgenic plants	Somatic embryogenesis	Cabrera-Ponce et al., 1996
Carica papaya	Hypocotyl	K599	*Ex vitro*	Composite plants		Hoang et al., 2022
Casuarina glauca	Seedlings	A4RS	*In vitro*	Composite plants		Diouf et al., 1995
Cherry rootstock Colt	Shoots	NCPPB 1855	*In vitro*	Transgenic plants	Somatic embryogenesis	Gutierrez-Pesce et al, 1998
Citrus	Stem cuttings	K599	*Ex vitro*	Composite plants		Ma et al., 2022
Citrus	Seedlings	ATCC 43056	*In vitro*	Transgenic plants	Organogenesis	Ramasamy et al., 2023
Coffea arabica	Roots, hypocotyls, cotyledons	A4RS, Rqua1, 1724, 2659, 8196	*In vitro*	Composite plants		Alpizar et al., 2006
Coffea canephora	Seedlings	A4	*In vitro*	Transgenic plants	Somatic embryogenesis	Kumar et al., 2006
Discaria trinervis	Seedlings, stem cuttings	A4RS, ARqua1	*In vitro* and *ex vitro*	Composite plants		Imanishi et al., 2011
Dryas drummondii, *Dryas octopetala*	Seedlings	AR1193	*In vitro*	Composite plants		Billault-Penneteau et al., 2019

(Continued)

10.3389/fpls.2023.1196561

TABLE 1 Continued

Plant species	Infected explants	*A. rhizogenes* strain(s)	Infected condition	Transformants	Regeneration pathway	References
Duboisia myoporoides x *D. leichhardtii*	Leaves	A4	*In vitro*	Transgenic plants	Organogenesis*	Roig Celma et al., 2001
Eucalyptus camaldulensis	Seedlings	A4RS	*In vitro*	Composite plants		Bosselut et al., 2011
Eucalyptus grandis	Seedlings	A4RS	*In vitro*	Hairy root and composite plants		Plasencia et al., 2016
Ginkgo biloba	Zygotic embryos	A4	*In vitro*	Hairy root		Ayadi and Trémouillaux-Guiller, 2003
Justicia gendarussa	Leaf petiole	A4 and MTCC 532	*In vitro*	Hairy root		Largia et al., 2022
Larix decidua	Hypocotyl	11325	*In vitro*	Transgenic plants	Organogenesis	Huang et al., 1991
Litchi chinensis	Leaves and stem segments	MSU440	*In vitro*	Hairy root		Qin et al., 2021
Malus baccata	Shoots	8196	*In vitro*	Transgenic plants	Organogenesis	Wu et al., 2012
Malus pumila	Shoots	A4	*In vitro*	Transgenic plants	Organogenesis	Lambert and Tepfer, 1992
Malus pumila	Shoots	8196, A4, 15834	*In vitro*	Transgenic plants	Organogenesis	Pawlicki-Jullian et al., 2002
Malus pumila	Stem	MAFF 02-10266, 03-01724, 03-01725	*In vitro*	Transgenic plants	Organogenesis	Yamashita et al., 2004
Morus indica	Seedlings	MAFF 210268 and 720001, MAFF 210265	*In vitro*			Oka and Tewary, 2000
Parasponia andersonii, *Trema tomentosa*	Shoots	MSU440	*In vitro*	Composite plants		Cao et al., 2012
Persea americana	Seedlings	K599 or ARqua1	*Ex vitro*	Composite plants		Prabhu et al., 2017
Pinus contorta	Seedlings	LBA 9402, A4RSII	*In vitro*	Composite plants		Yibrah et al., 1996
Poncirus trifoliata	Leaves, epicotyls	MSU440, K599	*In vitro*	Transgenic plants	Organogenesis	Xiao et al., 2014
poplars	Shoot cuttings	1724, K599, 8196, 15834	*In vitro*	Composite plants		Neb et al., 2017
Prunus persica	Cuttings	A4R	*In vitro*	Composite plants		Bosselut et al., 2011
Prunus persica	Hypocotyl, leaves, seedlings	MSU440	*In vitro*	Hairy root and composite plants		Xu et al., 2020
Punica granatum	radicle, cotyledon, leaves	MSU440, A4, 15834	*In vitro*	Hairy root		Ono et al., 2012
Rauwolfia serpentina	Leaves	A4	*In vitro*	Transgenic plants	Organogenesis*	Mehrotra et al., 2013
Robinia pseudoacacia	Hypocotyls	RI601	*In vitro*	Transgenic plants	Organogenesis	Han et al., 1993

(Continued)

10.3389/fpls.2023.1196561

TABLE 1 Continued

Plant species	Infected explants	*A. rhizogenes* strain(s)	Infected condition	Transformants	Regeneration pathway	References
Salix purpurea and *Salix* spp.	Shoots	A4RS	*In vitro*	Hairy root and composite plants		Gomes et al., 2019
Solanum erianthum	Leaves	A4	*In vitro*	Hairy root		Sarkar et al., 2020
Taxus baccata	Seedlings, shoots	A4	*In vitro* and *ex vitro*	Composite plants		He et al., 2022

*Automatically organogenesis from hairy root.

skilled operators and the identification of suitable bacteriostatic agents for different strains and explants. To address these challenges, Collier et al (2005) developed a novel method to induce *A. rhizogenes* infection in plants without the use of tissue culture, thereby generating composite plants (Collier et al., 2005). This approach not only is cost-effective and efficient, but also enables the production of composite plants in a short time frame without the need for tissue culture. Furthermore, it is simple to execute, boasts a short cycle, and does not require complex sterilization procedures. Collier and colleagues successfully applied this technique to 14 dicotyledonous herbs from nine genera spanning four families. Following infection with the same strain, 14 composite dicotyledonous plants were generated, with transformation efficiencies ranging from 56% to 100% (Collier et al., 2005). This method has also been successfully implemented in woody plants. Using seedlings as explants, composite plants have been generated from a wide range of woody plants, including *Camellia sinensis* (Alagarsamy et al., 2018), *Discaria trinervis* (Imanishi et al., 2011), *Persea americana* (Prabhu et al., 2017), *Taxus baccata* (He et al., 2022), *Carica papaya* (Hoang et al., 2022), *Citrus* (Ma et al., 2022), *Ailanthus altissima* (Cao et al., 2023), *Aralia elata* (Cao et al., 2023), *Clerodendrum chinense* (Cao et al., 2023), *Caragana sinica* (Diouf et al., 1995), and *Malus pumila* (Pawlicki-Jullian et al., 2002; Yamashita et al., 2004) (Table 1). Although seedlings can be used as explants for woody plants, the genetic heterozygosity of these plants requires the use of a sufficient number of transformed lines to ensure the accuracy of experimental results, particularly in studies of stress resistance and other biological phenomena. In addition, stem cutting represents a crucial means of reproducing woody plants. In this regard, Ma et al (2022) successfully generated composite plants using citrus stem cuttings as explants, resulting in composite plants that possess the same genetic background, and are therefore more suitable for biological research applications (Ma et al., 2022).

3.2 Impact factors

The core factors that impact the transformation efficiency of *A. rhizogenes* are the same as those that influence root growth promotion when *A. rhizogenes* is employed as a rooting agent. These factors include the type of *A. rhizogenes* strains and the genetic or genotype or states of the host plants. For instance, in the context of poplar stem segment transformation, the four strains (1724, K599, 8196, and 15834, representing mikimopine, cucumopine, mannopine, and agropine strains, respectively) exhibit varying differences in hairy root formation time, number of formations, and transformation efficiency. Hence, it is essential to identify suitable strains for specific plants (Neb et al., 2017). Moreover, variations can also arise between species, varieties, and clones. For instance, a study involving five woody plants found that *Aquilaria sinensis* failed to produce hairy roots, while the remaining five plants generated composite plants with efficiencies ranging from 30% to 85% (Meng et al., 2019). In 22 citrus species, transformation efficiencies fluctuated from 0 to 95% (Ma et al., 2022).

Other factors, such as the tissue and physiological conditions of explants, bacterial concentration, acetosyringone concentration, hormone recipes, and additional treatments such as vacuuming will affect transformation efficiency. Leaves, petioles, stem segments, and adventitious buds can be used as explants. Suitable explants vary from plant to plant. For example, in *Cajanus cajan*, leaves have the highest transformation efficiency, reaching 70.92% (Jiao et al., 2020). In *L. chinensis*, there is no significant difference between the transformation efficiency of stem segments and leaves (Qin et al., 2021). In *Coffea arabica*, hypocotyls of cultivars of Caturra and IAPAR-59 had the highest infection efficiencies, reaching 82% and 51%, respectively (Alpizar et al., 2006). The age of the plants also affects the transformation efficiency; the transformation efficiency of 8-week-old seedlings was significantly higher than that of 20-week-old *P. americana* (Prabhu et al., 2017). After *A. rhizogenes* infection, hairy roots are often produced through cortical cells (Falasca et al., 2000). The older the physiological age, the higher the degree of lignification, the lower the proportion of cortical cells, and the longer the time used for hairy root induction. In addition, the concentration of the bacterial solution also affects transformation efficiency. In general, the concentration of *A. rhizogenes* infecting plants is between 0.4 and 1.0 (Ma et al., 2022; Cao et al., 2023). At the same time, full induction of the *vir* gene in the strain requires acetosyringone, a chemoattractant for *Agrobacterium*, at a commonly used concentration of 100–200 μM (Ma et al., 2022; Cao et al., 2023). To improve the transformation efficiency, additional treatments such as vacuum treatment have been applied to facilitate the infiltration *A. rhizogenes* into plant cells (Ma et al., 2022).

3.3 Regeneration

Hairy roots induced by A. rhizogenes have the potential to facilitate the generation of transgenic plants through various ways, including spontaneous organ regeneration, organogenesis, or somatic embryogenesis (Figure 1). Nevertheless, many woody species were induced by regeneration plants via organogenesis or somatic embryogenesis. Some woody plants, such as *Actinidia deliciosa* (Yazawa et al., 1995), *Allocasuarina verticillate* (Phelep et al., 1991), *Duboisia myoporoides x D. leichhardtii* (Trovato et al., 2001), and *Rauwolfia serpentina* (Mehrotra et al., 2013), are capable of regeneration via spontaneous organ regeneration from their hairy roots, which can be completed in hormone-free medium. On the other hand, other woody plants such as *Aesculus hippocastanum* (Zdravkovic-Korac et al., 2004), *A. pseudoalhagi* (Wang et al., 2001), *Larix decidua* (Huang et al., 1991), *M. pumila*

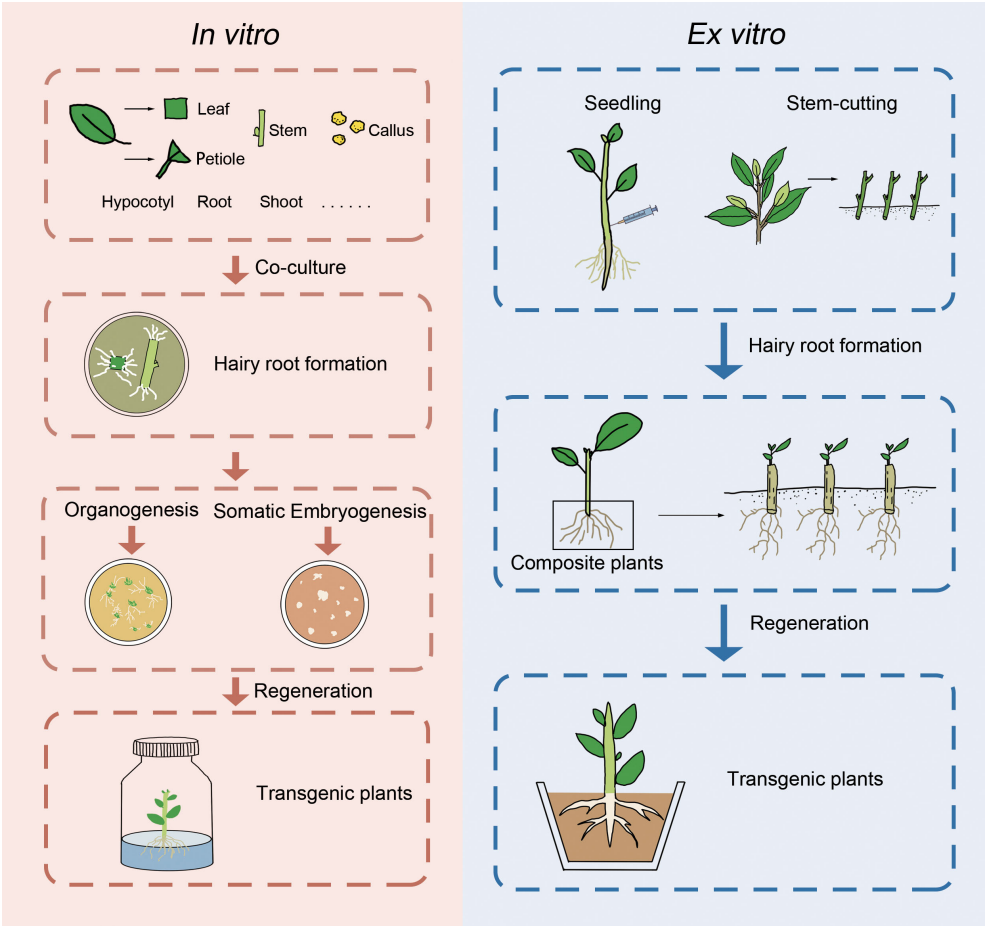


FIGURE 1
Presentation of *in vitro* and *ex vitro* genetic transformation of woody plants mediated by *A. rhizogenes*.

(Pawlicki-Jullian et al., 2002), *P. trifoliata* (Xiao et al., 2014), *Robinia pseudoacacia* (Han et al., 1993), and citrus (Ramasamy et al., 2023) require hormone ratios to induce the production of calli, from which regenerated plants can be obtained. It is worth noting that in the study of *Malus baccata*, regenerated plants were only obtained when the hairy roots remained attached to the mother plant (non-transformed aerial part) (Wu et al., 2012). In addition to organogenesis, some woody plants can also produce somatic embryos through the induction of hairy roots, which germinate into regenerated plants. Examples of such plants include *A. elata* (Kang et al., 2006), *C. papaya* (Cabrera-Ponce et al., 1996), cherry rootstock *Colt* (Gutierrez-Pesce et al., 1998), and *Coffea canephora* (Kang et al., 2006). However, despite multiple attempts, some studies have failed to obtain regenerated plants through hairy roots. Recently, a cut-dip-budding method has been successfully applied to three woody plants, namely, *A. altissima*, *A. elata*, and *C. chinense*, which allowed the generation of *A. rhizogenes*-mediated transgenic plants (Cao et al., 2023) (Figure 1). The regeneration of woody plants from roots is a common phenomenon and can be enhanced by pruning aboveground parts (Wan et al., 2006). These approaches would facilitate woody plant transformation mediated by *A. rhizogenes*.

4 Applications in woody plants

4.1 Basic biological research

Currently, *A. tumefaciens*-mediated genetic transformation systems have been successfully implemented in several woody species, such as poplar (Wang et al., 2011), apple (Schropfer et al., 2022), kiwifruit (Uematsu et al., 1991), and walnut (McGranahan et al., 1988). However, the number of woody plants that have been sequenced far exceeds the number of plants that have established transformation systems. With advancements in sequencing technology and the enhanced level of genome assembly, there is a growing potential for the sequencing of numerous genetically intricate woody plant genomes. An example of this progress is the successful completion of the Chinese pine (*Pinus tabuliformis*) genome, which has been assembled at the chromosome level and spans a size of 25.4 gigabases (Gb) (Niu et al., 2022). Consequently, effective use of the substantial amount of sequence information will become an imperative endeavor. Transgenic technology for resolving gene function is emerging as a key tool to address this challenge. In recent years, the utilization of *A. rhizogenes* for the production of composite plants has gained significant traction in the realm of herbaceous plant investigation, particularly in the case of soybean. This development has provided a promising avenue for exploring woody plant research.

The application of *A. rhizogenes*-mediated transformation in soybean has proven to be highly successful in various biological contexts. A particularly efficient and rapid method for this transformation has been established (Kereszt et al., 2007). In soybean research, three main types of applications have been extensively utilized. These include the construction of composite plants through overexpression, RNAi, and CRISPR-Cas9 binary vectors for the purpose of gene function analysis (Traubenik et al., 2020; Li et al., 2021). Additionally, basic molecular analyses, such as promoter analysis (Yang et al., 2021), subcellular localization studies (Brear et al., 2020), ChIP-PCR (Pi et al., 2019), GST pull-down assays (Wang et al., 2015), Co-IP assays (Vadivel et al., 2021), protein ubiquitination and degradation studies (Zhang et al., 2021), and *in vivo* kinase assays (Gao et al., 2022), have been conducted. Of particular interest is mutant complementation (Feng et al., 2021; Jiang et al., 2021), which serves as a genetic validation of gene function.

A variety of studies have used *A. rhizogenes*-mediated transformation technology for basic biological investigations in woody plants, including binary vector generation of composite plants, basic molecular analysis, and genetic analysis. *A. rhizogenes* possesses the ability to carry binary vectors, which facilitates gene overexpression, RNAi, and gene editing. These transformants have been utilized to examine nutrient uptake, abiotic stresses, nodule development, mycorrhizal interactions, allelopathy, biosynthesis, and wood formation. For instance, in apple, the function of the *MdPRP6* gene was elucidated under low nitrogen conditions through *A. rhizogenes*-mediated knockout or overexpression (Zhang et al., 2022). In *Eucalyptus*, the STOP-like gene was disrupted to assess its aluminum resistance function (Sawaki et al., 2014). In poplar, *A. rhizogenes*-mediated transformation yielded *PtJAZ6* knockout and overexpression materials, which were subsequently employed to analyze the mutualistic interaction between *PtJAZ6* and effectors in ectomycorrhizal *Laccaria bicolor* (Plett et al., 2014). This technology has also been applied to *R. pseudoacacia* to identify *Rpf41* as a critical regulator of symbiotic nodulation in legumes (Chou et al., 2016). In apple, the function of *BvSTI* in allelopathy was examined using the hairy root system (Stanisic et al., 2019). Although genetic transformation of tea plants poses challenges, the function of *CsTSI* in theanine biosynthesis was successfully investigated using the *A. rhizogenes* system (She et al., 2022). Furthermore, it is worth noting that woody plant roots, similar to stems, undergo secondary growth, making them a valuable resource for investigating the process of wood formation (Plasencia et al., 2016).

Gene function studies involve analyzing tissue expression, protein interactions, and downstream target genes. The use of hairy roots enables researchers to conduct related research within homologous species, thereby providing a more accurate reflection of gene function. For instance, in citrus, the *CsSUC2* promoter was analyzed in hairy roots (Roig Celma et al., 2001), while in Eucalyptus, the *EgCCR1* and *EgCAD2* promoters were studied in hairy roots (Rugini et al., 1991). Bimolecular fluorescence complementation (BiFC) was used to verify the interaction between *CcCIPK14* and *CcCBL1* in hairy roots of cowpea (Mehrotra et al., 2013). Furthermore, hairy roots can be utilized to screen downstream target genes of transcription factors. For instance, downstream target genes of *MYB15* were identified in grape (White et al., 1985). In poplar, the *A. rhizogenes* system can be used to verify the downstream target genes of transcription factors through ChIP-PCR (Schmülling et al., 1988; Filippini et al., 1996). It is worth noting that genetic manipulation of woody plants, especially the genetic analysis of upstream and downstream genes, is challenging. The use of the *A. rhizogenes* system has overcome this challenge in some species. For example, Ma (2018) used *A.*

rhizogenes to knock out *MdSUT2.2* based on *MdCIPK22* transgenic apple plants, which confirmed that *MdCIPK22* depends on *MdSUT2.2* for drought tolerance (Zhu et al., 2003).

4.2 Ri breeding

Plants regenerated from hairy roots exhibit phenotypes, including vigorous large root growth, lateral root development, root geotropism loss, loss of shoot apical dominance, internode shortening, and plant dwarfing (Han et al., 1993; Yazawa et al., 1995; Yamashita et al., 2004; Zdravkovic-Korac et al., 2004; Wu et al., 2012; Rugini et al., 2015). These characteristics have been observed in the regenerated plants of numerous species, such as *R. pseudoacacia* (Han et al., 1993), *A. deliciosa* (Rugini et al., 1991), papaya (Cabrera-Ponce et al., 1996), apple rootstock Jork 9 (Pawlicki-Jullian et al., 2002; Yamashita et al., 2004; Wu et al., 2012), and sweet cherry (Rugini et al., 2015). While the majority of research has concentrated on the initial characteristics of regenerated plants, Mehrotra et al (2013) conducted an investigation into the flowering traits of *R. serpentina*, a shrub characterized by a brief growth cycle (Mehrotra et al., 2013). Their findings revealed that transgenic plants, when compared to nontransformed plants, exhibited normal flowering patterns albeit with a reduced quantity of inflorescences and flowers. Another study conducted in cherries by Rugini et al (2015) tracked *A. rhizogenes*-transformed cherries for 10 years and observed a slight decrease in the number of flowers, while flower morphology, ovule differentiation, and flowering time remained unaltered. The fruit traits of transgenic materials did not change significantly, with the exception of reduced fruit yield due to the decreased number of flowers. The transgenic plants were grafted as rootstocks, reducing the size of the two plants to varying degrees, but fruit quality remained unchanged (Mehrotra et al., 2013). It should be noted that considerable differences in traits exist among the obtained transgenic plants due to strains of *A. rhizogenes* used in transformation and different copy numbers or insertion positions of T-DNA following transformation. Thus, a sufficient number of transgenic plants is required for stable, long-term observation.

The hairy root phenotype of transgenic plants is primarily regulated by *rol* genes. In 1985, White et al. identified and analyzed four *rol* genes, specifically *rolA*, *rolB*, *rolC*, and *rolD*, which play pivotal roles in neoplastic disease induction and hairy root formation (White et al., 1985). Transgenic plants obtained through transformation by *A. rhizogenes* exhibited phenotypes analogous to plants transformed by single or multiple *rol* genes. The *rolB* gene possesses tyrosine phosphatase activity and plays an essential role in root initiation and elongation, contributing significantly to the hairy root phenotype (Filippini et al., 1996). Plants transformed with *rolB* demonstrated typical *A. rhizogenes*-induced phenotypes including increased adventitious roots, loss of apical dominance, and shortened internodes. Examples of such transformed plants include *Pyrus communis*, *Kalanchoe diagremontiana*, and grape rootstocks "Richter 110" (Schmülling et al., 1988; Zhu et al., 2003; Geier et al., 2008). The *rolB* gene specifically upregulates *ARF7* and *ARF19* to promote root initiation in *Nicotiana tabacum* (Bose et al., 2022). The *rolC* is a glucosidase, and its expression correlates with enhanced cytokinin activity in tobacco (Estruch et al., 1991). Plants transformed with *rolC* exhibited dwarfing, dark green leaves, and increased branching, which were associated with reduced gibberellin levels. Furthermore, the introduction of *rolC* into plants such as *P. trifoliata* and *Diospyros kaki* displayed stronger rooting ability (Kaneyoshi and Kobayashi, 1999; Koshita et al., 2002). Similar to *rolB* and *rolC* genes, the *rolA* gene has the ability to induce leaf rooting in *K. diagremontiana* without the need for exogenous hormones (Schmülling et al., 1988). Plants transformed with *rolA* exhibited plant dwarfing and leaf shrinkage, which were associated with decreased gibberellin levels in tobacco (Schmülling et al., 1993). The *rolA* gene encodes a DNA-binding protein akin to the HPV-1 E2 DNA-binding protein (Rigden and Carneiro, 1999). The *rolD* gene promotes plant flowering, increases axillary inflorescence formation and elongation, and facilitates adventitious root formation in both tobacco and *Arabidopsis* (Mauro et al., 1996; Falasca et al., 2010). The *rolD* gene encodes ornithine cyclodeaminase, an enzyme that catalyzes the conversion of ornithine to proline (Trovato et al., 2001).

One notable characteristic of Ri plants is their compact plant structure, distinguished by shortened internodes, reduced plant height, increased branching, and enhanced axillary bud growth. These traits hold significant value in breeding of flowers and fruit trees. The Ri phenotype is heritable, with the inserted T-DNA being transmitted through meiosis (Tepfer, 1984) and inherited in a Mendelian dominant manner (Zhan et al., 1988; Christensen et al., 2019). Ri lines of *K. blossfeldiana*, characterized by increased branches, weakened apical dominance, and shortened internodes following backcross separation, have been successfully applied in commercial plant breeding (Christensen et al., 2019). It is widely recognized that rootstocks profoundly influence scion size. In fruit tree breeding, employing Ri plants as rootstocks is a good way to obtain dwarfed plants without compromising fruit quality (Rugini et al., 2015).

Another notable feature of Ri plants is the alteration of root morphology. Upon infection of host plants with *A. rhizogenes*, T-DNA insertion induces cell dedifferentiation, leading to the formation of hairy roots. Hairy roots exhibit characteristics such as rapid growth, a high degree of branching, and oblique development. Ri plants have demonstrated enhanced rooting ability under both greenhouse and field conditions in several plant species (Lambert and Tepfer, 1992; Pawlicki-Jullian et al., 2002; Casanova et al., 2005). This increased rooting ability offers several advantages. Firstly, it can lead to efficient asexual reproduction and improved adaptation to *in vitro* conditions. Secondly, Ri plants with enhanced root development hold significant promise in sustainable plant agriculture, as this can improve water and nutrient management, which can augment plant

TABLE 2 The representative naturally occurring (wild type) strains of *A. rhizogenes* and their plasmids.

Type of opine	Strain	Plasmid	References
Agropine	A4	pArA4a pArA4b = pRiA4 pArA4c	White and Nester, 1980b; Savka et al., 1990
	ATCC 15834	pAr15834a pAr15834b = pRi15834 pAr15834c	White and Nester, 1980b; Savka et al., 1990
	LBA 9402 (NCPPB 1855)	pRi1855	Savka et al., 1990
Mannopine	LBA 9365 = strain 8196	pAr8196a pAr8196b = pRi8196 pAr8196c	Chilton et al., 1982; Savka et al., 1990
Cucumopine	Strain 2659 = K599	pRi2659	Savka et al., 1990
Mikimopine	NIAES 1724	pRi1724	Akira et al., 1990

drought tolerance (Tepfer, 2017). Additionally, increased root biomass particularly benefits plants such that their roots were used for extracting specific metabolites (Mehrotra et al., 2013).

5 Conclusion and future perspectives

Over the past three decades, significant progress has been made in the transformation of woody plants by *A. rhizogenes*. Numerous woody plants have been transformed with *A. rhizogenes*, and regenerated plants can be obtained *in vitro* through hairy roots. The direct use of lignified hairy roots to acquire regenerated plants *ex vitro* offers broad prospects for the application of *A. rhizogenes* in woody plants. In addition, the incorporation of genome sequencing will significantly advance theoretical investigations in the field of woody plants. The utilization of diverse binary vectors with *A. rhizogenes* serves as a genetic tool for conducting *in vivo* investigations on gene functionality in woody plants (Bahramnejad et al., 2019).

The typical hairy root symptoms induced by *A. rhizogenes*, including compact plant types and enhanced rooting ability, have been used in flower breeding and rootstock modifications. Genome sequence analysis has revealed that *A. rhizogenes* undergoes horizontal gene transfer to plants, such as *Nicotiana*, *Linaria*, and *Ipomoea* species (White et al., 1983; Matveeva et al., 2012; Kyndt et al., 2015; Quispe-Huamanquispe et al., 2017). This implies that *A. rhizogenes*, during the process of pathogenesis, introduces multiple plasmid-encoded genes into its host through horizontal gene transfer, which occurs as a natural outcome. Currently, plants transformed with wild-type *Agrobacterium* strains are not classified as transgenic species (Lutken et al., 2012), which is advantageous for the application of Ri plants. The predominant natural *A. rhizogenes* strains used are listed in Table 2.

However, it is imperative to recognize the limitations associated with these applications. To enhance and exploit natural engineering capabilities, one promising avenue lies in the use of CRISPR-mediated base editing, which serves as a catalyst for "engineering the engineer" (Rodrigues et al., 2021). This approach holds promise for improving *A. rhizogenes*, thereby enabling more effective plant transformation and genome editing. According to the National Center for Biotechnology Information (NCBI) database, 96 strains of *A. rhizogenes* exist (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=359) as of the search conducted on 26 September 2023. However, only four strains (LBA9402, A4, K599, and CA75/95) have complete genome sequences. The transformation capabilities of these strains vary, necessitating the sequencing and engineering of strain genomes to improve infectivity.

Author contributions

Conceptualization, XW and BZ. Data curation, WY and GW. Writing—original draft preparation, WY, GW, XW, and BZ. Writing—review and editing, WX, HL, YH, HY, DY, FC, and JH. Supervision, XW and BZ. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Key Scientific and Technological Grant of Zhejiang for Breeding New Agricultural Varieties (2021C02066-12), the Opening Project of Zhejiang Provincial Key Laboratory of Forest Aromatic Plants-based Healthcare Functions (2022E10008), the State Key Laboratory of Subtropical Silviculture (KF202007), the General Foundation of Education of Zhejiang (Y202250159 and Y202147311), and the Ningbo Natural Science Foundation (202003N4016).

Conflict of interest

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

Publisher's note

References

Akira, I., Naoyuki, F., Miki, H., Hiroshi, K., Hiroshi, H., and Akinori, S. (1990). Mikimopine, an opine in hairy roots of tobacco induced by Agrobacterium rhizogenes. *Phytochemistry* 29 (10), 3131–3134. doi: 10.1016/0031-9422(90)80171-c

Alagarsamy, K., Shamala, L. F., and Wei, S. (2018). Protocol: high-efficiency in-planta *Agrobacterium*-mediated transgenic hairy root induction of *Camellia sinensis* var. *sinensis*. *Plant Methods* 14, 17. doi: 10.1186/s13007-018-0285-8

Alpizar, E., Dechamp, E., Espeout, S., Royer, M., Lecouls, A. C., Nicole, M., et al. (2006). Efficient production of *Agrobacterium rhizogenes*-transformed roots and composite plants for studying gene expression in coffee roots. *Plant Cell Rep.* 25 (9), 959–967. doi: 10.1007/s00299-006-0159-9

Ayadi, R., and Trémouillaux-Guiller, J. (2003). Root formation from transgenic calli of *Ginkgo biloba*. *Tree Physiol.* 23, 713–718. doi: 10.1093/treephys/23.10.713

Bahramnejad, B., Naji, M., Bose, R., and Jha, S. (2019). A critical review on use of *Agrobacterium rhizogenes* and their associated binary vectors for plant transformation. *Biotechnol. Adv.* 37 (7), 107405. doi: 10.1016/j.bioteChadv.2019.06.004

Balasubramanian, A., Venkatachalam, R., Selvakesavan, K. R., Mary, A. S., Gherbi, H., Svistoonoff, S., et al. (2011). Optimisation of methods for *Agrobacterium rhizogenes* mediated generation of composite plants in *Eucalyptus camaldulensis*. *BMC Proc.* 5 (7), 1–3. doi: 10.1186/1753-6561-5-s7-o45

Bassil, N. V., Proebsting, W. M., Moore, L. W., and Lightfoot, D. A. (1991). Propagation of hazelnut stem cuttings using *Agrobacterium rhizogenes*. *HortScience* 26 (8), 1058–1060. doi: 10.21273/hortsci.26.8.1058

Billault-Penneteau, B., Sandré, A., Folgmann, J., Parniske, M., and Pawlowski, K. (2019). *Dryas* as a model for studying the root symbioses of the rosaceae. *Front. Plant Sci.* 10, 661. doi: 10.3389/fpls.2019.00661

Bose, R., Sengupta, M., Basu, D., and Jha, S. (2022). The *rol*B-transgenic *Nicotiana tabacum* plants exhibit upregulated *ARF7* and *ARF19* gene expression. *Plant Direct* 6 (6), e414. doi: 10.1002/pld3.414

Bosselut, N., Van Ghelder, C., Claverie, M., Voisin, R., Onesto, J. P., Rosso, M. N., et al. (2011). *Agrobacterium rhizogenes*-mediated transformation of *Prunus* as an alternative for gene functional analysis in hairy-roots and composite plants. *Plant Cell Rep.* 30 (7), 1313–1326. doi: 10.1007/s00299-011-1043-9

Brear, E. M., Bedon, F., Gavrin, A., Kryvoruchko, I. S., Torres-Jerez, I., Udvardi, M. K., et al. (2020). GmVTL1a is an iron transporter on the symbiosome membrane of soybean with an important role in nitrogen fixation. *New Phytol.* 228, 667–681. doi: 10.1111/nph.16734

Caboni, E., Lauri, P., Tonelli, M., Falasca, G., and Damiano, C. (1996).). Root induction by *Agrobacterium rhizogenes* in walnut. *Plant Sci.* 118, 203–208. doi: 10.1016/0168-9452(96)04449-4

Cabrera-Ponce, J. L., Vegas-Garcia, A., and Herrera-Estrella, L. (1996). Regeneration of transgenic papaya plants via somatic embryogenesis induced by *Agrobacterium rhizogenes*. *In Vitro Cell. Dev. Biol.-Plant* 32, 86–90. doi: 10.1007/bf02823136

Cao, Q., Op Den Camp, R., Seifi Kalhor, M., Bisseling, T., and Geurts, R. (2012). Efficiency of *Agrobacterium rhizogenes*-mediated root transformation of *Parasponia* and *Trema* is temperature dependent. *Plant Growth Regul.* 68 (3), 459–465. doi: 10.1007/s10725-012-9734-y

Cao, X., Xie, H., Song, M., Lu, J., Ma, P., Huang, B., et al. (2023). Cut-dip-budding delivery system enables genetic modifications in plants without tissue culture. *Innovation (Camb)* 4 (1), 100345. doi: 10.1016/j.xinn.2022.100345

Capone, I., Spanò, L., Cardarelli, M., Bellincampi, D., Petit, A., and Costantino, P. (1989). Induction and growth properties of carrot roots with different complements of *Agrobacterium rhizogenes* T-DNA. *Plant Mol. Biol.* 13, 43–52. doi: 10.1007/bf00027334

Carmine, D., Simona, M., and Monticelli, S. (1998). *In vitro* fruit trees rooting by Agrobacterium rhizogenes wild type infection. *Electron. J. Biotechn.* 1 (3), 12–13. doi: 10.4067/s0717-34581998000300008

Casanova, E., Trillas, M. I., Moysset, L., and Vainstein, A. (2005). Influence of *rol* genes in floriculture. *Biotechnol. Adv.* 23 (1), 3–39. doi: 10.1016/j.bioteChadv.2004.06.002

Chilton, M.-D., Tepfer, D. A., Petit, A., David, C., Casse-Delbart, F., and Tempé, J. (1982). *Agrobacterium rhizogenes* inserts T-DNA into the genomes of the host plant root cells. *Nature* 295, 432–434. doi: 10.1038/295432a0

Chou, M., Xia, C., Feng, Z., Sun, Y., Zhang, D., Zhang, M., et al. (2016). A translationally controlled tumor protein gene Rpf41 is required for the nodulation of *Robinia pseudoacacia*. *Plant Mol. Biol.* 90 (4), 389–402. doi: 10.1007/s11103-015-0424-9

Christensen, E. M., Müller, R. P., Lütken, H. V., Hegelund, J. N., Jensen, L., Christensen, B., et al. (2019). *Agrobacterium rhizogenes transformation and expression of rol genes in Kalanchoë. U. S. Patent No.9,253,952* (Washington, DC: U.S. Patent and Trademark Ofiice).

Christophe, M., Hélène, B.-B., Angelo, S., Jacques, T., and Jean, G. (1991). Single *rol* genes from the *Agrobacterium rhizogenes* TL-DNA alter some of the cellular responses to auxin in *Nicotiana tabacum*. *Plant Physiol.* 97 (1), 212–216. doi: 10.1104/pp.97.1.212

Collier, R., Fuchs, B., Walter, N., Kevin Lutke, W., and Taylor, C. (2005). *Ex vitro* composite plants: an inexpensive, rapid method for root biology. *Plant J.* 43 (3), 449–457. doi: 10.1111/j.1365-313X.2005.02454.x

De Cleene, M., and De Ley, J. (1981). The host range of infectious hairy-root. *Bot. Rev.* 47, 147–194. doi: 10.1007/bf02868853

Diouf, D., Gherbi, H., Prin, Y., Franche, C., Duhoux, E., Bogusz, D., et al. (1995). Hairy root nodulation of *Casuarina glauca*: a system for the study of symbiotic gene expression in an actinorhizal tree. *Plant Microbe Interact.* 8 (4), 532–537. doi: 10.1094/mpmi-8-0532

Elisabeth, D., Annik, P., Max, E. T., Max, E. T., Maarten, H. R., Maarten, H. R., et al. (1988). Cucumopine—a new T-DNA-encoded opine in hairy root and crown gall. *Phytochemistry* 27, 2429–2433. doi: 10.1016/0031-9422(88)87007-9

Ellen, G. S., and Juvenal, G. L. (1993). Development anatomy of roots induced by *Agrobacterium rhizogenes* in *Malus pumila* 'M.26' shoots grown in *vitro*. *Int. J. Plant Sci.* 154, 59–67. doi: 10.1086/297090

Estruch, J. J., Chriqui, D., Grossmann, K., Schell, J., and Spena, A. (1991). The plant oncogene *rolC* is responsible for the release of cytokinins from glucoside conjugates. *EMBO J.* 10 (10), 2889–2895. doi: 10.1002/j.1460-2075.1991.tb07838.x

Falasca, G., Altamura, M. M., D'angeli, S., Zaghi, D., Costantino, P., and Mauro, M. L. (2010). The *rolD* oncogene promotes axillary bud and adventitious root meristems in *Arabidopsis*. *Plant Physiol. Biochem.* 48 (9), 797–804. doi: 10.1016/j.plaphy.2010.06.002

Falasca, G., Reverberi, M., Lauri, P., Caboni, E., De Stradis, A., Altamura, M. M., et al. (2000). How Agrobacterium rhizogenes triggers *de novo* root formation in a recalcitrant woody plant: an integrated histological, ultrastructural and molecular analysis. *New Phytol.* 145 (1), 77–93. doi: 10.1046/j.1469-8137.2000.00558.x

Feng, J., Lee, T., Schiessl, K., and Oldroyd, G. E. D. (2021). Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules. *Science* 374 (6567), 629–632. doi: 10.1126/science.abg2804

Filippini, F., Rossi, V., Marin, O., Trovato, M., Costantino, P., Downey, P. M., et al. (1996). A plant oncogene as a phosphatase. *Nature* 379 (6565), 499–500. doi: 10.1038/379499a0

FitzJohn, R. G., Pennell, M. W., Zanne, A. E., Stevens, P. F., Tank, D. C., and Cornwell, W. K. (2014). How much of the world is woody? *J. Ecol.* 102 (5), 1266–1272. doi: 10.1111/1365-2745.12260

Gao, H., Jiang, L., Du, B., Ning, B., Ding, X., Zhang, C., et al. (2022). GmMKK4-activated GmMPK6 stimulates GmERF113 to trigger resistance to *Phytophthora sojae* in soybean. *Plant J.* 111, 473–495. doi: 10.1111/tpj.15809

Geier, T., Eimert, K., Scherer, R., and Nickel, C. (2008). Production and rooting behaviour of *rol*B-transgenic plants of grape rootstock 'Richter 110' (*Vitis berlandieri*×*V. rupestris*). *Plant Cell Tiss. Org.* 94, 269–280. doi: 10.1007/s11240-008-9352-6

Gelvin, S. B. (2009). *Agrobacterium* in the genomics age. *Plant Physiol.* 150, 1665–1676. doi: 10.1104/pp.109.139873

Gelvin, S. B. (2010). Finding a way to the nucleus. *Curr. Opin. Microbiol.* 13, 53–58. doi: 10.1016/j.mib.2009.11.003

Gomes, C., Dupas, A., Pagano, A., Grima-Pettenati, J., and Paiva, J. (2019). Hairy root transformation: a useful tool to explore gene function and expression in Salix spp. recalcitrant to transformation. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01427

Guellec, V., David, C., Branchard, M., and Tempé, J. (1990). *Agrobacterium rhizogenes* mediated transformation of grapevine (*Vitis vinifera* L.). *Plant Cell Tiss. Org.* 20, 211–215. doi: 10.1007/BF00041883

Gutierrez-Pesce, P., Taylor, K., Muleo, R., and Rugini, E. (1998). Somatic embryogenesis and shoot regeneration from transgenic roots of the cherry rootstock Colt (*Prunus avium*×*P. pseudocerasus*) mediated by *p*Ri 1855 T-DNA of *Agrobacterium rhizogenes*. *Plant Cell Rep.* 17 (6-7), 574–580. doi: 10.1007/s002990050445

Han, K.-H., Keathley, D. E., Davis, J. M., and Gordon, M. P. (1993). Regeneration of a transgenic woody legume (*Robinia pseudoacacia* L., black locust) and morphological alterations induced by *Agrobacterium rhizogenes*-mediated transformation. *Plant Sci.* 88 (2), 149–157. doi: 10.1016/0168-9452(93)90086-f

Hansen, G., Larribe, M., Vaubert, D., Tempé, J., Biermann, B. J., Montoya, A. L., et al. (1991). Agrobacterium rhizogenes pRi8196 T-DNA: mapping and DNA sequence of functions involved in mannopine synthesis and hairy root differentiation. *Proc. Natl. Acad. Sci. U. S. A.* 88 (17), 7763–7767. doi: 10.1073/pnas.88.17.7763

He, J., Plácido, J. P. A., Pateraki, I., Kampranis, S., Favero, B. T., and Lütken, H. (2022). Hairy root induction of *Taxus baccata* L. by natural transformation with *Rhizobium rhizogenes*. *Horticulturae* 9 (1), 4. doi: 10.3390/horticulturae9010004

Hildebrand, E. (1934). Life history of the hairy-root organism in relation to its pathogenesis on nursery apple trees. *J.Agr.Res.* 48 (10), 857–885.

Hoang, T. H. T., Nguyen, N. H., Nguyen, L. T., Bui, T. P., Le, N. T., Dao, N. T., et al. (2022). Developing a robust in vivo hairy root system for assessing transgene expression and genome editing efficiency in papaya. *Plant Cell Tiss. Org.* 152, 661–667. doi: 10.1007/s11240-022-02421-2

Hooykaas, P. J., Hofker, M., Den Dulk-Ras, H., and Schilperoort, R. A. (1984). A comparison of virulence determinants in an octopine Ti plasmid, a nopaline Ti plasmid, and an Ri plasmid by complementation analysis of *Agrobacterium tumefaciens* mutants. *Plasmid* 11, 195–205. doi: 10.1016/0147-619x(84)90026-x

Hooykaas, M. J. G., and Hooykaas, P. J. J. (2021). The genome sequence of hairy root *Rhizobium rhizogenes* strain LBA9402: Bioinformatics analysis suggests the presence of a new opine system in the agropine Ri plasmid. *Microbiologyopen* 10 (2), e1180. doi: 10.1002/mbo3.1180

Huang, Y., Diner, A. M., and Karnosky, D. F. (1991). *Agrobacterium rhizogenes*-mediated genetic transformation and regeneration of a conifer:*Larix decidua*. *In Vitro Cell. Dev. Biol.-Plant* 27, 201–207. doi: 10.1007/bf02632217

Huffman, G. A., White, F. F., Gordon, M. P., and Nester, E. W. (1984). Hairy-root-inducing plasmid: physical map and homology to tumor-inducing plasmids. *J. Bacteriol.* 157, 269–276. doi: 10.1128/jb.157.1.269-276.1984

Imanishi, L., Vayssieres, A., Franche, C., Bogusz, D., Wall, L., and Svistoonoff, S. (2011). Transformed hairy roots of *Discaria trinervis*: a valuable tool for studying actinorhizal symbiosis in the context of intercellular infection. *Mol. Plant-Microbe Interact.* 24 (11), 1317–1324. doi: 10.1094/MPMI-03-11-0078

Jiang, S., Jardinaud, M.-F., Gao, J., Pecrix, Y., Wen, J., Mysore, K., et al. (2021). NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules. *Science* 374 (6567), 625–628. doi: 10.1126/science.abg5945

Jiao, J., Gai, Q. Y., Wang, X., Liu, J., Lu, Y., Wang, Z. Y., et al. (2020). Effective production of phenolic compounds with health benefits in pigeon pea [*Cajanus cajan* (L.) Millsp.] hairy root cultures. *J. Agric. Food Chem.* 68 (1), 8350–8361. doi: 10.1021/acs.jafc.0c02600

Jouanin, L. (1984). Restriction map of an agropine-type Ri plasmid and its homologies with Ti plasmids. *Plasmid* 12, 91–102. doi: 10.1016/0147-619x(84)90055-6

Kaneyoshi, J., and Kobayashi, S. (1999). Characteristics of transgenic trifoliate orange (*Poncirus trifoliata* Raf.) possessing the *rol*C gene of *Agrobacterium rhizogenes* Ri plasmid. *J. Jpn. Soc Hortic. Sci.* 68 (4), 734–738. doi: 10.2503/jjshs.68.734

Kang, H. J., Anbazhagan, V. R., You, X. L., Moon, H. K., Yi, J. S., and Choi, Y. E. (2006). Production of transgenic *Aralia elata* regenerated from *Agrobacterium rhizogenes*-mediated transformed roots. *Plant Cell Tiss. Org.* 85 (2), 187–196. doi: 10.1007/s11240-005-9070-2

Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C. D., Nontachaiyapoom, S., Kinkema, M., et al. (2007). *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nat. Protoc.* 2, 948–952. doi: 10.1038/nprot.2007.141

Koshita, Y., Nakamura, Y., Kobayashi, S., and Morinaga, K. (2002). Introduction of the *rolC* gene into the genome of the Japanese persimmon causes dwarfism. *J. Jpn. Soc Hortic. Sci.* 71 (4), 529–531. doi: 10.2503/jjshs.71.529

Kumar, V., Satyanarayana, K., Sarala Itty, S., Indu, E., Giridhar, P., Chandrashekar, A., et al. (2006). Stable transformation and direct regeneration in coffea canephora p ex. fr. by agrobacterium rhizogenes mediated transformation without hairy-root phenotype. *Plant Cell Rep.* 25 (3), 214–222. doi: 10.1007/s00299-005-0045-x

Kyndt, T., Quispe, D., Zhai, H., Jarret, R., Ghislain, M., Liu, Q., et al. (2015). The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: an example of a naturally transgenic food crop. *Proc. Natl. Acad. Sci. U. S. A.* 112 (18), 5844–5849. doi: 10.1073/pnas.1419685112

Lambert, C., and Tepfer, D. (1992). Use of *Agrobacterium rhizogenes* to create transgenic apple trees having an altered organogenic response to hormones. *Theor. Appl. Genet.* 85 (1), 105–109. doi: 10.1007/BF00223851

Largia, M. J. V., Pandian, S., Shilpha, J., Chitradevi, M., Kavikkuil, M., Sohn, S.-I., et al. (2022). Improved *in vitro* regeneration, genetic fidelity analysis, antioxidant potential, and hairy root induction of *Justicia gendarussa* burm. f. *Plant Biotechn. Rep.* 16 (6), 621–632. doi: 10.1007/s11816-022-00775-9

Li, M., Chen, R., Jiang, Q., Sun, X., Zhang, H., and Hu, Z. (2021). GmNAC06, a NAC domain transcription factor enhances salt stress tolerance in soybean. *Plant Mol. Biol.* 105, 333–345. doi: 10.1007/s11103-020-01091-y

Li, M., and Leung, D. (2003). Root induction in radiata pine using *Agrobacterium rhizogenes*. *Electron J. Biotechn.* 6 (3), 254–261. doi: 10.4067/S0717-34582003000300010

Lothar, W., Jose, J. S.-S., Ellen, B., and Jeff, S. (1982). DNA from *Agrobacterium-Rhizogenes* is transferred to and expressed in axenic hairy root plant-tissues. *Mol. Genet. Genomics* 186, 16–22. doi: 10.1007/bf00422906

Lutken, H., Clarke, J. L., and Muller, R. (2012). Genetic engineering and sustainable production of ornamentals: current status and future directions. *Plant Cell Rep.* 31 (7), 1141–1157. doi: 10.1007/s00299-012-1265-5

Ma, H., Meng, X., Xu, K., Li, M., Gmitter, F., Liu, N., et al. (2022). Highly efficient hairy root genetic transformation and applications in citrus. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1039094

Magnussen, D., Clapham, D. E., Grönroos, R., and Arnold, S. V. (1994). Induction of hairy and normal roots on *Picea abies*, *Pinus sylvestris* and *Pinus contorta* by *Agrobacterium rhizogenes*. *Scand. J. For. Res.* 9, 46–51. doi: 10.1080/02827589409382811

Masako, A., Masako, A., Takuma, I., Takuma, I., and Hiroji, S. (2004). Transformation of the monocot *Alstroemeria* by *Agrobacterium rhizogenes*. *Mol. Breeding.* 13, 69–78. doi: 10.1023/b:molb.0000012860.29731.9c

Matveeva, T. V., Bogomaz, D. I., Pavlova, O. A., Nester, E. W., and Lutova, L. A. (2012). Horizontal gene transfer from genus agrobacterium to the plant linaria in nature. *Mol. Plant-Microbe Interact.* 25, 1542–1551. doi: 10.1094/MPMI-07-12-0169-R

Mauro, M. L., Costantino, P., and Bettini, P. P. (2017). The never ending story of rol genes: a century after. *Plant Cell Tiss.Org.* 131, 201–212. doi: 10.1007/s11240-017-1277-5

Mauro, M. L., Trovato, M., Paolis, A. D., Gallelli, A., Costantino, P., and Altamura, M. M. (1996). The plant oncogene *rol*D stimulates flowering in transgenic tobacco plants. *Dev. Biol.* 180 (2), 693–700. doi: 10.1006/dbio.1996.0338

McAfee, B. J., White, E. E., Pelcher, L. E., and Lapp, M. (1993). Root induction in pine (*Pinus*) and larch (*Larix*) spp. using *Agrobacterium rhizogenes*. *Plant Cell Tiss. Org.* 34, 53–62. doi: 10.1007/BF00048463

McGranahan, G. H., Leslie, C. A., Uratsu, S. L., Martin, L. A., and Dandekar, A. M. (1988). *Agrobacterium*-Mediated transformation of walnut somatic embryos and regeneration of transgenic plants. *Nat. Biotechnol.* 6 (7), 800–804. doi: 10.1038/nbt0788-800

Mehrotra, S., Goel, M. K., Rahman, L. U., and Kukreja, A. K. (2013).). Molecular and chemical characterization of plants regenerated from Ri-mediated hairy root cultures of *Rauwolfia serpentina*. *Plant Cell Tiss. Org.* 114 (1), 31–38. doi: 10.1007/s11240-013-0302-6

Meng, D., Yang, Q., Dong, B., Song, Z., Niu, L., Wang, L., et al. (2019). Development of an efficient root transgenic system for pigeon pea and its application to other important economically plants. *Plant Biotechnol. J.* 17, 1804–1813. doi: 10.1111/pbi.13101

Mihaljevic, S., Stipkovic, S., and Jelaska, S. (1996). Increase of root induction in *Pinus nigra* explants using agrobacteria. *Plant Cell Rep.* 15 (8), 610–614. doi: 10.1007/BF00232463

Mochammad, H., Caula, A. B., Garton, S., and Diner, A. M. (1996). Induction of roots on jujube softwood cuttings using *Agrobacterium rhizogenes*. *J. Hortic. Sci.* 71 (6), 881–886. doi: 14620316.1996.11515470

Moore, L., Warren, G., and Strobel, G. (1979). Involvement of a plasmid in the hairy root disease of plants caused by *Agrobacterium rhizogenes*. *Plasmid* 2, 617–626. doi: 10.1016/0147-619x(79)90059-3

Neb, D., Das, A., Hintelmann, A., and Nehls, U. (2017). Composite poplars: a novel tool for ectomycorrhizal research. *Plant Cell Rep.* 36 (12), 1959–1970. doi: 10.1007/s00299-017-2212-2

Niu, S., Li, J., Bo, W., Yang, W., Zuccolo, A., Giacomello, S., et al. (2022). The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell* 185, 204–217.e214. doi: 10.1016/j.cell.2021.12.006

Oka, S., and Tewary, P. K. (2000). Induction of hairy roots from hypocotyls of mulberry (*Morus indica*) by japanese wild strains of *Agrobacterium rhizogenes*. *J. Seric. Sci. Jpn.* 69, 13–19. doi: 10.11416/kontyushigen1930.69.13

Ono, N. N., Bandaranayake, P. C., and Tian, L. (2012). Establishment of pomegranate (*Punica granatum*) hairy root cultures for genetic interrogation of the hydrolyzable tannin biosynthetic pathway. *Planta* 236 (3), 931–941. doi: 10.1007/s00425-012-1706-y

Pateña, L., Sutter, E. G., and Dandekar, A. M. (1988). Root induction by *Agrobacterium rhizogenes* in a difficult-to-root woody species. *Acta Hortic.* 227, 324–329. doi: 10.17660/ActaHortic.1988.227.59

Pawlicki-Jullian, N., Sedira, M., and Welander, M. (2002). The use of *Agrobacterium rhizogenes* transformed roots to obtain transgenic shoots of the apple rootstock Jork 9. *Plant Cell Tiss. Org.* 70 (2), 163–171. doi: 10.1023/a:1016387004712

Petersen, S. G., Stummann, B. M., Olesen, P., and Henningsen, K. W. (1989). Structure and function of root-inducing (Ri) plasmids and their relation to tumor-inducing (Ti) plasmids. *Physiol. Plant* 77 (3), 427–435. doi: 10.1111/j.1399-3054.1989.tb05664.x

Petit, A., David, C., Dahl, G. A., Ellis, J. G., Guyon, P., Casse-Delbart, F., et al. (1983). Further extension of the opine concept: plasmids in *Agrobacterium rhizogenes* cooperate for opine degradation. *Mol. Gen. Genet.* 190, 204–214. doi: 10.1007/bf00330641

Phelep, M., Petit, A., Martin, L., Duhoux, E., and Tempé, J. (1991). Transformation and regeneration of a nitrogen-fixing tree, *Allocasuarina Verticillata* Lam. *Nat. Biotechnol.* 9 (5), 461–466. doi: 10.1038/nbt0591-461

Pi, E., Xu, J., Li, H., Fan, W., Zhu, C., Zhang, T., et al. (2019). Enhanced salt tolerance of Rhizobia-inoculated soybean correlates with decreased phosphorylation of the transcription factor GmMYB183 and altered flavonoid biosynthesis. *Mol. Cell Proteomics* 18, 2225–2243. doi: 10.1074/mcp.RA119.001704

Pitzschke, A. (2013). *Agrobacterium* infection and plant defense-transformation success hangs by a thread. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00519

Plasencia, A., Soler, M., Dupas, A., Ladouce, N., Silva-Martins, G., Martinez, Y., et al. (2016). *Eucalyptus* hairy roots, a fast, efficient and versatile tool to explore function and expression of genes involved in wood formation. *Plant Biotechnol. J.* 14 (6), 1381–1393. doi: 10.1111/pbi.12502

Plett, J. M., Daguerre, Y., Wittulsky, S., Vayssieres, A., Deveau, A., Melton, S. J., et al. (2014). Effector MiSSP7 of the mutualistic fungus Laccaria bicolor stabilizes the *Populus* JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proc. Natl. Acad. Sci. U. S. A.* 111 (22), 8299–8304. doi: 10.1073/pnas.1322671111

Porter, J. R., and Flores, H. (1991). Host range and implications of plant infection by *Agrobacterium rhizogenes*. *Crit. Rev. Plant Sci.* 10 (4), 387–421. doi: 10.1080/07352689109382318

Prabhu, S. A., Ndlovu, B., Engelbrecht, J., and Van Den Berg, N. (2017). Generation of composite *Persea americana* (Mill.) (avocado) plants: A proof-of-concept-study. *PloS One* 12 (10), e0185896. doi: 10.1371/journal.pone.0185896

Qin, Y., Wang, D., Fu, J., Zhang, Z., Qin, Y., Hu, G., et al. (2021). *Agrobacterium rhizogenes*-mediated hairy root transformation as an efficient system for gene function analysis in *Litchi chinensis*. *Plant Methods* 17 (1), 103. doi: 10.1186/s13007-021-00802-w

Quispe-Huamanquispe, D. G., Gheysen, G., and Kreuze, J. F. (2017). Horizontal gene transfer contributes to plant evolution: the case of *Agrobacterium* T-DNAs. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02015

Ramasamy, M., Dominguez, M. M., Irigoyen, S., Padilla, C. S., and Mandadi, K. K. (2023). *Rhizobium rhizogenes*-mediated hairy root induction and plant regeneration for bioengineering citrus. *Plant Biotechnol. J.* 21, 1728–1730. doi: 10.1111/pbi.14096

Rana, M. M., Han, Z., Song, D., Liu, G., Li, D., Wan, X., et al. (2016). Effect of medium supplements on *Agrobacterium rhizogenes* mediated hairy root induction from the callus tissues of *Camellia sinensis* var. *sinensis*. *Int. J. Mol. Sci.* 17 (7), 1132. doi: 10.3390/ijms17071132

Rigden, D. J., and Carneiro, M. (1999).). A structural model for the *rol*A protein and its interaction with DNA. *Proteins: Struct. Funct. Genet.* 37 (4), 697–708. doi: 10.1002/(sici)1097-0134(19991201)37:4<697::Aid-prot18>3.0.Co;2-y

Riker, A. J., Banfield, W. M., Wright, W. H., Keitt, G. W., and Sagen, H. E. (1930). Studies on infectious hairy root of nursery apple trees. *J. Agric. Res.* 41 (7), 507–540.

Risuleo, G., Battistoni, P., and Costantino, P. (1982).). Regions of homology between tumorigenic plasmids from *Agrobacterium rhizogenes* and *Agrobacterium tumefaciens*. *Plasmid* 7, 45–51. doi: 10.1016/0147-619X(82)90025-7

Rodrigues, S. D., Karimi, M., Impens, L., Van Lerberge, E., Coussens, G., Aesaert, S., et al. (2021). Efficient CRISPR-mediated base editing in Agrobacterium spp. *Proc. Natl. Acad. Sci. U. S. A.* 118 (2), e2013338118. doi: 10.1073/pnas.2013338118

Roig Celma, C., Palazon, J., Cusido, R. M., Pinol, M. T., and Keil, M. (2001). Decreased scopolamine yield in field-grown Duboisia plants regenerated from hairy roots. *Planta Med.* 67 (3), 249–253. doi: 10.1055/s-2001-12006

Rugini, E., Pellegrineschi, A., Mencuccini, M., and Mariotti, D. (1991). Increase of rooting ability in the woody species kiwi (*Actinidia deliciosa* A. *Chev.*) by transformation with *Agrobacterium rhizogenes rol* genes. *Plant Cell Rep.* 10 (6), 291–295. doi: 10.1007/BF00193144

Rugini, E., Silvestri, C., Cristofori, V., Brunori, E., and Biasi, R. (2015). Ten years field trial observations of ri-TDNA cherry Colt rootstocks and their effect on grafted sweet cherry cv Lapins. *Plant Cell Tiss. Org.* 123 (3), 557–568. doi: 10.1007/s11240-015-0860-x

Sarkar, J., Misra, A., and Banerjee, N. (2020). Genetic transfection, hairy root induction and solasodine accumulation in elicited hairy root clone of *Solanum erianthum* D. Don. *J. Biotechnol.* 323, 238–245. doi: 10.1016/j.jbiotec.2020.09.002

Sarmast, M. K., Kordkatoli, R., Rezaei, Z., and Ghasemnezhad, A. (2019). Effect of seasons, gender and *Agrobacterium rhizogenes* strains on adventitious root induction of male and female *Juniperus communis* L. *J. Ornamental Plants* 9, 23–31.

Savka, M. A., Ravillion, B., Noel, G. R., and Farrand, S. K. (1990).). Induction of hairy roots on cultivated soybean genotypes and their use to propagate the soybean cyst nematode. *Phytopathology* 80 (5), 503–508.

Sawaki, Y., Kobayashi, Y., Kihara-Doi, T., Nishikubo, N., Kawazu, T., Kobayashi, M., et al. (2014). Identification of a STOP1-like protein in *Eucalyptus* that regulates transcription of Al tolerance genes. *Plant Sci.* 223, 8–15. doi: 10.1016/j.plantsci.2014.02.011

Schmülling, T., Fladung, M., Grossmann, K., and Schell, J. (1993). Hormonal content and sensitivity of transgenic tobacco and potato plants expressing single rol genes of *Agrobacterium rhizogenes* T-DNA. *Plant J.* 3, 371–382. doi: 10.1046/j.1365-313X.1993.t01-20-00999.x

Schmülling, T., Schell, J., and Spena, A. (1988). Single genes from *Agrobacterium rhizogenes* influence plant development. *EMBO J.* 7 (9), 2621–2629. doi: 10.1002/j.1460-2075.1988.tb03114.x

Schropfer, S., Lempe, J., Emeriewen, O. F., and Flachowsky, H. (2022). Recent Developments and Strategies for the Application of *Agrobacterium*-Mediated Transformation of Apple Malus x domestica Borkh. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.928292

Sharma, P., Padh, H., and Shrivastava, N. (2013). Hairy root cultures: A suitable biological system for studying secondary metabolic pathways in plants. *Eng. Life Sci.* 13, 62–75. doi: 10.1002/elsc.201200030

She, G., Yu, S., Li, Z., Peng, A., Li, P., Li, Y., et al. (2022). Characterization of CsTSI in the biosynthesis of theanine in tea plants (*Camellia sinensis*). *J. Agric. Food Chem.* 70 (3), 826–836. doi: 10.1021/acs.jafc.1c04816

Shen, W. H., Petit, A., Guern, J., and Tempe, J. (1988).). Hairy roots are more sensitive to auxin than normal roots. *Proc. Natl. Acad. Sci. U. S. A.* 85, 3417–3421. doi: 10.1073/pnas.85.10.3417

Spanò, L., Mariotti, D., Cardarelli, M., Branca, C., and Costantino, P. (1988). Morphogenesis and auxin sensitivity of transgenic tobacco with different complements of Ri T-DNA. *Plant Physiol.* 87, 479–483. doi: 10.1104/pp.87.2.479

Spiess, L. D., Turner, J. C., Mahlberg, P. G., Lippincott, B. B., and Lippincott, J. A. (1977). Adherence of Agrobacteria to moss protonema and gametophores viewed by scanning electron microscopy. *Am. J. Bot.* 64 (10), 1200–1208. doi: 10.2307/2442482

Stanisic, M., Cosic, T., Savic, J., Krstic-Milosevic, D., Misic, D., Smigocki, A., et al. (2019). Hairy root culture as a valuable tool for allelopathic studies in apple. *Tree Physiol.* 39 (5), 888–905. doi: 10.1093/treephys/tpz006

Tepfer, D. (1984). Transformation of several species of higher plants by *Agrobacterium rhizogenes*: sexual transmission of the transformed genotype and phenotype. *Cell* 37, 959–967. doi: 10.1016/0092-8674(84)90430-6

Tepfer, D. (2017). "DNA Transfer to plants by *Agrobacterium rhizogenes*: a model for genetic communication between species and biospheres," in *Transgenesis and secondary metabolism*. Ed. S. Jha (Cham: Springer). doi: 10.1007/978-3-319-28669-3_19

Tisserant, L. P., Aziz, A., Jullian, N., Jeandet, P., Clement, C., Courot, E., et al. (2016). Enhanced stilbene production and excretion in *Vitis vinifera* cv pinot noir hairy root cultures. *Molecules* 21 (12), 1703. doi: 10.3390/molecules21121703

Traubenik, S., Reynoso, M. A., Hobecker, K., Lancia, M., Hummel, M., Rosen, B., et al. (2020). Reprogramming of root cells during nitrogen-fixing symbiosis involves dynamic polysome association of coding and noncoding RNAs. *Plant Cell* 32, 352–373. doi: 10.1105/tpc.19.00647

Trovato, M., Maras, B., Linhares, F., and Costantino, P. (2001). The plant oncogene *rol*D encodes a functional ornithine cyclodeaminase. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13449–13453. doi: 10.1073/pnas.231320398

Trumbore, S., Brando, P., and Hartmann, H. (2015). Forest health and global change. *Science* 349, 814–818. doi: 10.1126/science.aac6759

Uematsu, C., Murase, M., Ichikawa, H., and Imamura, J. (1991). *Agrobacterium*-mediated transformation and regeneration of kiwi fruit. *Plant Cell Rep.* 10 (6-7), 286–290. doi: 10.1007/BF00193143

Vadivel, A. K. A., Mcdowell, T., Renaud, J. B., and Dhaubhadel, S. (2021). A combinatorial action of GmMYB176 and GmbZIP5 controls isoflavonoid biosynthesis in soybean (*Glycine max*). *Commun. Biol.* 4 (1), 356. doi: 10.1038/s42003-021-01889-6

Van Altvorst, A. C., Bino, R. J., Van Dijk, A. J., Lamers, A. M. J., Lindhout, W. H., van der Mark, F., et al. (1992). Effects of the introduction of *Agrobacterium rhizogenes rol* genes on tomato plant and flower development. *Plant Sci.* 83, 77–85. doi: 10.1016/0168-9452(92)90064-s

Wan, X., Landhausser, S. M., Lieffers, V. J., and Zwiazek, J. J. (2006). Signals controlling root suckering and adventitious shoot formation in aspen (*Populus tremuloides*). *Tree Physiol.* 26 (5), 681–687. doi: 10.1093/treephys/26.5.681

Wang, F., Chen, H., Li, Q., Wei, W., Li, W., Zhang, W., et al. (2015). GmWRKY27 interacts with GmMYB174 to reduce expression of GmNAC29 for stress tolerance in soybean plants. *Plant J.* 83, 224–236. doi: 10.1111/tpj.12879

Wang, H., Wang, C., Liu, H., Tang, R., and Zhang, H. (2011). An efficient *Agrobacterium*-mediated transformation and regeneration system for leaf explants of two elite aspen hybrid clones *Populus alba*× *P. berolinensis* and *Populus davidiana*× *P. bolleana*. *Plant Cell Rep.* 30, 2037–2044. doi: 10.1007/s00299-011-1111-1

Wang, Y., Wang, J., Luo, D., and Jia, J. (2001). Regeneration of plants from callus tissues of hairy roots induced by *Agrobacterium rhizogenes* on *Alhagi pseudoalhagi*. *Cell Res.* 11 (4), 279–284. doi: 10.1038/sj.cr.7290097

White, F. F., Garfinkel, D. J., Huffman, G. A., Gordon, M. P., and Nester, E. W. (1983). Sequences homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature* 301, 348–350. doi: 10.1038/301348a0

White, F. F., and Nester, E. W. (1980a). Hairy root: plasmid encodes virulence traits in Agrobacterium rhizogenes. *J. Bacteriol.* 141 (3), 1134–1141. doi: 10.1128/jb.141.3.1134-1141.1980

White, F. F., and Nester, E. W. (1980b). Relationship of plasmids responsible for hairy root and crown gall tumorigenicity. *J. Bacteriol.* 144 (2), 710–720. doi: 10.1128/jb.144.2.710-720.1980

White, F. F., Taylor, B. H., Huffman, G. A., Gordon, M. P., and Nester, E. W. (1985). Molecular and genetic analysis of the transferred DNA regions of the root-inducing plasmid of *Agrobacterium rhizogenes*. *J. Bacteriol.* 164 (1), 33–44. doi: 10.1128/jb.164.1.33-44.1985

Wu, J., Wang, Y., Zhang, L., Zhang, X., Kong, J., Lu, J., et al. (2012). High-efficiency regeneration of *Agrobacterium rhizogenes*-induced hairy root in apple rootstock *Malus baccata* (L.) Borkh. *Plant Cell Tiss. Org.* 111, 183–189. doi: 10.1007/s11240-012-0182-1

Xiao, X., Ma, F., Chen, C., and Guo, W. (2014). High efficient transformation of auxin reporter gene into trifoliate orange via *Agrobacterium rhizogenes*-mediated co-transformation. *Plant Cell Tiss. Org.* 118, 137–146. doi: 10.1007/s11240-014-0469-5

Xu, S., Lai, E., Zhao, L., Cai, Y., Ogutu, C., Cherono, S., et al. (2020). Development of a fast and efficient root transgenic system for functional genomics and genetic engineering in peach. *Sci. Rep.* 10 (1), 2836. doi: 10.1038/s41598-020-59626-8

Yamashita, H., Daimon, H., Akasaka-Kennedy, Y., and Masuda, T. (2004). Plant regeneration from hairy roots of apple rootstock, *Malus prunifolia* Borkh. var. ringo Asami, strain Nagano No.1, transformed by *Agrobacterium rhizogenes*. *J. Jpn. Soc Hortic. Sci.* 73 (6), 505–510. doi: 10.2503/jjshs.73.505

Yang, Z., Gao, Z., Zhou, H., He, Y., Liu, Y., Lai, Y., et al. (2021). GmPTF1 modifies root architecture responses to phosphate starvation primarily through regulating GmEXPB2 expression in soybean. *Plant J.* 107, 525–543. doi: 10.1270/jsbbs1951.45.241

Yazawa, M., Suginuma, C., Ichikawa, K., Kamada, H., and Akihama, T. (1995). Regeneration of transgenic plants from hairy root of kiwi fruit (*Actinidia deliciosa*) induced by *Agrobacterium rhizogenes*. *Jpn. J. Breed.* 45 (2), 241–244. doi: 10.1270/jsbbs1951.45.241

Yibrah, H. S., Grönroos, R., Lindroth, A., Franzén, H., Clapham, D. E., and Arnold, S. (1996). Agrobacterium rhizogenes-mediated induction of adventitious rooting from Pinus contorta hypocotyls and the effect of 5-azacytidine on transgene activity. *Transgenic Res.* 5, 75–85. doi: 10.1007/BF01969425

Zarei, M., Salehi, H., and Jowkar, A. (2020). Controlling the barriers of cloning mature *Picea abies* (L.) H. Karst. via tissue culture and co-cultivation with *Agrobacterium rhizogenes*. *Trees* 34, 637–647. doi: 10.1007/s00468-019-01945-z

Zavattieri, M. A., Ragonezi, C., and Klimaszewska, K. (2016). Adventitious rooting of conifers: influence of biological factors. *Trees* 30, 1021–1032. doi: 10.1007/s00468-016-1412-7

Zdravkovic-Korac, S., Muhovski, Y., Druart, P., Calic, D., and Radojevic, L. (2004).). *Agrobacterium rhizogenes*-mediated DNA transfer to *Aesculus hippocastanum* L. and the regeneration of transformed plants. *Plant Cell Rep.* 22, 698–704. doi: 10.1007/s00299-004-0756-4

Zhan, X., Jones, D. A., and Kerr, A. (1988). Regeneration of flax plants transformed by *Agrobacterium rhizogenes*. *Plant Mol. Biol.* 11, 551–559. doi: 10.1007/BF00017455

Zhang, C., Cheng, Q., Wang, H., Gao, H., Fang, X., Chen, X., et al. (2021). GmBTB/POZ promotes the ubiquitination and degradation of LHP1 to regulate the response of soybean to Phytophthora sojae. *Commun. Biol.* 4 (1), 372. doi: 10.1038/s42003-021-01907-7

Zhang, X., Gong, X., Cheng, S., Yu, H., Li, D., Su, X., et al. (2022). Proline-rich protein *MdPRP6* alters low nitrogen stress tolerance by regulating lateral root formation and anthocyanin accumulation in transgenic apple (*Malus domestica*). *Environ. Exp. Bot.* 197, 104841. doi: 10.1016/j.envexpbot.2022.104841

Zhu, L., Li, X., Ahlman, A., and Welander, M. (2003). The rooting ability of the dwarfing pear rootstock BP10030 (*Pyrus communis*) was significantly increased by introduction of the *rol*B gene. *Plant Sci.* 165 (4), 829–835. doi: 10.1016/s0168-9452(03)00279-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 December 2023
DOI 10.3389/fpls.2023.1302536

OPEN ACCESS

EDITED BY
Meng-Zhu Lu,
Zhejiang Agriculture &Forestry University, China

REVIEWED BY
Igor Cesarino,
University of São Paulo, Brazil
Yoshimi Nakano,
National Institute of Advanced Industrial Science and Technology (AIST), Japan

*CORRESPONDENCE
Marie Baucher
marie.baucher@ulb.be

RECEIVED 26 September 2023
ACCEPTED 29 November 2023
PUBLISHED 21 December 2023

CITATION
Guérin C, Behr M, Sait J, Mol A, El Jaziri M and Baucher M (2023) Evidence for poplar PtaPLATZ18 in the regulation of plant growth and vascular tissues development. *Front. Plant Sci.* 14:1302536. doi: 10.3389/fpls.2023.1302536

Evidence for poplar PtaPLATZ18 in the regulation of plant growth and vascular tissues development

Claire Guérin, Marc Behr, Julie Sait, Adeline Mol, Mondher El Jaziri and Marie Baucher*

Laboratoire de Biotechnologie Végétale, Université libre de Bruxelles, Gosselies, Belgium

Introduction: Plant A/T-rich protein and zinc-binding protein (PLATZ) are plant-specific transcription factors playing a role in plant development and stress response. To assess the role of PLATZs in vascular system development and wood formation in poplar, a functional study for *PtaPLATZ18*, whose expression was associated with the xylem, was carried out.

Methods: Poplar dominant repressor lines for *PtaPLATZ18* were produced by overexpressing a *PtaPLATZ18-SRDX* fusion. The phenotype of three independent transgenic lines was evaluated at morphological, biochemical, and molecular levels and compared to the wild type.

Results: The PtaPLATZ18-SRDX lines showed increased plant height resulting from higher internode length. Besides, a higher secondary xylem thickness was also evidenced in these dominant repression lines as compared to the wild type suggesting an activation of cambial activity. A higher amount of lignin was evidenced within wood tissue as compared to the wild type, indicating an alteration in cell wall composition within xylem cell types. This latter phenotype was linked to an increased expression of genes involved in lignin biosynthesis and polymerization.

Discussion: The phenotype observed in the PtaPLATZ18-SRDX lines argues that this transcription factor targets key regulators of plant growth and vascular tissues development.

KEYWORDS

PLATZ, transcription factor, poplar, vascular tissue development, lignin, secondary growth

1 Introduction

PLATZ are plant-specific transcription factors (TFs) that bind A/T-rich DNA sequences. They are characterised by two conserved domains required for their zinc-dependent DNA-binding, C-X_2-H-X_{11}-C-X_2-C-$X_{(4-5)}$-C-X_2-C-$X_{(3-7)}$-H-X_2-H and C-X_2-C-$X_{(10-11)}$-C-X_3-C (Nagano et al., 2001). Some PLATZ proteins have been localized both in the nucleus, which is consistent with their function as transcriptional regulators, and in the cytoplasm, suggesting a role in cellular signaling pathways (So et al., 2015; Li et al., 2017; Kim et al., 2018; Zhou and Xue, 2020; Li et al., 2021; Han et al., 2022; Wai et al., 2022; Zhao et al., 2022). PLATZ TFs are central elements in protein accumulation by acting at two distinct regulation levels. On the one hand, they regulate the biosynthesis of transfer RNAs (tRNAs) and 5S ribosomal RNA (5S rRNA) through protein-protein interactions within the RNA polymerase III complex (Pol III) (Li et al., 2017; Wang et al., 2018). On the other hand, several PLATZ can control transcription by binding A/T-rich sequences in a Pol II-dependent process (Kim et al., 2018). PLATZ have been identified either as transcriptional activators (Kim et al., 2018; Hu et al., 2023) or repressors (Nagano et al., 2001; Liu et al., 2020; Zhao et al., 2022; Zhang et al., 2023). The size of the PLATZ family varies between species. For instance, 12 *PLATZ* genes have been identified in Arabidopsis (Wang et al., 2018), 17 in *Zea mays* (Wang et al., 2018), and 18 in *Populus trichocarpa* (Ma et al., 2023).

The characterization of several *PLATZ* from monocot and dicot plants has highlighted their role in the regulation of a large panel of stress responses and/or developmental processes. First, the expression of several *PLATZ* is induced under abiotic stress, and they play a role in regulating abiotic stress tolerance. For instance, in Arabidopsis, AtPLATZ1 and AtPLATZ2 have been identified as regulators of seed desiccation tolerance (González-Morales et al., 2016). AtPLATZ1/AIN1 is also involved in ABA-mediated inhibition of root elongation by modulating ROS homeostasis (Dong et al., 2021). AtPLATZ2 is involved in the regulation of stress tolerance as lines overexpressing *AtPLATZ2* had an increased sensitivity to salt stress (Liu et al., 2020). AtPLATZ4 was found to promote drought tolerance by directly repressing *Plasma membrane Intrinsic Protein 2;8* (*PIP2;8*) as well as by regulating genes involved in ABA signalling (Liu et al., 2023). In poplar, *PtPLATZ1-4* and *PtPLATZ8-9* are involved in cadmium tolerance as evidenced by heterologous expression in yeast and overexpression in poplar (Ma et al., 2023).

Second, PLATZ were found to regulate developmental processes including seed filling, floral development, and organ as well as plant growth. For instance, the *Z. mays* semi-dominant mutant *floury 3* (*fl3*) showed a defect in endosperm development (Li et al., 2017). FL3 (also named ZmPLATZ12) interacts with proteins of the Pol III complex including RPC53, TbBRF1, NRPC2, and TFC1, and the *fl3* mutant has defects in Pol III activity as shown by reduced accumulation of 5S rRNA and several tRNAs (Li et al., 2017; Zhao et al., 2020). Another maize PLATZ, ZmPLATZ2, was found to directly bind to the *ZmSSI* starch synthetic gene promoter and when *ZmPLATZ2* was transiently expressed in maize endosperm or overexpressed in rice, an upregulation of starch synthesis genes was observed (Li et al., 2021). In *O. sativa*, the *Osfl3* CRISPR/Cas9 knockout lines were also impacted in grain development as shown by their reduced grain weight and size (Guo et al., 2022). *FL3* was positively regulated by the *Triticum aestivum* Positive regulator of Grain Size1 (PGS1) basic helix-loop-helix TF through binding to the E-box motif in the *FL3* promoter of both wheat and rice (Guo et al., 2022). The rice GLABROUS GENE 6 (GL6)/SHORT GRAIN 6 (SG6) was found to participate in the Pol III transcription machinery by interacting with the RPC53 and TFC1 subunits to promote cell division in the spikelet hull, which positively controls grain length (Wang et al., 2019; Zhou & Xue, 2020). Finally, the ectopic expression of the soybean *GmPLATZ* in Arabidopsis triggers larger seeds with a 12.5-49% increase in the 1000-seed weight (Hu et al., 2023). On the opposite, these authors showed that the knockout *gmpla/b* soybean mutants had a ~10% lower 100-seed weight and smaller seeds with reduced cell numbers but no difference in seed nutrient composition compared to the control.

The *Vitis vinifera* VviPLATZ1 is involved in the control of female flower morphology as in the CRISPR/Cas9 mutants, female flowers with reflex stamens were observed instead of hermaphrodite flowers in the wild type (WT) (Iocco-Corena et al., 2021). The Arabidopsis *AtPLATZ10* expression was reported to be controlled by the Agamous-mediated biotimer suggesting a role in flower development (Pelayo et al., 2023).

Phenotypic and molecular analyses of Arabidopsis activation-tagged lines and knockout mutants demonstrated that AtPLATZ3 (ORESARA 15, ORE15) controls cell proliferation during growth as well as leaf senescence and directly regulates the genes encoding the TFs GROWTH-REGULATING FACTOR (GRF) 1/4 in a Pol II-dependent process (Kim et al., 2018). Consistently, the expression of *ORE15* peaks in young leaves and progressively decreases during leaf aging (Kim et al., 2018; Jun et al., 2019). In addition, ORE15 was reported to be a regulator of root growth and development as it positively controls root apical meristem size in a process involving auxin and cytokinin signalling-related pathways (Timilsina et al., 2022). Transgenic Arabidopsis overexpressing *AtPLATZ4* exhibited a lower fresh weight than the WT, and it was suggested that AtPLATZ4 acts as a negative regulator of growth by suppressing the expression of *EXP6* and *EXP16 expansin* genes thereby inhibiting the expansion of the cell wall (Liu et al., 2023).

A role of PLATZs in developmental processes linked to perennial growth, such as wood formation has not been reported yet. However, there are indications of the involvement of PLATZ in secondary cell wall (SCW) formation in several plant species. Actually, in hemp and nettle, two species harboring extraxylary fibers with a gelatinous-type SCW, several *PLATZs* were upregulated in thickening bast fibers, either as compared to elongating bast fibers or to the xylem core of the stem (Guerriero et al., 2017; Xu et al., 2019). These genes were among the most highly expressed TF-encoding genes in the investigated tissues, supporting a role in SCW thickening. In addition, a transcriptomic analysis made in *Populus deltoides* × *P. euramericana* along a gradient from the shoot apex to the fifth internode, where secondary growth is well developed, highlighted the preferential expression of several *PLATZs* in internodes with cells undergoing SCW thickening (Chao et al., 2019). As this expression pattern is

comparable to those of well-described regulators of SCW biosynthesis, such as homologs of Arabidopsis *NST2*, *MYB46*, and *MYB83*, these authors suggested a role for some *PLATZ* genes in the transition from primary to secondary growth. Finally, Potri.013G078500 (PLATZ14) was detected in a shotgun analysis of the nuclear-fractionated proteome of stem-differentiating xylem tissue of *P. trichocarpa* (Loziuk et al., 2017).

To assess the role of *PLATZs* in vascular system development and wood formation in poplar (*P. tremula x P. alba*), a functional study for *PtaPLATZ18*, whose expression was associated with the xylem, was investigated. The *PtaPLATZ18* dominant repression (SRDX) lines produced by expressing full-length *PtaPLATZ18* in frame with the *SUPERMAN repression domain X* (*SRDX*) displayed an enhanced plant elongation. Besides, a higher secondary xylem thickness was also evidenced in the dominant repression lines of *PtaPLATZ18* as compared to WT suggesting an activation of cambial activity in the transgenic lines. Finally, a significantly higher amount of lignin was evidenced within wood tissue as compared to the WT, indicating an alteration in cell wall composition within xylem cell types. This report evidences a role for *PtaPLATZ18* in vascular tissues development in poplar, a woody perennial species.

2 Materials and methods

2.1 *In silico* expression analysis

The RNA-seq data used in this study were retrieved from Sundell et al. (2017). Hierarchical clustering and heatmapping of *P. tremula PtPLATZ* genes were conducted using Mev software (Howe et al., 2011) with Euclidean distance and complete linkage clustering.

2.2 Plant material

Both WT and transgenic poplars (*P. tremula* x *P. alba* clone INRA 717-1B4) were grown aseptically and then transferred in soil in the phytotron for three months, as described previously (25°C, 16h day/8h night, 55% humidity) (Baldacci-Cresp et al., 2015).

2.3 Cloning of *PtaPLATZ18*, gene constructs, plant transformation, and selection of homogeneous transgenic poplar lines

PtaPLATZ18 was amplified from *P. tremula* x *P. alba* 717-1B4 cDNA (primers are given in Supplementary Table 1), cloned into the pCR BLUNT II TOPO (Life Technologies, Carlsbad, CA, USA) and sequenced. For the *PtaPLATZ18-SRDX* fusion construct, *PtaPLATZ18* was further amplified by PCR using a forward primer containing *att*B1 and a reverse primer containing the sequence encoding both the SRDX (LDLDLELRLGFA) peptide (Hiratsu et al., 2003) and the *att*B2 sequence (Supplementary Table 1), and cloned in the pK7WG2D (Karimi et al., 2002) plasmid by Gateway technology (Supplementary Figure 1). The binary vector was transformed into poplar *via Agrobacterium tumefaciens* ($C58^{Rif}$PMP90) (Leple et al., 1992). A minimum of 20 independent transgenic lines were produced and the expression level of the *PLATZ18-SRDX* construct was analyzed for 15 of them (Supplementary Figure 2). Subsequently, six lines were multiplied *in vitro* and grown for three months in soil in a phytotron (25°C, 16h day/8h night, 55% of humidity), with five biological replicates. Among them, three lines (SRDX2, SRDX19, and SRDX20), showing a significant increase in height, were selected for further characterization (Supplementary Figure 2).

2.4 Expression profile analyses by RT-qPCR

For the expression profile analysis of *PLATZ* genes in different organs and tissues, three biological replicates and two trees per biological replicate were sampled. Samples from the apex (first 2 cm of the stem), young stem (internodes from 2-7 cm under the apex), young leaves (numbers 10-12 under the first expanded leaf), young roots and lignified roots, as well as secondary vascular tissues from a portion of the stem taken at 15-45 cm below the apex *i.e.* xylem (debarked stem, scratched with a scalpel) and phloem (bark) of 3-month-old WT poplars, were collected and immediately frozen in liquid nitrogen.

For the analysis of SRDX2, SRDX19, and SRDX20 lines, young stem (internode number 1 to 4 under the first expanded leaf) and xylem (debarked stem from 30 to 60 cm from the ground, scratched with a scalpel) were sampled and immediately frozen in liquid nitrogen. Five biological replicates, made of one plant per replicate, were sampled for each SRDX line and the associated WT.

Total RNA was extracted according to Muoki et al. (2012), followed by a DNAse treatment using the TURBO DNA-Free kit (Fisher Scientific), according to the manufacturer's instructions. For each RNA sample, reverse transcription was performed on 2 µg of total RNA, using the Protoscript II First Strand cDNA Synthesis Kit (New England Biolabs). RT-qPCR were carried out in three technical replicates in a 20 µl volume containing 25 ng of cDNA, 0.5 µl primers (10 µM), and 10 µl of Luna Universal qPCR Master Mix (NEB #M3003) using a thermocycler LightCycler 480 system (Roche). The thermal cycle was set up as follows: pre-incubation at 95°C/1 min, 40 amplification cycles of 95°C/15 s then 60°C/15 s, a melting curve step at 95°C/5 s, 65°C/1 min then heat up to 97°C at 0.11°C/s, and a cooling step to 40°C. Relative gene expression was determined by the $2^{-\Delta\Delta C}{}_{T}$ method (Livak and Schmittgen, 2001), using adjusted primer efficiencies, and *CDC2* as a reference gene (Baldacci-Cresp et al., 2016) for expression normalization. Monolignol biosynthesis genes were identified by Wang et al. (2018) and the genes putatively involved in monolignol polymerization were identified by Sundell et al. (2017). Genes coding for TFs associated with secondary cell wall biosynthesis in *P. trichocarpa* were taken from Li et al. (2012) (*SND1-A2* and *VND6-C1*) and Payyavula et al. (2022) (*MYB3* and *MYB28*), and primers for corresponding genes in *P. tremula x P. alba* were designed using PrimerQuest. Primer sequences are given in Supplementary Table 1.

2.5 Plant phenotyping

Poplar plant height was measured each week for three months. Then, mature leaves (numbers 19 to 21 under the first expanded leaf) from WT and transgenic lines were scanned and analyzed to estimate the mean leaf area, using Digimizer (https://www.digimizer.com/). The internode number was counted starting from the first expanded leaf and the internode mean length was determined by dividing plant height by internode numbers. The stem diameter was measured using a calliper at the basis of the stem. A 2 cm sample was retrieved from the internode number 6 and kept at 4°C in 70% ethanol. Stem cross sections (110 µm thick) were made using an HM 650 V vibratome and photographed by a stereomicroscope (Leica M165-FC). The pictures were analyzed using ImageJ (Schneider et al., 2012) to estimate the percentage of bark, xylem, and pith in the stem.

Stem cross sections were subjected to phloroglucinol staining as described in Speer (1987) or to Maüle staining (5 min in 4% $KMnO_4$, three water baths, 80 sec in 3% HCl, three water baths, observation under microscope in a drop of Tris-HCl 1M) adapted from Yamashita et al. (2016), then examined and photographed under a wide-field microscope (Axio Observer Z1).

2.6 Cell wall residue preparation

The basal part (30 cm long) of the debarked stem was dried at 50°C for 48 h and then ground using a ball mill. One gram of wood powder was mixed with 10 ml of methanol:water (80:20, v/v), sonicated for 10 min, and homogenised for 4 h at room temperature. After 2 min of centrifugation at 4000 g, the pellet was washed two times with ethanol 80% and three times with pure acetone, then dried at 35°C for 24 h. The cell wall residue (CWR) quantification was performed on 300 mg of wood powder, with two technical replicates and three biological replicates per line, and the same protocol as described above. The CWR powder was used as material for both the lignin and cellulose analyses.

2.7 Lignin content

The lignin content was determined using the cysteine-assisted sulfuric acid (CASA) method (Lu et al., 2021). Briefly, 10 mg of CWR was fully dissolved with 1 mL of 10% L-cysteine (w/v) in 72% sulfuric acid after an incubation of 1 h at room temperature under agitation. The solution was diluted to 100 mL with pure water and CASA lignin content was evaluated based on the absorbance at 283 nm and an absorption coefficient of 11.23 $g^{-1}{\cdot}L{\cdot}cm^{-1}$. Five biological replicates and two technical replicates were performed.

2.8 Cellulose content

To determine the cellulose content, an Updegraff treatment was performed on the CWR powder. The Updegraff reagent is composed of acetic acid, nitric acid, and pure water in 8:1:2 proportions. 1.5 ml of this reagent was added to the samples and heated to 100°C for 30 min. After a centrifugation of 15 min and 4000 g, the pellet was washed three times with pure water and two times with pure acetone, then dried at 35°C for 24 h.

The cellulose content was determined on 10 mg of the obtained powder, according to the protocol described by Dampanaboina et al. (2021). The equation used here is: $Cellulose\ content(\%\ CWR) = \frac{[Glu]\cdot A\cdot B}{1.1\cdot m}\cdot 100$ with [Glu] as the glucose concentration (µg/ml), A the dilution factor in anthrone solution (here: 0.5), B the sample volume taken to measure the glucose concentration (µl) (here: 0.1), and m the sample mass (mg).

3 Results

3.1 Several *PLATZ* are expressed within vascular tissues in poplar

As a preliminary investigation, the expression pattern of the 18 poplar *PLATZ* genes was studied by retrieving and examining data issued from an RNA-seq analysis performed on a series of laser-microdissected tangential stem sections from phloem to lignified xylem in a 45-year-old *P. tremula* (Sundell et al., 2017). As shown in Figure 1, 16 *PLATZs* are expressed within vascular tissues. Some of them are preferentially expressed in the cambial zone (such as *PtPLATZ9*) whereas the expression of others is linked to phloem, xylem, or not differential within the stem. Particularly, six of them (*PtPLATZ1, PtPLATZ2, PtPLATZ7, PtPLATZ10, PtPLATZ14*, and *PtPLATZ18*) display a high expression level in lignified xylem (secondary xylem) suggesting a role for these genes in xylem development and/or SCW formation.

To assess the expression profile of these six genes in the model poplar *P. tremula x P. alba* grown in the phytotron for three months, a RT-qPCR was performed on secondary vascular tissues, *i.e.*, debarked stem (made mainly of secondary xylem) and bark (including phloem). As shown in Figure 2, the six *PtaPLATZ* genes showed different expression profiles. Whereas *PtaPLATZ2* was lowly expressed in both tissues, the five other genes, and particularly *PtaPLATZ1, PtaPLATZ14*, and *PtaPLATZ18*, had a higher expression in the xylem compared to bark. To evaluate the specificity of the expression profile of these genes, a complementary expression analysis was made on different organs *i.e.* apex, leaf, young stem, as well as young roots (containing no secondary vascular tissues) and old roots (containing secondary vascular tissues) (Supplementary Figure 3). *PtaPLATZ1* was strongly expressed in leaves and roots compared to the apex and young stem, *PtaPLATZ2* had a low but preferential expression in the apex and young roots, *PtaPLATZ7* was expressed at a similar level in all samples, and *PtaPLATZ10* was preferentially expressed in the young root. *PtaPLATZ14* and *PtaPLATZ18* were expressed in all parts but with a 2-3-fold lower relative expression level (Supplementary Figure 3) than in the xylem (Figure 2). As a case study, *PtaPLATZ18* was selected as a relevant candidate gene for a functional study in the context of vascular tissue development and lignification in poplar.

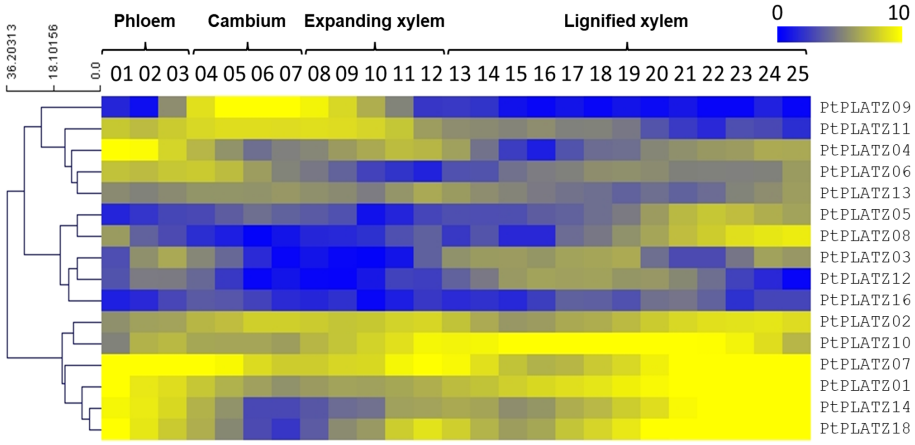


FIGURE 1
PLATZ expression profile as determined by RNA-seq on a series of tangential laser-microdissected sections from phloem to lignified xylem of a 45-year-old *Populus tremula* tree (data from Sundell et al. (2017)). There were no RNA-seq data for *PtPLATZ15* and *PtPLATZ17*.

3.2 The dominant repression of *PtaPLATZ18* enhances plant elongation and xylem development

To explore the function of *PtaPLATZ18* in vascular tissue development and/or lignification, we produced dominant repressor lines for *PtaPLATZ18* by generating poplar lines overexpressing a *PtaPLATZ18-SRDX* fusion. Six transgenic lines expressing *PtaPLATZ18-SRDX* were grown *in vitro* in phytotron for three months. Five of these lines were characterized by a significantly increased height (Supplementary Figure 2) and three of them (SRDX2, SRDX19, and SRDX20) were selected for further characterization. The expression level of both endogenous *PtaPLATZ18* and *PtaPLATZ18-SRDX* was analysed in the secondary xylem by RT-qPCR. As shown in Figure 3, the endogenous *PtaPLATZ18* expression level was similar for all lines whereas *PtaPLATZ18-SRDX* was highly expressed in the three SRDX lines and not detectable in the WT.

Figure 4 shows the 3-month-old PtaPLATZ18-SRDX lines (SRDX2, SRDX19, and SRDX20) which displayed an enhanced growth compared to the WT (Figure 4A) as their height was increased by 22 to 38% (Figure 4B). This increase in height was due to an increased mean length of the internodes by 25 to 45% (Figure 4C). There was no significant difference in the diameter measured at the basis of the stem of the three transgenic lines and the WT (Figure 4D). In addition, SRDX lines exhibit smaller leaves than the WT with a leaf area decreased by 32-38% (Figure 4E).

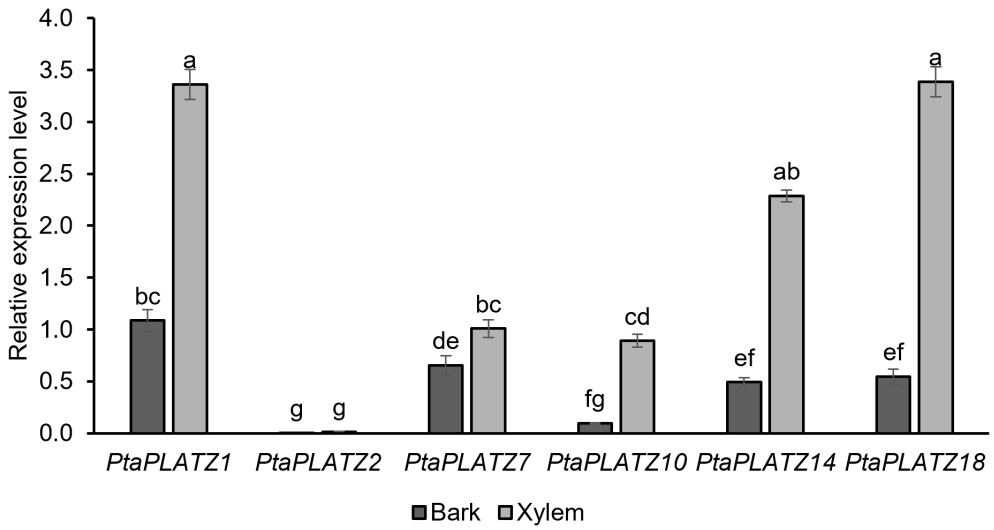


FIGURE 2
Expression level of six *PtaPLATZ* genes in bark and xylem of the stem of 3-month-old *P. tremula x P. alba* grown in the greenhouse, as measured by RT-qPCR. Bars represent the standard error, from five biological replicates and three technical replicates. Letters indicate the significantly different groups identified by a Kruskal-Wallis test with a Bonferroni correction ($P < 0.05$).

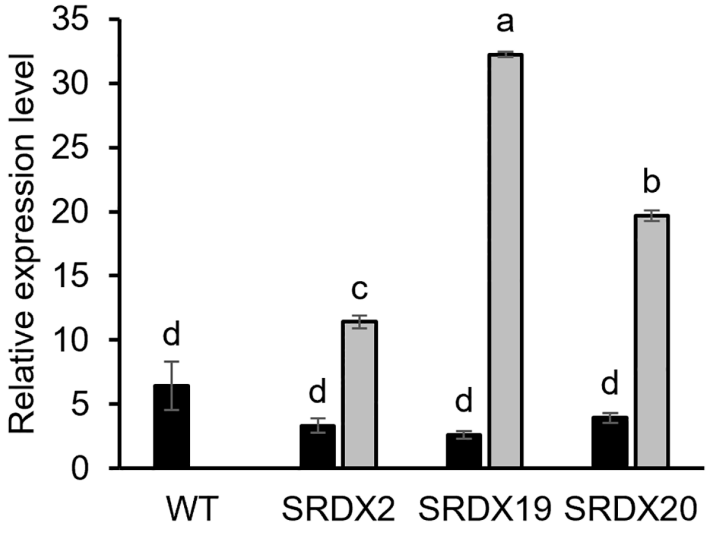


FIGURE 3
Relative expression level of endogenous *PtaPLATZ18* (black) and *PtaPLATZ18-SRDX* (grey) in the xylem of 3-month-old SRDX and WT lines, as measured by RT-qPCR. Values are the mean and the standard error of five plants (biological replicates) and three technical replicates. The values have been normalized by using the *CDC2* reference gene. Letters indicate the significantly different groups identified by an ANOVA and Tukey, $P < 0.05$.

To characterize the SRDX lines at the anatomical level, stem cross sections were realized in the 6th internode below the first expanded leaf and stained with phloroglucinol-HCl to visualize lignified tissues. As shown in Figure 5A and Supplementary Figure 4, transgenic lines displayed a more developed secondary xylem than the WT. To quantify this observation, bark, xylem, and pith proportions were measured, providing their percentage in the stem sections (Figure 5B). A significant increase in the proportion of xylem by 38%, 21%, and 26% was observed in SRDX2, SRDX19, and SRDX20 lines respectively, compared to the WT. This increase in xylem size takes place at the expense of bark, whose proportion decreases by 13%, 16%, and 29% in the three SRDX lines, respectively, compared to the WT.

3.3 PtaPLATZ18-SRDX lines have a higher lignin content in the stem xylem than the WT

Subsequently, the impact of the *PtaPLATZ18-SRDX* expression on the cell wall components accumulation was evaluated on CWR produced from wood, taken from the 30 cm basal portion of the stem. As shown in Table 1, the amount of recovered CWR was similar for the three SRDX lines and the WT. When compared to the WT, no significant modification was found regarding the cellulose content of the SRDX lines (Table 1). On the contrary, they contain a significantly higher lignin content, by 11 to 22%, depending on the lines. A Maüle staining was performed on stem cross section taken at a similar level to those for the phloroglucinol staining to evaluate possible change in lignin composition. This experiment did not reveal a clear difference in qualitative lignin composition between the WT and the SRDX lines (Supplementary Figure 5).

3.4 Higher expression of genes involved in the biosynthesis and polymerization of lignin in the xylem of PtaPLATZ18-SRDX lines

To evaluate whether the increased proportion of secondary xylem tissue and lignin accumulation in the SRDX lines were associated with a raised expression of genes involved in monolignol biosynthesis and/or polymerization, the expression level of several related genes (Supplementary Table 1) was subsequently monitored in both whole young stem (at a level comparable to stem sections of Figure 6, Supplementary Figures 4, 5) and in xylem (at a level comparable the material used for lignin quantification).

In young stems of the three SRDX lines, among the 14 studied genes (Figure 6A, Supplementary Figure 6), four of them showed a significative ~2-fold increase in their expression level, compared to the WT. These genes include *phenylalanine ammonia-lyase* 2 (*PAL2*), *hydroxycinnamoyl-CoA shikimate transferase 1* (*HCT1*), *ferulate-5 hydroxylase 1* (*F5H1*), and the gene homologous to the *A. thaliana peroxidase 72* (*PRX72*). The expression level of these four genes was further evaluated in xylem. As shown in Figure 6B, their expression level was higher than in the young stem and there was no difference between SRDX and WT lines. As there was no differential expression in lignin-related genes in the xylem, one can hypothesize that the higher lignin content in SRDX lines is due to the enhanced expression of lignin biosynthetic and polymerization genes within the young stem, containing mainly developing xylem.

To evaluate which regulation pathway may be altered by the expression of the *PtaPLATZ18-SRDX* construct, the expression level of several genes coding for TFs involved in the regulation of secondary cell wall differentiation was investigated in young stem tissues. Whereas no significant change was monitored for the

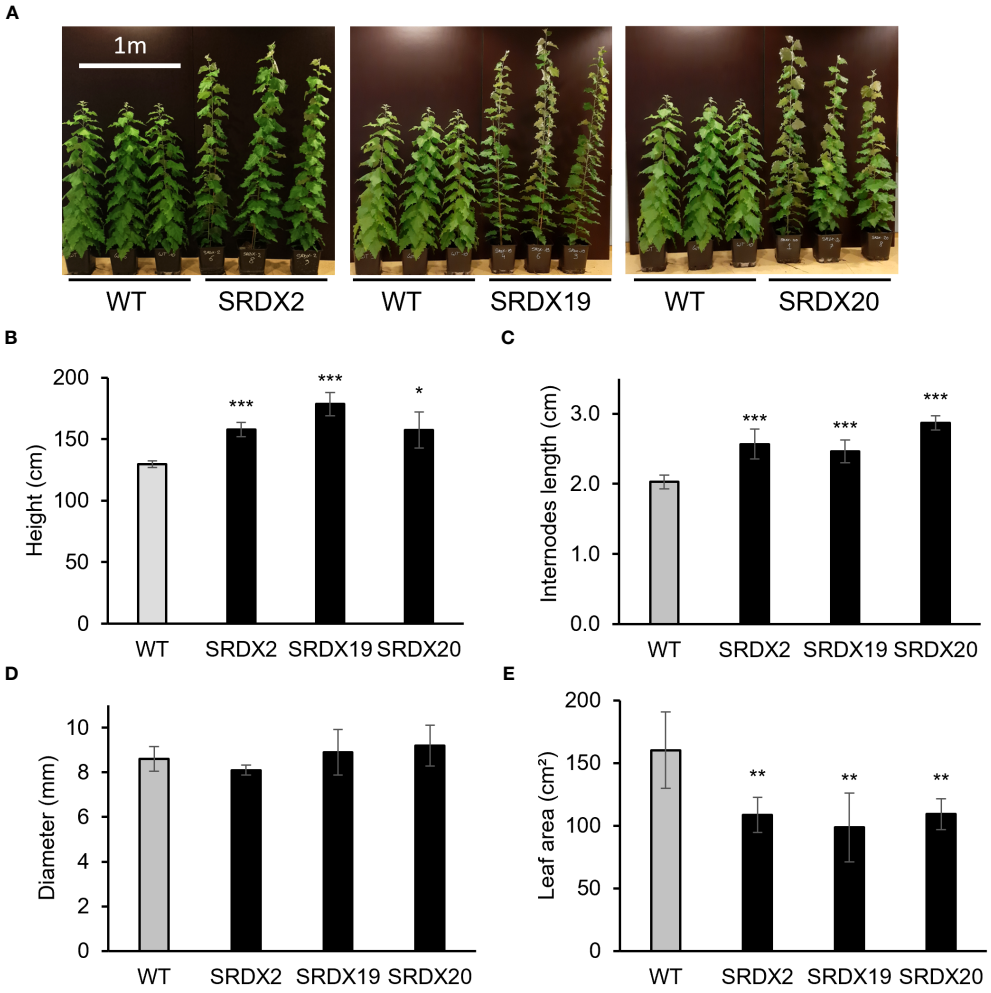


FIGURE 4
Phenotype of 3-month-old PtaPLATZ18-SRDX poplar lines and the WT. **(A)** Global phenotype. **(B)** Height. **(C)** Internodes mean length. **(D)** Diameter of the basal stem. **(E)** Mean area of leaves numbers 19 to 21. Five biological replicates. Bars = standard deviation. Student *t*-test was performed for each mutant line against the WT. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

expression of *SND1-A2*, *MYB3*, or *MYB28*, that of *VND6-C1* was 1.6 to 2.6-fold higher than in the WT, depending on the line (Supplementary Figure 7).

4 Discussion

Here, we analysed the expression profile of several *PLATZ* genes in different tissues and organs of poplar, and selected *PtaPLATZ18* as a candidate gene regulating secondary growth and/or secondary cell wall formation as it is preferentially expressed within xylem compared to other analysed plant tissues and organs (Figures 1, 2 and Supplementary Figure 3). We show that the dominant repression of *PtaPLATZ18* triggers phenotypical alterations at morphological, anatomical, and biochemical levels as discussed below.

At the morphological level, dominant repression of *PtaPLATZ18* results in an enhanced growth (by 22-38% depending on the line) associated with increased internode elongation (by 25-45% depending on the line) (Figure 4), suggesting that *PtaPLATZ18* controls plant growth and development as already reported for other *PLATZ* genes. For instance, the dominant Arabidopsis *ore15-1D* (*AtPLATZ3*) activation mutant had a 28% increase in plant height when compared to the WT (Kim et al., 2018). Likewise, the overexpression of *T. aestivum PLATZ1-A1* leads to an increased plant height and inversely its mutation caused a reduction of plant height compared to the WT. This gene was found to be the causative gene underlying the gibberellic acid (GA)-sensitive *REDUCED HEIGHT* (*RHT*) *25* dwarfing locus (Zhang et al., 2023). By using yeast two-hybrid and co-immunoprecipitation, these authors showed that PLATZ1-A1 interacts with RHT1, as well as with DELLA, a critical component of the GA growth-stimulating pathway. Recently, *gmpla/b* soybean mutants were described to have reduced leaf size and plant height, which was caused by a reduced internode elongation compared to the WT (Hu et al., 2023). GmPLATZ was found to regulate 3 A-type cyclin (*GmCYCA2;2, GmCYCA2;4a* and *GmCYCA2;4b*) and 3 D-type cyclin (*GmCYCD1;1, GmCYCD4;1* and *GmCYCD6;1*) genes. Besides, this study reported that the expression of *GmGA20OX*

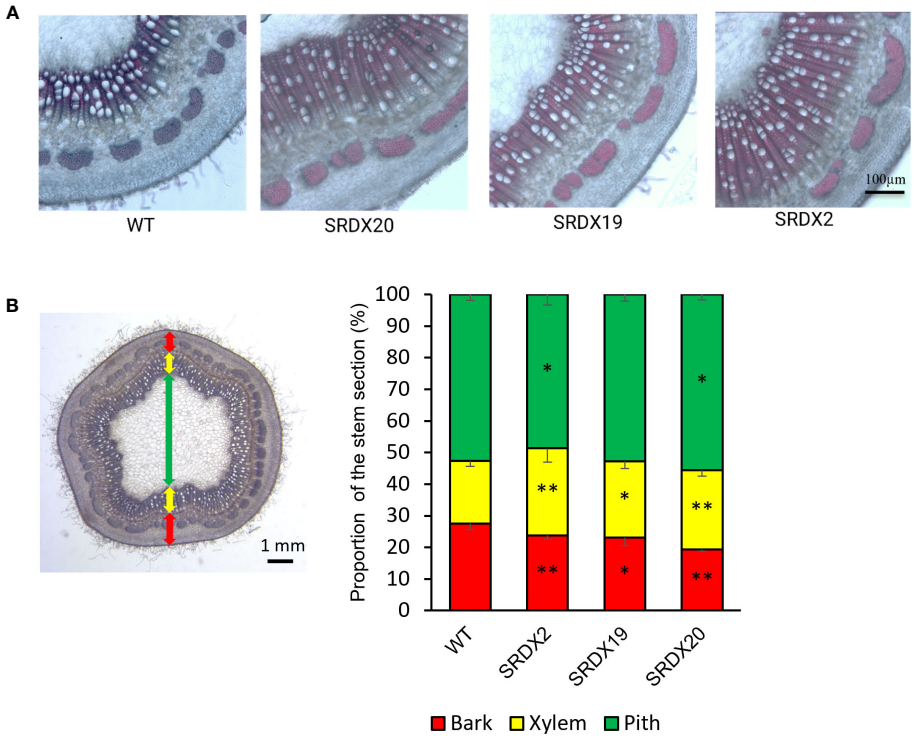


FIGURE 5
Enhanced xylem formation in 3-month-old PtaPLATZ18-SRDX lines compared to the WT. **(A)** Phloroglucinol-HCl stained stem cross sections (110 µm thickness) of the 6th internode from the apex visualized using bright field microscopy. **(B)** Percentage of bark, xylem, and pith stem sections. Two measures were performed on three stem sections per plant and five plants were studied per line. Student *t*-test was done for each mutant line against the WT. *, $P < 0.05$; **, $P < 0.01$. Bars = standard deviation.

was downregulated in the *gmpla/b* mutants and that GmPLATZ activates directly the promoter of *GmGA20OX* through the core element 5'-AATGGGCATT-3' (Hu et al., 2023).

Stem elongation and growth are known to be regulated by GA level (Davies, 2010). For instance, in poplar, the overexpression of Arabidopsis *GA20 oxidase*, a key enzyme for producing bioactive GA, was reported to increase plant growth and induce xylogenesis (Eriksson et al., 2000; Cho et al., 2019). Actually, 3-month-old transgenic poplars overexpressing the *Pinus densiflora PdGA20ox1* showed an increase of up to ~55% in height with no changes in stem diameter, when compared to the control (Cho et al., 2019).

Further, the dominant repression of *PtaPLATZ18* results in a leaf area decreased by 32-38%, depending on the line (Figure 4E). A similar observation was reported for the *ore15-2* knockout mutant (*AtPLATZ3*) characterized by a decreased leaf size mainly caused by an alteration of cell proliferation rather than cell expansion (Kim et al., 2018). ORE15 was found to interact directly with the promoters of *GRF1* and *GRF4*, which are major regulators of plant growth and development including leaf development (Gonzalez et al., 2012).

Pleiotropic effects of elevated GA concentration were noticed in transgenic poplars overexpressing *AtGA20ox1* such as smaller leaf area (Eriksson et al., 2000). In addition, a reduction of 38-40% in leaf area was measured in 2-month-old p35S-PdGA20ox1 poplar lines (Cho et al., 2019).

At the anatomical level, dominant repression of *PtaPLATZ18* results in a significant increase in the proportion of xylem tissue (Figure 5) indicating that PtaPLATZ18 plays a role in the formation of secondary growth by controlling cambial activity. Several other genes have been found to promote cambial activity and xylem development in poplar, together with plant height. For instance, the overexpression of *PagRabE1b* encoding a small guanosine triphosphate (GTP)-binding protein leads to an increased xylem width by 27-50% and thicker xylem fiber cell walls (Liu et al., 2021).

TABLE 1 Lignin, cellulose, and CWR content quantified in the WT and the three PtaPLATZ18-SRDX lines.

Line	WT	SRDX2	SRDX19	SRDX20
Lignin content (% CWR)	23.05 (± 1.78)	25.75 (± 0.95) *	28.95 (± 2.22) **	27.74 (± 1.41) ***
Cellulose content (% CWR)	37.28 (± 6.60)	45.49 (± 8.83)	38.12 (± 5.70)	39.31 (± 9.10)
CWR (% of wood powder)	85.14 (± 0.66)	85.42 (± 0.42)	86.16 (± 1.46)	85.14 (± 0.58)

At least three biological replicates and two technical replicates have been performed for each quantification. Data are mean ± standard deviation. *t*-test between the WT and the considered mutant line. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

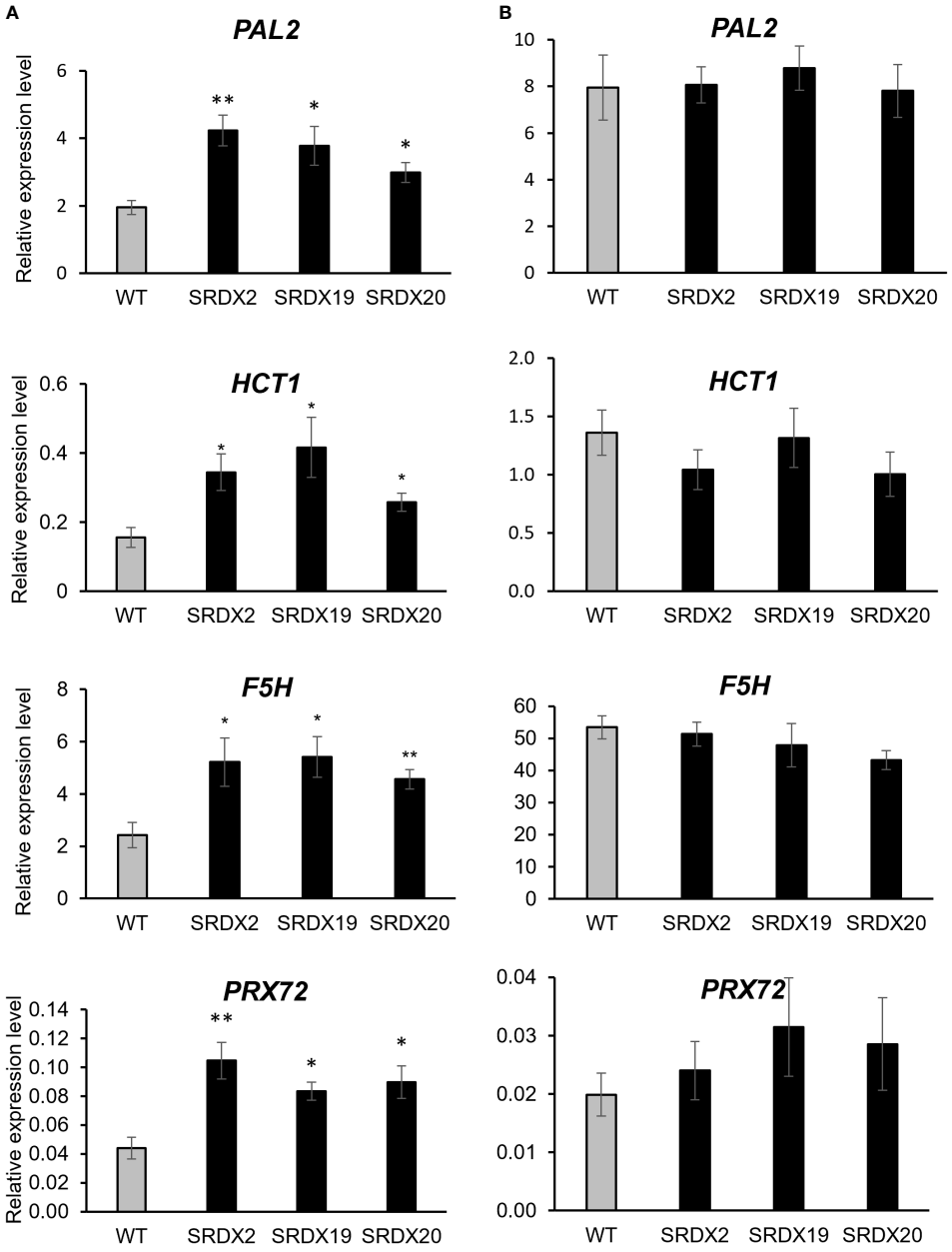


FIGURE 6
Relative expression level of genes involved in lignin biosynthesis and polymerization in 3-month-old WT and PtaPLATZ18-SRDX lines, as measured by RT-qPCR. **(A)** Young stem. **(B)** Xylem. Value for each line is an average of five plants (biological replicates) and three technical replicates, normalized by the *CDC2* reference gene. Bars = standard error. Student *t*-test was performed for each mutant line against the WT. *, $P < 0.05$; **, $P < 0.01$.

In another study, the overexpression of *PagSAG101a* induced increased plant height, as well as increased internode number and stem diameter, xylem width (by 5.1-14.7%), and secondary cell wall thickness while opposite phenotypes were observed for *PagSAG10a* knock-out plants (He et al., 2022). As shown by these authors, *PagSAG101a* gene expression is regulated by the PagC3H17 TF. The overexpression of *PagC3H17* (Chai et al., 2014) gave similar results than that of *PagSAG101a*, indicating that the PagC3H17-PagSAG101a module plays a role in the positive gene regulatory network of secondary vascular system development (He et al., 2022). Likewise, overexpression of *UNFERTILIZED EMBRYO SAC12 (UNE12)* encoding a basic helix-loop-helix (bHLH) TF promoted xylem development and increased the thickness of the cell walls and the lignin content (as analyzed by FTIR and Raman microspectrometry) of secondary xylem in poplar together with a reduced plant height, shorter internodes and curled leaves, when compared to the WT (Song et al., 2023).

At the biochemical level, dominant repression of *PtaPLATZ18* results in a higher lignin content as measured in the cell walls of the SRDX lines compared to the WT (Table 1). Compared to the WT, the expression of genes involved in lignin biosynthesis and polymerization of SRDX lines was higher in young stems, with

early developmental steps of secondary growth (upper portion of the stem), but not in the xylem from the stem with well-established secondary growth (lower portion of the stem) (Figure 6 and Supplementary Figure 6). These genes include *PAL2*, the first gene of the monolignol biosynthesis pathway (Wang et al., 2014), *HCT1*, encoding a central enzyme for the direction of the flux towards coniferyl and sinapyl alcohols (Hoffmann et al., 2003), *F5H1* encoding a key enzyme for the synthesis of sinapyl alcohol (Meyer et al., 1998) and a gene homologous to the *A. thaliana peroxidase 72* (*PRX72*), which has been shown to play a role in monolignol polymerization (Herrero et al., 2013). Therefore, the increased lignin accumulation in older xylem tissue of the SRDX lines seems to originate from lignin metabolism activation at earlier stages of secondary growth. Besides, the gene encoding VND6-C1, belonging to the VND6 family of NAC TFs which are master regulators of xylem vessel element differentiation, was upregulated in the SRDX lines (Supplementary Figure 7), suggesting that PtaPLATZ18 acts upstream of the complex NAC-MYB network regulating xylogenesis in both Arabidopsis and poplar (Kubo et al., 2005; Ohtani et al., 2011; Li et al., 2012). A more in-depth study through RNA-seq analysis would give a comprehensive overview of the transcriptional changes induced by the alteration of *PtaPLATZ18* expression.

Future research on the transcriptional activity of PtaPLATZ18, *i.e.* as an activator or as a repressor, on its protein interactions, and on the biological processes regulated by this TF are required to unravel its role in plant vegetative growth, secondary growth, and wood formation.

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. Nucleotide sequence of *PtaPLATZ18* can be found in the GenBank database under the accession number OR730833.

Author contributions

CG: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. MBe: Conceptualization, Investigation, Methodology, Supervision, Visualization, Writing – review & editing. JS: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. AM: Investigation, Writing – review & editing. ME: Supervision, Writing – review & editing. MBa: Conceptualization, Funding acquisition, Methodology, Project administration, 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. MBe and CG are supported by Belgian Fonds de la Recherche Scientifique (FRS-FNRS) research projects T.0068.18 and T.0010.22, respectively. MBa is a Senior Research Associate of the FRS-FNRS.

Acknowledgments

The authors acknowledge Ricardo Gameiro Da Costa, Nathanaël Sanne, and Aubry Fenestre for their technical assistance. Thanks to Louise Conrard and the CMMI team for the access to the wide-field microscope. A special thanks to Annabelle Dupas and Nathalie Ladouce from the LRSV team, for their collaboration in the wood powder preparation.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1302536/full#supplementary-material

References

Baldacci-Cresp, F., Moussawi, J., Leplé, J.-C., Van Acker, R., Kohler, A., Candiracci, J., et al. (2015). PtaRHE1, a *Populus tremula x Populus alba* RING-H2 protein of the ATL family, has a regulatory role in secondary phloem fibre development. *Plant J.* 82, 978–990. doi: 10.1111/tpj.12867. Article 6.

Baldacci-Cresp, F., Sacré, P.-Y., Twyffels, L., Mol, A., Vermeersch, M., Ziemons, E., et al. (2016). Poplar–root knot nematode interaction: A model for perennial woody species. *Mol. Plant Microbe Interact.* 29, 560–572. doi: 10.1094/MPMI-01-16-0015-R

Chai, G., Qi, G., Cao, Y., Wang, Z., Yu, L., Tang, X., et al. (2014). Poplar PdC3H17 and PdC3H18 are direct targets of PdMYB3 and PdMYB21, and positively regulate secondary wall formation in Arabidopsis and poplar. *New Phytol.* 203, 520–534. doi: 10.1111/nph.12825

Chao, Q., Gao, Z.-F., Zhang, D., Zhao, B.-G., Dong, F.-Q., Fu, C.-X., et al. (2019). The developmental dynamics of the *Populus* stem transcriptome. *Plant Biotechnol. J.* 17, 206–219. doi: 10.1111/pbi.12958

Cho, J.-S., Jeon, H.-W., Kim, M.-H., Vo, T. K., Kim, J., Park, E.-J., et al. (2019). Wood forming tissue-specific bicistronic expression of *PdGA20ox1* and *PtrMYB221* improves both the quality and quantity of woody biomass production in a hybrid poplar. *Plant Biotechnol. J.* 17, 1048–1057. doi: 10.1111/pbi.13036

Dampanaboina, L., Yuan, N., and Mendu, V. (2021). Estimation of crystalline cellulose content of plant biomass using the Updegraff method. *J. Vis. Exp.* 171, 62031. doi: 10.3791/62031

Davies, P. J. (2010). *Plant Hormones* (Dordrecht: Springer Netherlands). doi: 10.1007/978-1-4020-2686-7

Dong, T., Yin, X., Wang, H., Lu, P., Liu, X., Gong, C., et al. (2021). ABA-INDUCED expression 1 is involved in ABA-inhibited primary root elongation via modulating ROS homeostasis in *Arabidopsis*. *Plant Sci.* 304, 110821. doi: 10.1016/j.plantsci.2021.110821

Eriksson, M. E., Israelsson, M., Olsson, O., and Moritz, T. (2000). Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. *Nat. Biotech.* 18, 784–788. doi: 10.1038/77355

Gonzalez, N., Vanhaeren, H., and Inzé, D. (2012). Leaf size control: Complex coordination of cell division and expansion. *Trends Plant Sci.* 17, 332–340. doi: 10.1016/j.tplants.2012.02.003

González-Morales, S. I., Chávez-Montes, R. A., Hayano-Kanashiro, C., Alejo-Jacuinde, G., Rico-Cambron, T. Y., de Folter, S., et al. (2016). Regulatory network analysis reveals novel regulators of seed desiccation tolerance in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 113, E5232–E5241. doi: 10.1073/pnas.1610985113

Guerriero, G., Behr, M., Hausman, J.-F., and Legay, S. (2017). Textile hemp vs. Salinity: insights from a targeted gene expression analysis. *Genes* 8, 242. doi: 10.3390/genes8100242

Guo, X., Fu, Y., Lee, Y. J., Chern, M., Li, M., Cheng, M., et al. (2022). The PGS1 basic helix-loop-helix protein regulates *Fl3* to impact seed growth and grain yield in cereals. *Plant Biotechnol. J.* 20, 1311–1326. doi: 10.1111/pbi.13809

Han, X., Rong, H., Tian, Y., Qu, Y., Xu, M., and Xu, L. (2022). Genome-wide identification of PLATZ transcription factors in *Ginkgo biloba* L. and their expression characteristics during seed development. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.946194

He, H., Song, X.-Q., Jiang, C., Liu, Y.-L., Wang, D., Wen, S.-S., et al. (2022). The role of senescence-associated gene101 (*PagSAG101a*) in the regulation of secondary xylem formation in poplar. *J. Integr. Plant Biol.* 64, 73–86. doi: 10.1111/jipb.13195

Herrero, J., Fernández-Pérez, F., Yebra, T., Novo-Uzal, E., Pomar, F., Pedreño, M. Á., et al. (2013). Bioinformatic and functional characterization of the basic peroxidase 72 from *Arabidopsis thaliana* involved in lignin biosynthesis. *Planta* 237, 1599–1612. doi: 10.1007/s00425-013-1865-5

Hiratsu, K., Matsui, K., Koyama, T., and Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant J.* 34, 733–739. doi: 10.1046/j.1365-313X.2003.01759.x

Hoffmann, L., Maury, S., Martz, F., Geoffroy, P., and Legrand, M. (2003). Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J. Biol. Chem.* 278, 95–103. doi: 10.1074/jbc.M209362200

Howe, E. A., Sinha, R., Schlauch, D., and Quackenbush, J. (2011). RNA-Seq analysis in MeV. *Bioinform* 27, 3209–3210. doi: 10.1093/bioinformatics/btr490

Hu, Y., Liu, Y., Lu, L., Tao, J.-J., Cheng, T., Jin, M., et al. (2023). Global analysis of seed transcriptomes reveals a novel PLATZ regulator for seed size and weight control in soybean. *New Phytol.* 240, 2436–2454. doi: 10.1111/nph.19316

Iocco-Corena, P., Chaïb, J., Torregrosa, L., Mackenzie, D., Thomas, M. R., and Smith, H. M. (2021). VviPLATZ1 is a major factor that controls female flower morphology determination in grapevine. *Nat. Commun.* 12, 6995. doi: 10.1038/s41467-021-27259-8

Jun, S. E., Kim, J. H., Hwang, J. Y., Huynh Le, T. T., and Kim, G.-T. (2019). ORESARA15 acts synergistically with ANGUSTIFOLIA3 and separately from AINTEGUMENTA to promote cell proliferation during leaf growth. *Int. J. Mol. Sci.* 21, 241. doi: 10.3390/ijms21010241

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3

Kim, J. H., Kim, J., Jun, S. E., Park, S., Timilsina, R., Kwon, D. S., et al. (2018). ORESARA15, a PLATZ transcription factor, mediates leaf growth and senescence in *Arabidopsis*. *New Phytol.* 220, 609–623. doi: 10.1111/nph.15291

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., et al. (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.* 19, 1855–1860. doi: 10.1101/gad.1331305

Leple, J. C., Brasileiro, A. C. M., Michel, M. F., Delmotte, F., and Jouanin, L. (1992). Transgenic poplars: expression of chimeric genes using four different constructs. *Plant Cell Rep.* 11, 137–141. doi: 10.1007/BF00232166

Li, Q., Lin, Y.-C., Sun, Y.-H., Song, J., Chen, H., Zhang, X.-H., et al. (2012). Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14699–14704. doi: 10.1073/pnas.1212977109

Li, H., Wang, Y., Xiao, Q., Luo, L., Zhang, C., Mao, C., et al. (2021). Transcription factor ZmPLATZ2 positively regulate the starch synthesis in maize. *Plant Growth Regul.* 93, 291–302. doi: 10.1007/s10725-020-00687-0

Li, Q., Wang, J., Ye, J., Zheng, X., Xiang, X., Li, C., et al. (2017). The maize imprinted gene *Floury3* encodes a PLATZ protein required for tRNA and 5S rRNA transcription through interaction with RNA Polymerase III. *Plant Cell* 29, 2661–2675. doi: 10.1105/tpc.17.00576

Liu, M., Wang, C., Ji, Z., Lu, J., Zhang, L., Li, C., et al. (2023). Regulation of drought tolerance in *Arabidopsis* involves the PLATZ4 -mediated transcriptional repression of plasma membrane aquaporin *PIP2;8*. *Plant J.* 115, 434–451. doi: 10.1111/tpj.16235

Liu, Y.-L., Wang, L.-J., Li, Y., Guo, Y.-H., Cao, Y., and Zhao, S.-T. (2021). A small guanosine triphosphate binding protein PagRabE1b promotes xylem development in poplar. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.686024

Liu, S., Yang, R., Liu, M., Zhang, S., Yan, K., Yang, G., et al. (2020). PLATZ2 negatively regulates salt tolerance in Arabidopsis seedlings by directly suppressing the expression of the *CBL4*/*SOS3* and *CBL10*/*SCaBP8* genes. *J. Exp. Bot.* 71, 5589–5602. doi: 10.1093/jxb/eraa259

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

Loziuk, P., Hecht, E., and Muddiman, D. (2017). *N*-linked glycosite profiling and use of Skyline as a platform for characterization and relative quantification of glycans in differentiating xylem of *Populus trichocarpa*. *Anal. Bioanal. Chem.* 409, 487–497. doi: 10.1007/s00216-016-9776-5

Lu, F., Wang, C., Chen, M., Yue, F., and Ralph, J. (2021). A facile spectroscopic method for measuring lignin content in lignocellulosic biomass. *Green Chem.* 23, 5106–5112. doi: 10.1039/D1GC01507A. Article 14.

Ma, X., Yang, H., Bu, Y., Wu, X., Sun, N., Xiao, J., et al. (2023). Genome-wide identification of *PLATZ* genes related to cadmium tolerance in *Populus trichocarpa* and characterization of the role of *PtPLATZ3* in phytoremediation of cadmium. *Int. J. Biol. Macromol.* 228, 732–743. doi: 10.1016/j.ijbiomac.2022.12.176

Meyer, K., Shirley, A. M., Cusumano, J. C., Bell-Lelong, D. A., and Chapple, C. (1998). Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6619–6623. doi: 10.1073/pnas.95.12.6619

Muoki, R. C., Paul, A., Kumari, A., Singh, K., and Kumar, S. (2012). An improved protocol for the isolation of RNA from roots of tea (*Camellia sinensis* (L.) O. Kuntze). *Mol. Biotechnol.* 52, 82–88. doi: 10.1007/s12033-011-9476-5

Nagano, Y., Furuhashi, H., Inaba, T., and Sasaki, Y. (2001). A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. *Nucleic Acids Res.* 29, 4097–4105. doi: 10.1093/nar/29.20.4097

Ohtani, M., Nishikubo, N., Xu, B., Yamaguchi, M., Mitsuda, N., Goué, N., et al. (2011). A NAC domain protein family contributing to the regulation of wood formation in poplar. *Plant J.* 67, 499–512. doi: 10.1111/j.1365-313X.2011.04614.x

Payyavula, R. S., Badmi, R., Jawdy, S. S., Rodriguez, M. Jr., Gunter, L., and Sykes, R. W. (2022). Biomass formation and sugar release efficiency of *Populus* modified by altered expression of a NAC transcription factor. *Plant Direct* 6 (8), e419. doi: 10.1002/pld3.419

Pelayo, M. A., Morishita, F., Sawada, H., Matsushita, K., Iimura, H., He, Z., et al. (2023). AGAMOUS regulates various target genes via cell cycle–coupled H3K27me3 dilution in floral meristems and stamens. *Plant Cell* 35, 2821–2847. doi: 10.1093/plcell/koad123

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. Article 7.

So, H.-A., Choi, S. J., Chung, E., and Lee, J.-H. (2015). Molecular characterization of stress-inducible *PLATZ* gene from soybean (*Glycine max* L.). *Plant Omics* 8, 479–484. doi: 10.3316/informit.263226429190188

Song, C., Guo, Y., Shen, W., Yao, X., Xu, H., Zhao, Y., et al. (2023). *PagUNE12* encodes a basic helix-loop-helix transcription factor that regulates the development of secondary vascular tissue in poplar. *Plant Physiol.* 192, 1046–1062. doi: 10.1093/plphys/kiad152

Speer, E. O. (1987). A method of retaining phloroglucinol proof of lignin. *Stain Technol.* 62, 279–280. doi: 10.3109/10520298709108008

Sundell, D., Street, N. R., Kumar, M., Mellerowicz, E. J., Kucukoglu, M., Johnsson, C., et al. (2017). AspWood: high-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in *Populus tremula*. *Plant Cell* 29, 1585–1604. doi: 10.1105/tpc.17.00153

Timilsina, R., Kim, Y., Park, S., Park, H., Park, S.-J., Kim, J. H., et al. (2022). ORESARA 15, a PLATZ transcription factor, controls root meristem size through auxin and cytokinin signalling-related pathways. *J. Exp. Bot.* 73, 2511–2524. doi: 10.1093/jxb/erac050

Wai, A. H., Rahman, M. M., Waseem, M., Cho, L.-H., Naing, A. H., Jeon, J.-S., et al. (2022). Comprehensive genome-wide analysis and expression pattern profiling of

PLATZ gene family members in *Solanum Lycopersicum* L. under multiple abiotic stresses. *Plants* 11, 3112. doi: 10.3390/plants11223112

Wang, A., Hou, Q., Si, L., Huang, X., Luo, J., Lu, D., et al. (2019). The PLATZ transcription factor GL6 affects grain length and number in rice. *Plant Physiol.* 180, 2077–2090. doi: 10.1104/pp.18.01574

Wang, J., Ji, C., Li, Q., Zhou, Y., and Wu, Y. (2018). Genome-wide analysis of the plant-specific PLATZ proteins in maize and identification of their general role in interaction with RNA polymerase III complex. *BMC Plant Biol.* 18, 221. doi: 10.1186/s12870-018-1443-x

Wang, J. P., Naik, P. P., Chen, H.-C., Shi, R., Lin, C.-Y., Liu, J., et al. (2014). Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in *Populus trichocarpa*. *Plant Cell* 26 (3), 894–914. doi: 10.1105/tpc.113.120881

Xu, X., Backes, A., Legay, S., Berni, R., Faleri, C., Gatti, E., et al. (2019). Cell wall composition and transcriptomics in stem tissues of stinging nettle (*Urtica dioica* L.): Spotlight on a neglected fibre crop. *Plant Direct* 3 (8), e00151. doi: 10.1002/pld3.151

Yamashita, D., Kimura, S., Wada, M., and Takabe, K. (2016). Improved Mäule color reaction provides more detailed information on syringyl lignin distribution in hardwood. *J. Wood. Sci.* 62, 131–137. doi: 10.1007/s10086-016-1536-9

Zhang, J., Li, C., Zhang, W., Zhang, X., Mo, Y., Tranquilli, G. E., et al. (2023). Wheat plant height locus *RHT25* encodes a PLATZ transcription factor that interacts with DELLA (RHT1). *Proc. Natl. Acad. Sci. U.S.A.* 120, e2300203120. doi: 10.1073/pnas.2300203120

Zhao, H., Qin, Y., Xiao, Z., Li, Q., Yang, N., Pan, Z., et al. (2020). Loss of function of an RNA Polymerase III subunit lads to impaired maize kernel development. *Plant Physiol.* 184, 359–373. doi: 10.1104/pp.20.00502

Zhao, J., Zheng, L., Wei, J., Wang, Y., Chen, J., Zhou, Y., et al. (2022). The soybean PLATZ transcription factor GmPLATZ17 suppresses drought tolerance by interfering with stress-associated gene regulation of GmDREB5. *Crop J.* 10, 1014–1025. doi: 10.1016/j.cj.2022.03.009

Zhou, S., and Xue, H. (2020). The rice PLATZ protein SHORT GRAIN6 determines grain size by regulating spikelet hull cell division. *J. Integr. Plant Biol.* 62, 847–864. doi: 10.1111/jipb.12851

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 January 2024
DOI 10.3389/fpls.2023.1297235

OPEN ACCESS

EDITED BY
Yicun Chen,
Chinese Academy of Forestry, China

REVIEWED BY
Yuepeng Song,
Beijing Forestry University, China
Daisuke Todaka,
RIKEN Yokohama, Japan

*CORRESPONDENCE
Xiao Han
hanx2017@zafu.edu.cn
Zaikang Tong
zktong@zafu.edu.cn

†These authors share first authorship

RECEIVED 19 September 2023
ACCEPTED 18 December 2023
PUBLISHED 08 January 2024

CITATION
Yu J, Yin K, Liu Y, Li Y, Zhang J, Han X and Tong Z (2024) Co-expression network analysis reveals PbTGA4 and PbAPRR2 as core transcription factors of drought response in an important timber species *Phoebe bournei*.
Front. Plant Sci. 14:1297235.
doi: 10.3389/fpls.2023.1297235

Co-expression network analysis reveals PbTGA4 and PbAPRR2 as core transcription factors of drought response in an important timber species *Phoebe bournei*

Jinjian Yu†, Ke Yin†, Yan Liu, Yuhui Li, Junhong Zhang, Xiao Han* and Zaikang Tong*

State Key Laboratory of Subtropical Silviculture, College of Forestry and Biotechnology, Zhejiang A&F University, Hangzhou, China

Phoebe bournei is one of the main afforestation tree species in subtropical regions of China and is famous for its timber. Its distribution and growth are significantly impaired by water conditions. Thus, it is essential to understand the mechanism of the stress response in *P. bournei*. Here, we analyzed the phenotypic changes and transcriptomic rearrangement in the leaves and roots of *P. bournei* seedlings grown for 0 h, 1 h, 24 h, and 72 h under simulated drought conditions (10% PEG 6000). The results showed that drought stress inhibited plant photosynthesis and increased oxidoreductase activity and abscisic acid (ABA) accumulation. Spatio-temporal transcriptomic analysis identified 2836 and 3704 differentially expressed genes (DEGs) in leaves and roots, respectively. The responsive genes in different organs presented various expression profiles at different times. Gene co-expression network analysis identified two core transcription factors, *TGA4* and *APRR2*, from two modules that showed a strong positive correlation with ABA accumulation. Our study investigated the different responses of aboveground and belowground organs of *P. bournei* to drought stress and provides critical information for improving the drought resistance of this timber species.

KEYWORDS

***Phoebe bournei*, drought stress, abscisic acid, transcriptome, transcription factor, co-expression network**

Introduction

Woody plants are an abundant source of lignocellulosic biomass for bioenergy production (An et al., 2021). Trees, as sessile organisms, need to cope with constantly changing environments (Wang et al., 2017). Drought stress is one of the major environmental factors influencing the geographical distribution and biomass accumulation of trees (Zhu, 2016). The frequency and intensity of drought will continue to increase (Gupta et al., 2020). When water is insufficient, water from different tissues of the plant converges toward mature tissues, resulting in a lack of water in young tissues and retarded growth (Wakamori and Mineno, 2019). To reduce water loss under drought stress, plants reduce stomatal opening and transpiration rates, which restrict the entry of carbon dioxide into the leaves and inhibit photosynthesis to impair biomass accumulation (Zhang et al., 2009; Li et al., 2018). At the same time, the reduced transpiration rate also leads to the obstruction of mineral transport in plants, which further aggravates physiological and metabolic disorders in plant cells, leading to the destruction of the cell membrane structure and impaired cell growth (Lu et al., 2018; Lin et al., 2022). Cell elongation is most sensitive to drought, and a slight water deficit can cause a significant decrease in the growth rate, resulting in dwarf plants (Marmagne et al., 2020).

Roots and leaves are the most critical organs regulating the water balance between soil and plants (Li et al., 2021). Roots regulate water and nutrient uptake and are the first to sense water deficits (Hein et al., 2016). When plants experience drought conditions, a decrease in water transport from the roots to the aboveground parts through the xylem directly signals drought stress to the plant (Mansinhos et al., 2022). Roots can be traced to coordinated cell division, elongation, and differentiation events to enhance their ability to absorb water (Rellan-Alvarez et al., 2016). Drought also induces plant roots to produce small peptides and transmit them through the vascular system to leaves, driving the synthesis of ABA in leaves (Christmann and Grill, 2018). Leaves are the primary organs for transpiration, which is the primary mechanism of water loss and is monitored by stomates (Rodrigues et al., 2019). The regulation of stomatal movement by ABA is an important way for plants to regulate water balance under drought stress (Li et al., 2021). Moreover, inhibited photosynthesis and induced ABA in plants under drought stress will accumulate a large amount of reactive oxygen species (ROS) in cells, resulting in cell damage (Zhang et al., 2021). Plants defend against excessive ROS damage to cells by increasing the activity of antioxidant enzymes such as peroxidase (POD) and superoxide dismutase (SOD) (Wu et al., 2022).

ABA is a classic plant stress-related hormone, and at least two ABA-dependent signaling pathways are involved in response to water stress (Valliyodan and Nguyen, 2006; Guo et al., 2021). Previous research has shown that members of the MYB, MYC, bZIP, and B3 families are directly induced to express ABA (Yang et al., 2019). These genes compose a complicated regulatory network for the conduction of ABA signals. Transcriptomic sequencing provides technical support for the elucidation of this complex biological process. This approach is used to understand the expression response of genes under stress at the whole genome level. Transcriptomic sequencing is significant for constructing the stress-responsive transcriptional regulatory network (Wang et al., 2014). Based on this approach, we found that several genes encoding bHLH, WRKY, NAC, DREB, SnRK, protein phosphatases, protein kinases, and protein-related enzymes involved in phosphatidylinositol synthesis and metabolism are involved in the drought stress response (Sakuraba et al., 2015; Vaziriyeganeh et al., 2021). The elucidation of these genes and related regulatory networks contributes to breeding drought-resistant plant genotypes.

Phoebe bournei (Hemsl.) Yang is an important commercial and ornamental tree endemic to southern China. Its wood is regarded as "noble wood" due to its high corrosion resistance, special scent, and visible golden-tinted texture (Zhang J. H. et al., 2018). A recent survey of species distribution patterns showed that *P. bournei* is mainly distributed in the subtropics, with an annual average temperature of 15°C to 20°C and yearly precipitation of approximately 1505 mm (Ge et al., 2014). Frequent extreme climate conditions and water scarcity have reduced the suitable areas for *P. bournei*, making afforestation difficult (Li et al., 2022). Thus, it is essential to understand the regulatory mechanism of the stress response in *P. bournei*. However, current research on drought stress in *P. bournei* has focused on the patterns of changes in biomass, photosynthesis, and antioxidant enzyme activity (Ge et al., 2014; Tang et al., 2020). Little is known about the coordination of the belowground and aboveground parts of *P. bournei* trees and the corresponding molecular mechanisms involved in the stress response. In this study, changes in photosynthesis, antioxidant enzyme activity, ABA content, and gene expression profiling in the leaves and roots of *P. bournei* under different treatment periods of drought stress were systematically analyzed from a temporal and an organ-level perspective. Gene co-expression analysis was employed to identify hub genes and key modules in *P. bournei* in response to drought stress. This study may inform future studies on the potential mechanisms underlying *P. bournei* drought resistance and provide theoretical support for genetic engineering to breed drought-resistant *P. bournei* varieties.

Materials and methods

Plant material and PEG treatment

Seedlings of *P. bournei* cv. 'WY8' were collected from the plantation base at Zhejiang Agriculture and Forestry University in Zhejiang Province in China (30°15′N, 119°43′E). Plants with uniform growth were selected and hydroponically cultured in ¼-strength Hoagland's solution. Seedlings were cultured in a greenhouse at 25°C, 16 h/8 h (light/dark), and 60% humidity. The culture solution was renewed every four days. After one month of preculture, plants that showed consistent and vigorous growth were chosen for further experiments. PEG treatment was imposed on the plants by transplanting them into ¼-strength Hoagland's solution containing 10% PEG 6000 for 0 h, 1 h, 24 h, and 72 h. To avoid the effects of the exploitation process on the experimental results, the experiment end time and the sampling period were standardized

(Yang and Wei, 2015). Leaves and roots were collected from each group of ten individual plants and stored at -80 °C for physiological tests and RNA-Seq analysis.

Measurements of the physiological indices of *P. bournei*

In this study, the photosynthetic parameters of the seedlings were monitored using a Li-6400 Portable Photosynthesis System (LI-COR Biosciences, Lincoln, NE, USA), with a controlled reference CO_2 concentration of 400 μmol·mol^{-1} in the leaf chamber and photosynthetically active radiation set to 1200 μmol·mol^{-1}·s^{-1}. The measurements were carried out from 9:30 a.m. to 11:30 a.m. The photosynthetic parameters of 3–5 healthy and mature leaves of four groups of seedlings were automatically recorded by the experimental apparatus. The photosynthetic parameters included the net photosynthetic rate, transpiration rate, and stomatal conductance. Concurrently, the maximum photochemical efficiency (Fv/FM) of PSII after 30 minutes of dark adaptation was measured by a portable modulated chlorophyll fluorometer (PAM-2500, Heinz Walz GmbH, Germany). Superoxide dismutase (SOD) and peroxidase (POD) activities were determined with the corresponding kits (Nanjing Jincheng Bioengineering Institute, Nanjing, China). The ABA content was measured by MetWare (http://www.metware.cn/) based on the AB Sciex QTRAP 6500 LC-MS/MS platform (AB SCIEX, Foster City, CA, USA). Three replicates of each assay were performed. In this research, each physiological parameter was measured in triplicate.

RNA extraction, library preparation, and mRNA sequencing

Total RNA was extracted from 200 mg of ground leaf or root tissue with the M5 HiPer Plant RNeasy Mini Kit (Mei5bio, Beijing, China) following the manufacturer's protocol, with each tissue being analyzed in triplicate. The quantity and purity of RNA were determined with a NanoDrop 2000 (Thermo Fisher Scientific, New York, USA) and 1% agarose gel electrophoresis. The qualified RNA was used for library construction and transcriptomic sequencing. Eight treatments, L_0h (leaves of plants treated with 10% PEG for 0 h), L_1h (leaves of plants treated with 10% PEG for 1 h), L_24h (leaves of plants treated with 10% PEG for 24 h), L_72h (leaves of plants treated with 10% PEG for 72 h), R_0h (roots of plants treated with 10% PEG for 0 h), R_1h (roots of plants treated with 10% PEG for 1 h), R_24h (roots of plants treated with 10% PEG for 24 h), R_72h (roots of plants treated with 10% PEG for 72 h), totaling twenty-four cDNA libraries were constructed according to the manufacturer's instructions (Illumina, Inc., San Diego, CA, USA). Using oligo dT25 magnetic beads, 10 μg of total RNA was converted into mRNA, which was subsequently fragmented into small fragments. With random hexamer primers, these fragments were used to create first-strand complementary DNA (cDNA), followed by second-strand cDNA using a dUTP mixture. Magnetic beads were used to separate the double-stranded cDNA, then treated with three single adenylations and end repair. Following size selection using the adenylated fragments, sequencing adaptors were ligated to them and then enriched by PCR amplification. During the quality control phase, an Agilent 2100 Bioanalyzer was used to evaluate the quantity and quality of the sample library. Following the vendor's suggested methodology, paired-end 150-bp sequencing on an Illumina HiSeq2000 was carried out at LC Sciences (Houston, TX, USA). Transcriptomic sequencing was commissioned by LC-Bio Technologies Co., Ltd. (Hangzhou, China). Sequencing data have been uploaded to the National Genomics Data Center (NGDC) (Chen et al., 2021; Xue et al., 2022), and the assigned accession of the submission is CRA011591.

Sequence mapping and annotation and differentially expressed gene analysis

After the adapter sequences were trimmed and the low-quality reads were filtered, the valid data were mapped onto the *P. bournei* genome (Han et al., 2022) by Tophat2 v2.0.6. The transcript abundance estimation was processed by Cufflinks v2.0.2. Genomic locations and UMI tags were used to remove redundancy and perform calibration to obtain comprehensive transcript information. We utilized protein sequence similarity, the Kyoto Encyclopedia of Genes and Genomes (KEGG), cluster of orthologous groups of proteins (COG), and gene ontology (GO) analyses to perform the functional annotation of unigenes. We aligned all paired-end reads back to the final assembly using Perl scripts in Trinity to investigate the expression level of each unigene in different samples and expressed the abundance of unigenes in transcripts per million (TPM). The differentially expressed genes (DEGs) were screened based on $|\log2(\text{fold change})|>1$ and $P<0.05$. The GO term enrichment analysis of the DEGs was carried out using the GOseq R package and plotted using the GOplot R package. PlantTFDB (http://planttfdb.gao-lab.org/) was used to analyze transcription factors (TFs) in DEGs.

Weighted gene co-expression network analysis

The co-expression network and the correlation of modules and physiological data were analyzed by the R package WGCNA according to a previously described method (Langfelder and Horvath, 2008). The co-expression network was generated using genes with TPM > 1. We analyzed the correlation of module traits by combining the co-expression network, antioxidant enzyme activity, and ABA content. Key modules were screened through the correlation between modules and phenotypes. We selected the hub gene according to the module membership and gene significance of each gene in the key module. Cytoscape v3.7.2 was used to plot the co-expression network between the core TFs and co-expressed genes.

qRT-PCR validation

The genes of interest were validated by qRT-PCR. Primer3 software (http://frodo.wi.mit.edu/primer3/input.htm) was used to design primers with an output production of 150–300 bp (Supplementary Table 1). *PbUBQ4* and *PbEF1α* were used as internal controls. We utilized the SYBR Premix EX Taq Kit (TaKaRa, Dalian, China) for qRT-PCR amplification on the CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA). The $2^{-\triangle\triangle Ct}$ method was used to calculate the relative expression of the target gene, and qRT-PCR analysis was conducted with three replicates of each cDNA sample. The resulting data are presented as the average value and standard deviation.

Statistical analyses

SPSS22.0 (SPSS Inc., Chicago, IL, USA) was used for the statistical analysis of both physiological data and relative expression, with the significance of the difference between the data tested using One-way ANOVA and Duncan's multiple comparisons ($P<0.05$).

Results

Morphological and physiological evaluation of *P. bournei* in response to PEG

To better present the change in roots, hydroponic culture with PEG was employed for stimulated drought stress treatment. PEG-induced stress affected normal plant growth with browning roots (Figure 1A). The color gradually deepened with increasing treatment time. However, the leaves did not change significantly during the treatment.

Photosynthesis is an important indicator for evaluating the growth status of plants. Compared with the control, the photosynthetic parameters of *P. bournei* decreased to varying degrees after drought stress (Figure 1B). With increasing treatment time, the net photosynthetic and transpiration rates of *P. bournei* gradually decreased and were only 26.72% and 49.28%, respectively, of those of the control group after 72 h of treatment. The net photosynthetic and transpiration rates presented significant differences at each time point. The stomatal conductance of *P. bournei* leaves under stress conditions was not significantly different from that of the control at 1 h, while it decreased to 51.46% of that of the control at 72 h. The photochemical efficiency of PSII in the light (Fv/Fm) decreased significantly immediately after stress induction (1 h). However, there was no significant correlation between the stress duration and the Fv/Fm value.

Dehydration induces oxidative stress in plant cells, while plants themselves produce antioxidants such as POD and SOD to mitigate this damage. The changes in POD and SOD in *P. bournei* presented a similar pattern (Figure 1C). In the leaves, there were no significant differences in the POD and SOD activities between the control and treated groups after 1 h. However, the POD and SOD activities of the treated groups were significantly higher than those of the control group at 24 h and 72 h. The highest activities were observed at 24 h, with 52.22% and 42.46% increases in POD and SOD activities, respectively, compared to the control group. In the roots, the activities of POD and SOD gradually increased with treatment time, and the antioxidant enzyme activity was highest at 72 h.

Since ABA is a significant phytohormone that regulates the plant stress response, we monitored the levels of ABA in the leaves and roots of the stressed plants (Figure 1D). The ABA content in leaves gradually increased with increasing treatment time. The ABA concentration was much lower in roots than in leaves. The change in the ABA content in roots lagged behind that in leaves and significantly differed between the control and treated groups at 72 h.

Transcriptomic changes in *P. bournei* under PEG treatments

To investigate the molecular response of *P. bournei* to drought stress, we conducted a comparative analysis of transcriptomic changes in the leaves and roots under three time points. After rRNA and low-quality reads were filtered out, the clean reads ranged from 36.73 to 54.81 Mb across the libraries, with 71.06 to 89.53% of sequenced reads being accurately mapped to the *P. bournei* genome (Supplementary Table 2).

A total of 5688 DEGs (Supplementary Table 3) were identified, and the distribution analysis across the *P. bournei* chromosomes indicated that they were not significantly enriched in any specific chromosome region (Figure 2A). The number of DEGs in *P. bournei* was significantly positively correlated with stress duration. Significantly fewer DEGs were identified in the leaves (58, 1318, and 2141) than in the roots (466, 2376, and 2322) at the respective time points of 1 h, 24 h, and 72 h, indicating a slower stress response rate in leaves. The stress-responsive transcriptomic program was predominant in the roots.

We then compared the DEGs overlapping among different stress stages between *P. bournei* leaves and roots (Figure 2B). The results showed that the stress-responsive genes in the leaves and roots presented independent expression profiles. Only two genes, *Kiss Me Deadly 3* (*KMD3*), which controls phenylpropanoid biosynthesis, and *Arabidopsis Toxicos En Levadura 16* (*ATL16*), which encodes a RING/U-box superfamily protein, exhibited differential expression in both tissues at 1 h. As the duration of stress increased, the number of genes that acted in unison in leaves and roots also gradually increased. Functional annotation analysis revealed that the overlapping DEGs between leaves and roots were mainly involved in biological processes such as protein phosphorylation, oxidation-reduction processes, and carbohydrate metabolic processes.

The overlaps in each comparison were shown in a DiVenn diagram to identify identical DEGs in different stress stages (Figure 2C). The number of up-regulated and downregulated DEGs in the overlap of L_24h and L_72h was almost the same. A

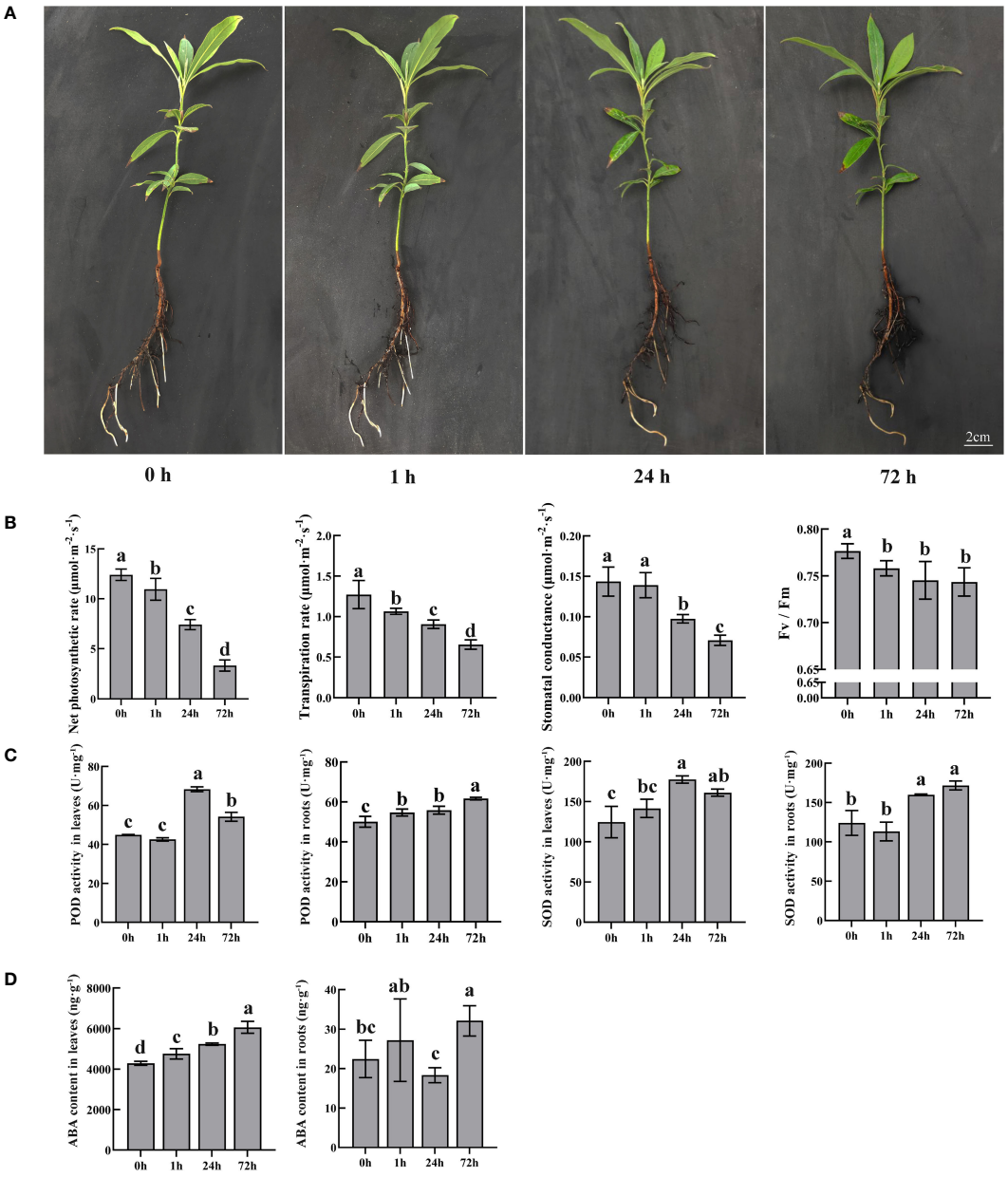


FIGURE 1
Morphological and physiological changes in *P. bournei* in response to PEG. **(A)** Morphological analysis of *P. bournei* under different PEG treatment times. The seedling was taken under the PEG treatment at 0, 1, 24, and 72 h. **(B)** Photosynthesis changes in *P. bournei* in response to PEG. **(C)** Oxidoreductase activity changes in *P. bournei* in response to PEG. **(D)** ABA content changes in *P. bournei* in response to PEG. Data are means and standard deviation calculated from three independent experiments. Different lowercase letters represent significant differences ($P<0.05$).

total of 502 up-regulated genes were enriched in L_24h. In contrast, the overlap of the DEGs in the roots was mainly downregulated. The highest overlap of the DEGs was between R_24h and R_72h (991). There were 855 downregulated genes, 126 up-regulated genes, and 10 genes with varying expression patterns. The overlap of the three conditions consisted of 135 DEGs in roots and 37 DEGs in leaves. Compared with leaves, there were more genes with changed expression trends in the three overlapping conditions of the roots. This result suggests that the responsive genes at different stress times are unique.

Gene ontology enrichment in DEGs

GO was utilized as a universal functional classification system to identify the primary molecular functions of the DEGs. The DEGs of leaves and roots were used separately for GO analysis and screen representative terms (Figure 3). We found that the DEGs were enriched in many biological processes associated with the drought-stress response. The response to stimulus (GO:0050896) was the most significantly enriched biological process (BP) term in the DEGs in both leaves and roots. The BP of DEGs in leaves mainly

10.3389/fpls.2023.1297235

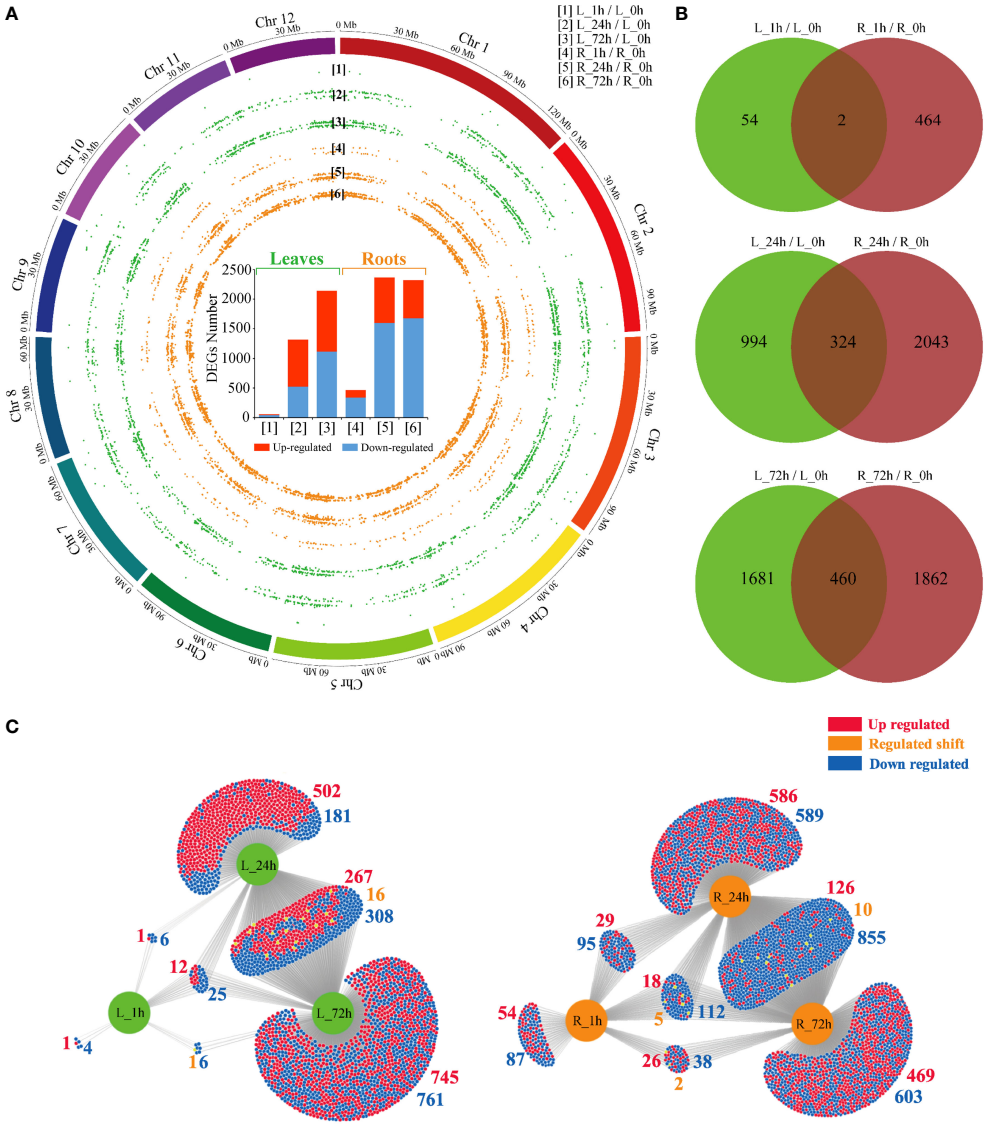


FIGURE 2
Overview of PEG treatment effects on *P. bournei* gene expression. **(A)** Statistical histogram of the DEGs and distribution in 12 P. *bournei* chromosomes from pairwise comparisons. **(B)** Venn diagram analysis of DEGs among different treatment periods between leaves and roots. **(C)** DiVenn diagram analysis of DEGs among different treatment periods in leaves and roots.

included photosynthesis (GO:0015979, GO:0019684, GO:0010119, and GO:0009768), the hormonal response (GO:0009737, GO:0009725, GO:0032870, GO:0009755, GO:0009751, GO:0009753, and GO:0010105) and carbohydrate metabolism (GO:0044262, GO:0044264, GO:0044042, GO:0044264, GO:0044262, and GO:0044264) (Figure 3A). Cellular components (CC), such as the chloroplast stroma (GO:0009570), photosynthetic membrane (GO:0034357), and photosystem (GO:0009521), were enriched in the DEGs in leaves. The DEGs in leaves were primarily involved in molecular functions (MFs) closely related to the drought stress response, such as oxidoreductase activity (GO:0016491), glutathione transferase activity (GO:0004364), and trehalose-phosphatase activity (GO:0004805). Unlike the GO-enriched terms in leaves, the DEGs in roots were mainly related to transport and signal transduction (Figure 3B). The terms cell communication (GO:0007154), signal transduction (GO:0007165), anion transport (GO:0006820), and ion transport (GO:0006811) all had a relatively high degree of enrichment. Among the CC of DEGs in roots, six terms were related to the membrane (GO:0005886, GO:0016020, GO:0016021, GO:0031226, GO:0009532, and GO:0005887), and three terms related to cell junctions (GO:0009506, GO:0005911, and GO:0030054) were highly enriched. In addition to the activities of carbohydrate synthase (GO:0008194, GO:0004805, and GO:0052692) related to drought stress, many MFs related to transporter activities were also enriched in the DEGs in roots.

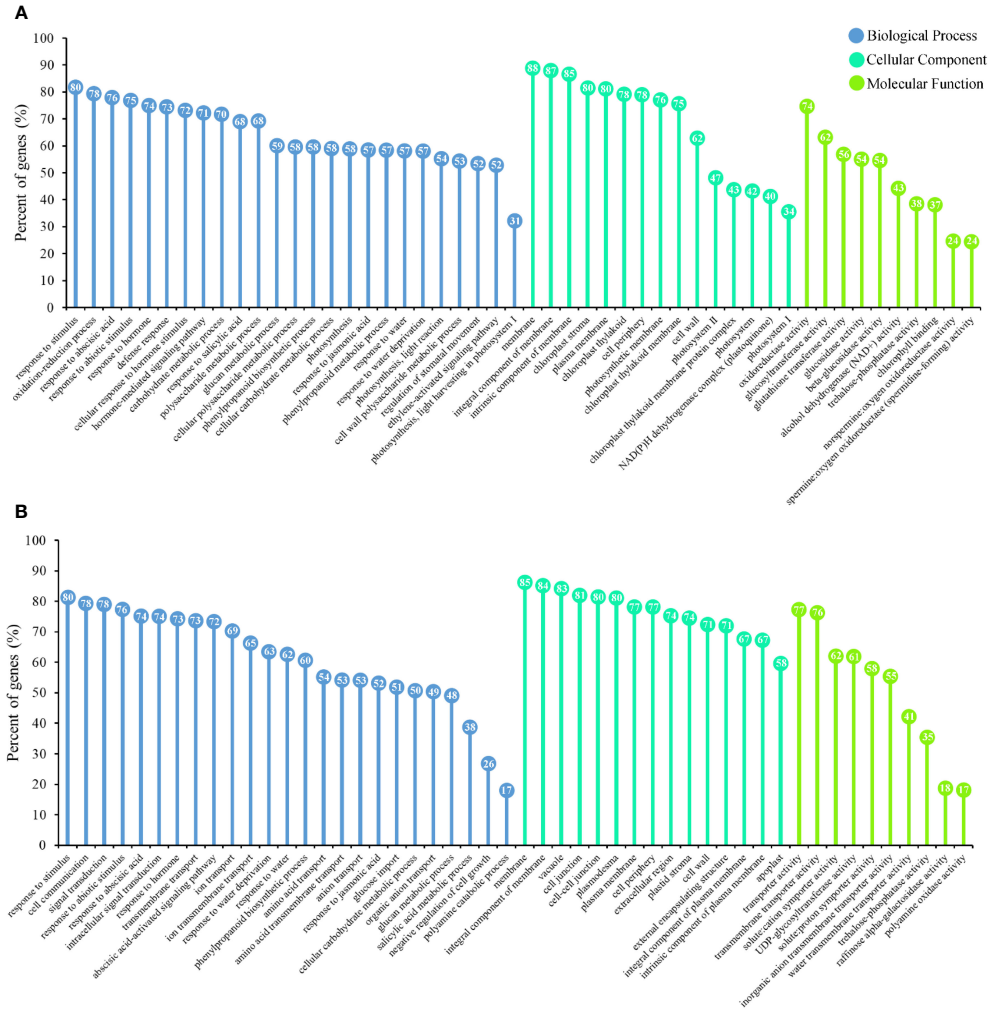


FIGURE 3
P. bournei DEGs functional classification by gene ontology (GO) analysis. **(A)** Enrichment lollipop diagram of GO terms assigned to DEGs in leaves. **(B)** Enrichment lollipop diagram of GO terms assigned to DEGs in roots.

Transcription factor identification

Drought stress significantly affected the expression of 447 TFs. The MYB, ERF, WRKY, bHLH, and NAC genes accounted for a significant proportion of *P. bournei* (Figure 4). In leaves, there were 207 TFs in 30 gene families, with MYB exhibiting the highest number of genes (Figure 4A). Compared to leaves, a greater diversity of transcription factor families was found in roots, with ERF exhibiting the highest number of genes. The BBR-BPC (Phbou.02G0563), CO-like (Phbou.05G2008, Phbou.03G2012, and Phbou.10G1237), SBP (Phbou.05G1740 and Phbou.05G1326) and SRS (Phbou.02G2683) family members were explicitly up-regulated in roots. The GRF (Phbou.01G2410), M-type MADS (Phbou.09G1428), and SAP (Phbou.03G1232) family members were explicitly up-regulated in leaves (Figure 4B). Every gene family of TFs had variable expression patterns in the different treatment periods and organs. Some exceptions, such as the B3 (Phbou.01G2687), C3H (Phbou.03G3182), HD-ZIP (Phbou.07G1018) and NAC (Phbou.07G0803, Phbou.03G0565) family members, had the same up-regulated pattern during all three treatment periods in both leaves and roots, and the TPM levels all exceeded twice those of the 0 h.

Weighted gene co-expression network analysis module generation and core TF analysis

We utilized WGCNA to analyze the gene co-expression patterns of *P. bournei* in response to drought to investigate the correlation between genes and physiological indices, as well as intramodular and intermodular genes. A total of 4584 DEGs (TPM>1) were analyzed, and 11 co-expression modules and correlation coefficients were identified and obtained (Figure 5A). The turquoise module was positively correlated with ABA, and the correlation coefficient was 0.86. The brown module has a significant positive correlation with SOD and ABA. Therefore, the two modules were selected to screen the hub genes and analyze the network.

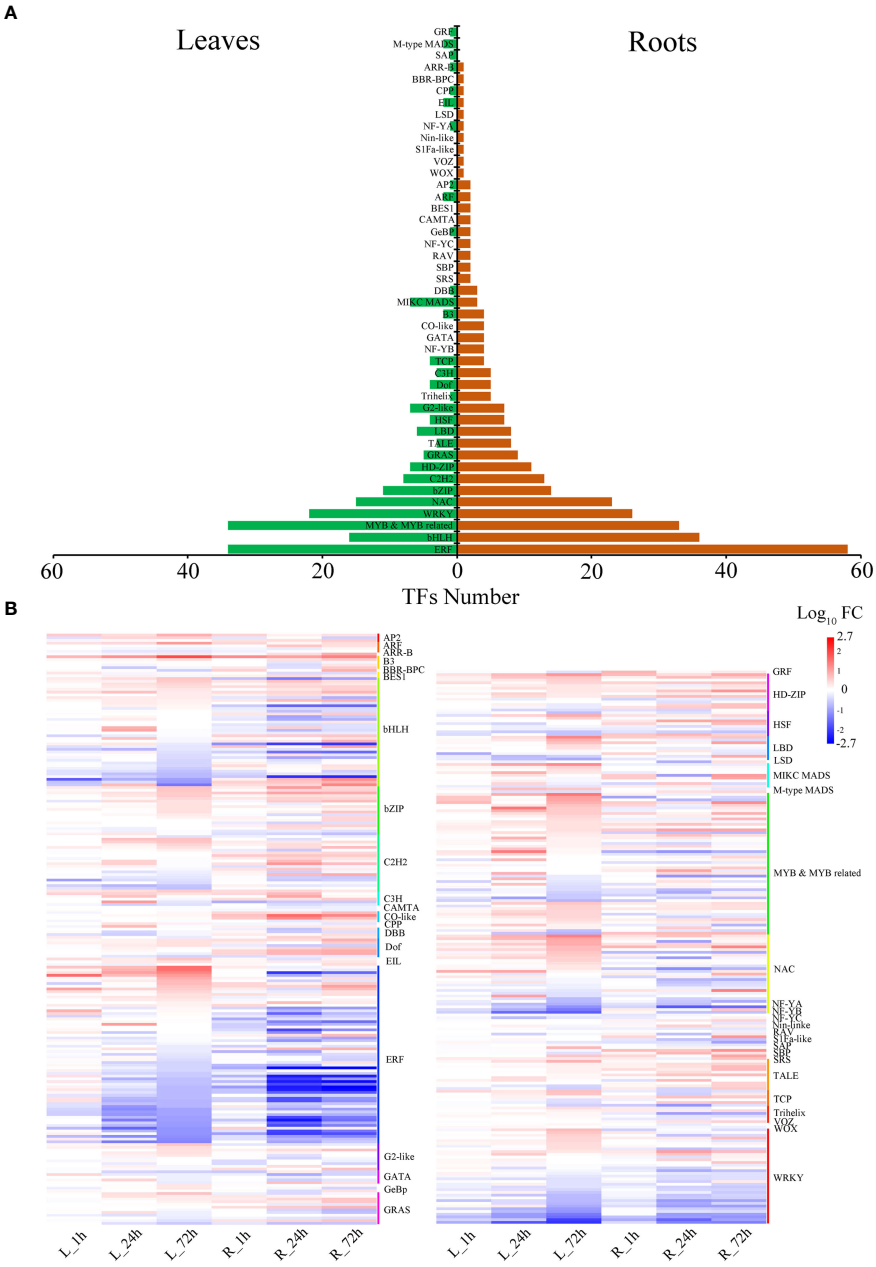


FIGURE 4
Distribution of transcription factors responsive to drought stress in *P. bournei*. **(A)** The number of predicted TFs in the leaves and roots. Each row in the figure corresponds to a TF gene family. The green represents the number of TFs in the leaves, and the brown represents the number of TFs in the roots. **(B)** Heatmap of the expression of all TFs at different treatment periods in the TF family with cluster assignments shown on the right. Heatmap indicates log 10 (fold change) of genes where red indicates up-regulated and blue indicates down-regulated genes.

For each module, TFs in this module were selected as prior regulators. After transcription factor-binding site (TFBS) scanning, 21 TFs belonging to 13 TF families had TFBSs of genes in the brown module, and 24 TFs belonging to 14 TF families had TFBSs of genes in the turquoise module. A total of 333 genes in the brown module all had TFBSs belonging to *TGACG Motif-Binding Factor 4* (*PbTGA4*) of the bZIP family, and 904 out of 905 genes in the turquoise module had TFBSs belonging to *Arabidopsis Pseudo-Response Regulator 2* (*PbAPRR2*) of the ARR-B family. These two TFs were selected as core TFs in each module (Figure 5B, Supplementary Table 4).

The relative expression of two core TFs was strongly induced by drought stress in plants (Figure 5C). The results of transcriptomic sequencing and qRT-PCR showed the same trend. This result further verified the model of the hub gene in response to drought and the reliability of the transcriptomic data.

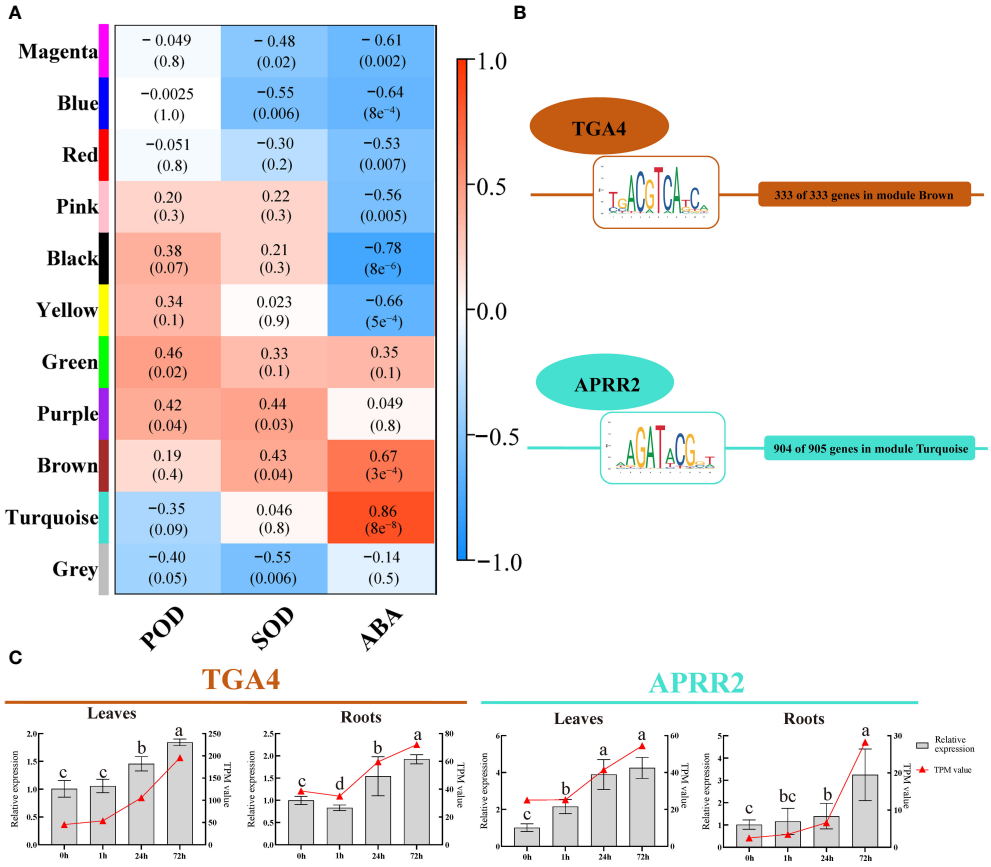


FIGURE 5
WGCNA and core TFs and their binding sites (TFBS) identification. **(A)** Module-Trait correlations. Each row in the table corresponds to a module and each column to a physiological index. The module name is shown on the left side of each cell. Numbers in the table represent the correlations between the corresponding module eigengenes and physiological indices, with p values listed below the correlations in parentheses. The table is color-coded to represent correlation strength according to the color legend. The intensity and direction of each correlation are indicated on the right side of the heatmap (The red indicates a positive correlation, and the blue indicates a negative correlation). **(B)** The most abundant TFBS and its corresponding TF in brown and turquoise modules. **(C)** Expression validation of the *PbTGA4* and *PbAPRR2* using qRT-PCR in 4 periods of leave and roots. Data are means and standard deviation calculated from three independent experiments. Different lowercase letters represent significant differences ($P<0.05$).

Core TF co-expression network analysis and functional enrichment analysis

Plants' response to drought stress is a collaborative process. Core TFs and their interconnected TFs and functional genes should be considered when regulating the plant response to drought. Based on two core TFs screened by WGCNA, we proposed a regulatory network of hub and edge genes closely related to them. In the brown module, *TGA4* showed a complex co-expression relationship with 330 genes in the module (Supplementary Table 5). Thirty-five genes from 17 TF families, such as the HD-ZIP and MIKC MADS families, showed a high co-expression relationship with *PbTGA4*. The functional genes co-expressed with *PbTGA4* mainly participated in the oxidation-reduction process (GO:0055114), transmembrane transport (GO:0055085), protein phosphorylation (GO:0006468), carbohydrate metabolic process (GO:0005975) and proteolysis (GO:0006508) (Figure 6A), such as *cytochrome P450-71* (*CYP71*, Phbou.010C003), *CYP76* (Phbou.08G1384), *CYP97* (Phbou.07G1158), *CYP707* (Phbou.09G1221), *aldehyde dehydrogenase 7B4* (*ALDH7B4*, Phbou.12G1648), *PIN-Formed 3* (*PIN3*, Phbou.10G1119), *polyol & monosaccharide transporter 1* (*PMT1*, Phbou.11G0533) and *open stomata 1* (*OST1*, Phbou.06G2001). In the turquoise module, *PbAPRR2* exhibited a high relationship with 380 edge genes. The TFs with a high weight value came from the bHLH family, C2H2 family, and WRKY family (Figure 6B, Supplementary Table 6). The genes co-expressed with *PbAPRR2* were mainly involved in the oxidation-reduction process (GO:0055114), protein phosphorylation (GO:0006468), photosynthesis (GO:0015979), proteolysis (GO:0006508) and cell redox homeostasis (GO:0045454), such as *CYP706* (Phbou.02G0977), *glutamate synthase 1* (*GLU1*, Phbou.12G1524), cysteine-rich *receptor-like protein kinase 10* (*CRK10*, Phbou.07G0003), *thioredoxin M-type 4* (*TRX-M4*, Phbou.02G3206) and *photosynthetic NDH subcomplex B4* (*PNSB4*, Phbou.03G3345). By conducting qPCR experiments on the high co-expression levels of transcription factors, we discovered that these genes demonstrated comparable expression patterns under stress conditions in both tissues (Supplementary Figure 1). This finding strengthens the credibility of the transcriptome sequencing outcomes.

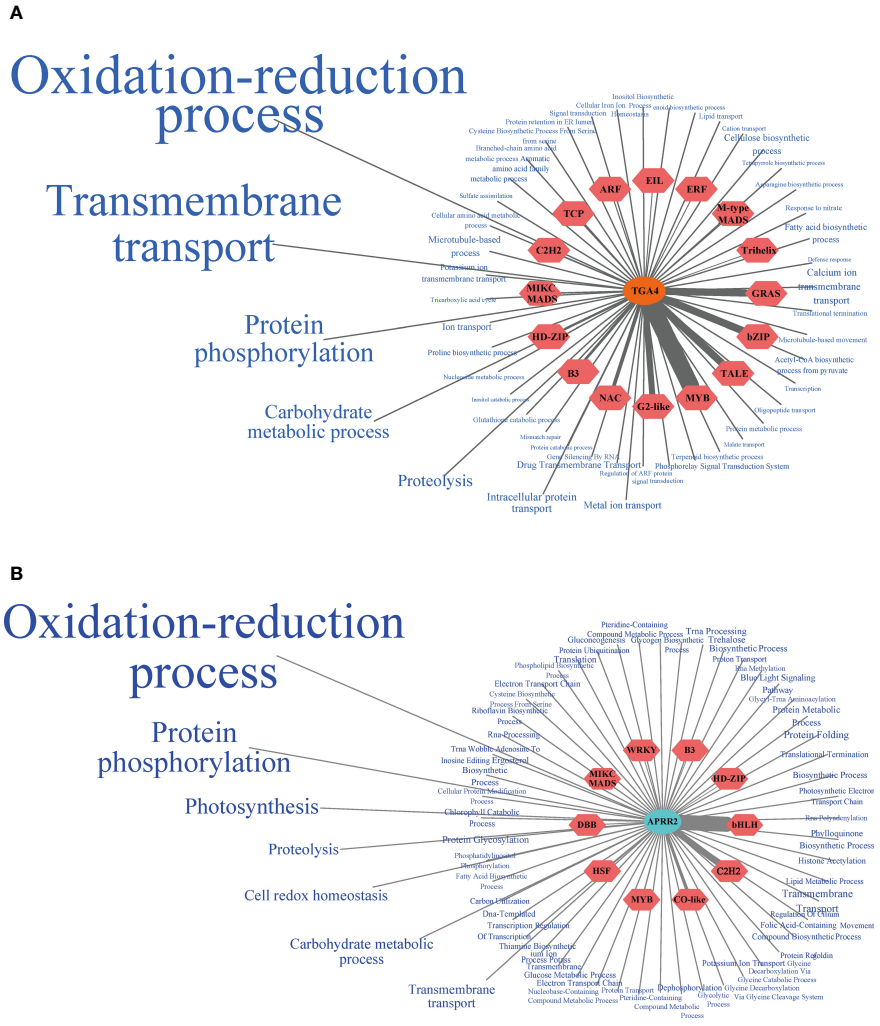


FIGURE 6
Core TFs co-expression network. **(A)** *PbTGA4* co-expression network of the brown module. **(B)** *PbAPARR2* co-expression network of the turquoise module. The TFs co-expressing with the core TF were located in the red hexagon, and the thickness of the line connection between them reflects the weight value. The blue font on the periphery was the biological process that co-expresses functional genes with core TF, and the font size reflects the weight value of the genes contained in each GO term and core TF.

Discussion

Distinguishable from the short life cycle of annual herbaceous plants to escape stress, perennial woody plants suffer prolonged and repeated stresses. The growth and biomass accumulation of woody plants is inhibited under water shortage, which becomes more evident with increased stress duration (Zlobin, 2022). Thus, trees evolve complex regulatory networks to adjust their growth patterns and biomass allocation for adaptation to water-limited environments (Jia et al., 2022). This study combined morphological, physiological, and transcriptomic changes to identify genes differentially expressed under a drought-mimicking treatment. The regulatory networks of *P. bournei* in response to PEG treatments were initially elucidated.

The inconsistent responses of morphological and physiological changes in *P. bournei* under short-term drought stress

Plants undergo physiological and morphological changes during drought stress (Fang et al., 2022). Extensive research has shown that drought stress increases ABA accumulation in leaves to induce stomatal closure and reduce water loss (Yu et al., 2021). Photosynthesis is impaired due to stomatal closure and limited carbon dioxide intake (Salomon et al., 2022). Drought induction significantly reduced the net photosynthetic rate, stomatal conductance, and transpiration rate in *P. bournei*. At the same time, GO enrichment analysis suggested an enrichment of

photosynthesis-related terms in the leaves under drought stress induction. The results are consistent with previous conclusions (Razi and Muneer, 2021). The Fv/Fm ratio reflects the function of the whole PSII, and its reduction indicates the damage or photoinhibition of PSII in response to environmental stress (Yan K. et al., 2022). *P. bournei* had a significantly lower Fv/Fm value than the control at 1 h of drought induction, indicating that 10% PEG caused drought stress.

Drought stress causes lipid peroxidation and cell membrane damage due to ROS accumulation. Plants prevent cellular damage from ROS accumulation by increasing the activities of SOD and POD (Yan W. P. et al., 2022). However, severe drought can lead to a decrease in the efficiency of the antioxidant defense system in plants, decreasing the activities of antioxidant enzymes such as SOD and POD (Chen et al., 2017). The SOD and POD activities of *P. bournei* leaves initially increased and then decreased in response to drought induction. At the same time, the activity of SOD in roots showed a positive correlation with the duration of drought stress. Furthermore, the activity of POD in roots was higher than in leaves. These results indicate that the initial increase in SOD and POD activities in leaves helps to resist drought-induced damage. However, prolonged drought conditions lead to a decrease in SOD and POD activities due to the decline in the efficiency of the antioxidant defense system.

ABA plays an essential role in the response to drought stress in plants (Xue et al., 2018). ABA alleviates the damage of drought stress caused to plants by closing stomata, mediating root architecture, and regulating stress-responsive gene expression (Gupta et al., 2020; Muhammad Aslam et al., 2022). In this study, the content of ABA in leaves was negatively correlated with photosynthesis, and the accumulation of ABA might be the main reason for the decrease in the stomatal conductance, transpiration rate, and net photosynthetic rate of *P. bournei*. The response to abscisic acid (GO:0009737), the cellular response to hormone stimulus (GO:0032870), and photosynthesis (GO:0015979) were enriched in leaves. A previous study showed that ABA is rapidly synthesized in leaves after plants are subjected to drought stress (Zhang F. P. et al., 2018). ABA synthesized in the leaves can be transported through the phloem to the roots, then enters the xylem and returns to the shoot (Yoshida et al., 2019). The ABA content in *P. bournei* roots exhibited a similar variation pattern to the above results. The biological process of response to abscisic acid (GO:0009737) was also enriched in roots. However, in contrast to leaves, signal transduction (GO:0007165) exhibits the highest enrichment in roots. We observed a significant increase in the expression level of genes in signal transduction, such as *sucrose non-fermenting 1-related protein kinase 3* (*SnRK3*, Phbou.03G0313), after 1 hour of treatment. *SnRK3* has a role as a component of calcium signaling, mediating plant development, and regulating responses to abiotic stress (Zhang et al., 2020). Therefore, we consider the genes involved in signal transduction to be crucial factors in coordinating the response of both aboveground and belowground plant parts to drought stress.

Unlike the solid physiological and molecular responses to drought stress, the leaves of *P. bournei* showed no significant morphological changes during the treatment period. Previous research showed different morphological responses to drought stress between leaves of evergreen and deciduous species (De Oliveira et al., 2015). To reduce the negative effects of drought stress and diminish variations in water potential, evergreen species adopt a water-use strategy that is highly conservative to maintain leaf shape (De Souza et al., 2020). In comparison, deciduous species tend to prioritize their water potential by facilitating leaf wilting and detachment to ensure their survival during a drought period (Lima et al., 2018). *P. bournei* is a typical evergreen species with an epicuticular waxy layer covering leaves that enhances drought stress tolerance (Al-Abdallat et al., 2014; Han et al., 2022). Transcriptomic analysis showed that the response speed of the differential expression of genes in leaves was slower than that in roots. In summary, the lack of significant morphological changes in the leaves of *P. bournei* during drought treatment might be attributed to the characteristics of evergreen tree species, such as the wax layer of leaves and the relatively slow response of gene expression. Due to this characteristic, it may not be possible to determine promptly whether *P. bournei* experienced drought stress only by observing changes in the aboveground parts during seedling cultivation and afforestation.

ABA-related core TFs that were differentially regulated in response to drought stress

To harvest the potential biological significance from the abundant DEGs, categorizing these genes into modules using co-expression profiling could help us better identify key genes involved in the response to drought stress. We identified two core TFs, *PbTGA4* and *PbAPRR2*, from the ABA-related brown and turquoise modules, respectively. *PbTGA4* exhibited a co-expression relationship with many CYP family member genes involved in the oxidation-reduction process and *ALDH7B4* in the co-expression network. *TGA* transcription factors belong to the D subfamily of the basic region-leucine zipper (bZIP) family. As a TF directly regulated by ABA, *TGA* plays a significant role in various abiotic stresses and signal transduction pathways (Sakuraba et al., 2015). *TGA* transcription factor family members regulate gene expression and participate in the plant stress response by binding to TGACG motifs on DNA promoters (Ju et al., 2012; Van Der Does et al., 2013). In addition, they can also interact with signaling transduction proteins in the ABA and jasmonic acid (JA) pathways and regulate their activity, thus participating in plant resistance to drought stress (Zander et al., 2010). Cytochrome P450 is an important class of enzymes present in plants. Under drought stress, cytochrome P450 may regulate the synthesis and decomposition of antioxidants in plant cells, thus regulating the oxidative/reductive state in cells to reduce ROS accumulation and promote plant adaptation to drought stress (Zhou et al., 2023). Plants will increase the expression level of *ALDH* to prevent the accumulation of aldehyde substances that may cause damage in plant cells after being subjected to abiotic stress (Hou and Bartels, 2015). This result suggested that *PbTGA4* might regulate genes with oxidation-reduction process functions to alleviate ROS accumulation and the toxicity of aldehyde substances caused by drought stress in *P. bournei*.

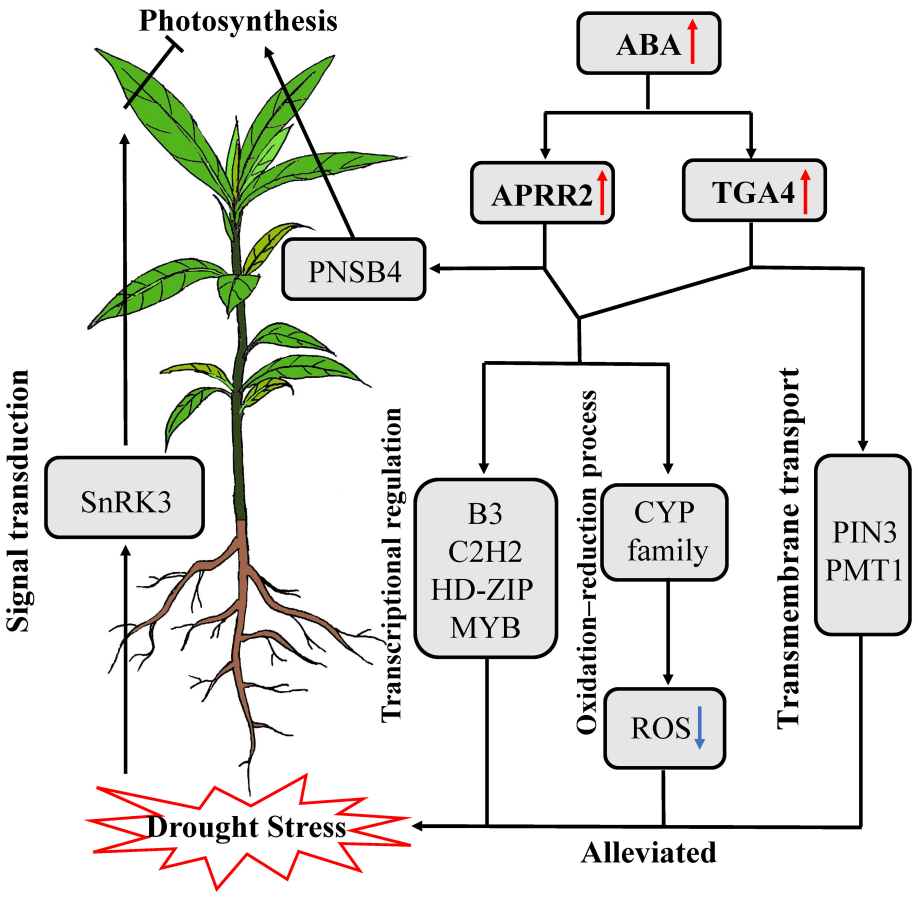


FIGURE 7
The model for the response to short-term drought stress in *P. bournei*.

APRR2 is a transcription factor involved in phytochrome synthesis (Yang et al., 2015). Apart from co-expression with the CYP family in *P. bournei* under drought stress, it was observed in this study that *PbAPRR2* also showed a significant degree of co-expression with amino acid synthesis and protein kinase-related functional genes such as *GLU1*, *CRK10*, and *TRX-M4*. *MdAPR2* promoted chlorophyll synthesis by activating the expression of chlorophyll synthesis-related genes to alleviate stress damage to apple (*Malus domestica*) leaves (Wen et al., 2022). *APRR2* has been reported to regulate the synthesis of flavonoids and antioxidant enzymes, which can alleviate oxidative stress induced by drought stress in plants (Perochon et al., 2010). Additionally, *APRR2* participates in the ABA-dependent response to abiotic stress by interacting with transcription factors (Kurup et al., 2000; Wen et al., 2022). This result indicated that *PbAPRR2* carries out a complex regulatory function in the response of plants to drought stress.

Potential molecular mechanism of the response to drought stress in *P. bournei*

Based on current research, we proposed a model to explain how *P. bournei* alleviates drought stress (Figure 7). When *P. bournei* roots sense water deficiency, they transmit the stress signal to aboveground tissues by activating signal transduction genes such as *SnRK3*. Upon receiving the signal, *P. bournei* leaves synthesize ABA to regulate stomatal closure to maintain water potential, but this inhibits photosynthesis and causes the accumulation of ROS. ABA accumulation activates the downstream transcription factors *PbAPRR2* and *PbTGA4*. *PbAPRR2* and *PbTGA4* alleviate ROS damage to cells by regulating genes involved in oxidation-reduction processes in *P. bournei*. The upregulation of core TFs influences other TFs and functional genes, inducing physiological and biochemical changes in *P. bournei* to enhance drought stress tolerance. However, in the plant response to drought stress, *PbAPRR2* correlated more with photosynthesis genes, while *PbTGA4* was more inclined to be co-expressed with transmembrane transport genes. This result indicated that *PbAPRR2* and *PbTGA4* also involve different response pathways to help *P. bournei* resist drought stress.

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

JY: Data curation, Investigation, Methodology, Writing – original draft. KY: Data curation, Investigation, Methodology, Writing – original draft. YL: Investigation, Methodology, Writing – original draft. YHL: Investigation, Methodology, Writing – original draft. JZ: Investigation, Writing – review & editing. XH: Data curation, Supervision, Writing – review & editing. ZT: Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We acknowledge support from the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (Grant No. 2021C02070-10), the National Natural Science Foundation of China (Grant No. 32101545 and No. 32171828), and the research foundation of Zhejiang A&F University (2022LFR113).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1297235/full#supplementary-material

References

Al-Abdallat, A. M., Al-Debei, H. S., Ayad, J. Y., and Hasan, S. (2014). Over-expression of *SlSHN1* gene improves drought tolerance by increasing cuticular wax accumulation in tomato. *Int. J. Mol. Sci.* 15, 19499–19515. doi: 10.3390/ijms151119499

An, Y., Liu, Y., Liu, Y. J., Lu, M. Z., Kang, X. H., Mansfield, S. D., et al. (2021). Opportunities and barriers for biofuel and bioenergy production from poplar. *Global Change Biol. Bioenergy* 13, 905–913. doi: 10.1111/gcbb.12829

Chen, T. T., Chen, X., Zhang, S. S., Zhu, J. W., Tang, B. X., Wang, A. K., et al. (2021). The Genome Sequence Archive Family: Toward explosive data growth and diverse data types. *Genomics Proteomics Bioinf.* 19, 578–583. doi: 10.1016/j.gpb.2021.08.001

Chen, Y., Wang, M., Hu, L. L., Liao, W. B., Dawuda, M. M., and Li, C. L. (2017). Carbon monoxide is involved in hydrogen gas-induced adventitious root development in cucumber under simulated drought stress. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00128

Christmann, A., and Grill, E. (2018). Peptide signal alerts plants to drought. *Nature* 556, 178–179. doi: 10.1038/d41586-018-03872-4

De Oliveira, C. C., Zandavalli, R. B., De Lima, A. L. A., and Rodal, M. J. N. (2015). Functional groups of woody species in semi-arid regions at low latitudes. *Austral Ecol.* 40, 40–49. doi: 10.1111/aec.12165

De Souza, B. C., Carvalho, E. C. D., Oliveira, R. S., De Araujo, F. S., De Lima, A. L. A., and Rodal, M. J. N. (2020). Drought response strategies of deciduous and evergreen woody species in a seasonally dry neotropical forest. *Oecologia* 194, 221–236. doi: 10.1007/s00442-020-04760-3

Fang, X., Ma, J., Guo, F. C., Qi, D. Y., Zhao, M., Zhang, C. Z., et al. (2022). The AP2/ERF GmERF113 positively regulates the drought response by activating *GmPR10-1* in soybean. *Int. J. Mol. Sci.* 23, 8159. doi: 10.3390/ijms23158159

Ge, Y. J., He, X. Y., Wang, J. F., Jiang, B., Ye, R. H., and Lin, X. C. (2014). Physiological and biochemical responses of *Phoebe bournei* seedlings to water stress and recovery. *Acta Physiol. Plant.* 36, 1241–1250. doi: 10.1007/s11738-014-1502-3

Guo, Q., Li, X., Niu, L., Jameson, P. E., and Zhou, W. (2021). Transcription-associated metabolomic adjustments in maize occur during combined drought and cold stress. *Plant Physiol.* 186, 677–695. doi: 10.1093/plphys/kiab050

Gupta, A., Rico-Medina, A., and Cano-Delgado, A. I. (2020). The physiology of plant responses to drought. *Science* 368, 266–269. doi: 10.1126/science.aaz7614

Han, X., Zhang, J. H., Han, S., Li Chong, S., Meng, G. L., Song, M. Y., et al. (2022). The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. *Plant Commun.* 3, 100410. doi: 10.1016/j.xplc.2022.100410

Hein, J. A., Sherrard, M. E., Manfredi, K. P., and Abebe, T. (2016). The fifth leaf and spike organs of barley (*Hordeum vulgare* L.) display different physiological and metabolic responses to drought stress. *BMC Plant Biol.* 16, 248. doi: 10.1186/s12870-016-0922-1

Hou, Q. C., and Bartels, D. (2015). Comparative study of the aldehyde dehydrogenase (*ALDH*) gene superfamily in the glycophyte *Arabidopsis thaliana* and *Eutrema* halophytes. *Ann. Bot.* 115, 465–479. doi: 10.1093/aob/mcu152

Jia, Y. Q., Niu, Y. I., Zhao, H. M., Wang, Z. B., Gao, C. Q., Wang, C., et al. (2022). Hierarchical transcription factor and regulatory network for drought response in *Betula platyphylla*. *Hortic. Res.* 9, uhac040. doi: 10.1093/hr/uhac040

Ju, C., Yoon, G. M., Shemansky, J. M., Lin, D. Y., Ying, Z. I., Chang, J., et al. (2012). CTR1 phosphorylates the central regulator EIN2 to control ethylene hormone signaling from the ER membrane to the nucleus in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19486–19491. doi: 10.1073/pnas.1214848109

Kurup, S., Jones, H. D., and Holdsworth, M. J. (2000). Interactions of the developmental regulator ABI3 with proteins identified from developing *Arabidopsis* seeds. *Plant J.* 21, 143–155. doi: 10.1046/j.1365-313x.2000.00663.x

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

Li, Q. F., Wang, J. D., Xiong, M., Wei, K., Zhou, P., Huang, L. C., et al. (2018). iTRAQ-Based analysis of proteins co-regulated by brassinosteroids and gibberellins in rice embryos during seed germination. *Int. J. Mol. Sci.* 19, 3460. doi: 10.3390/ijms19113460

Li, X., Liu, L. L., Sun, S. X., Li, Y. M., Jia, L., Ye, S. L., et al. (2022). Leaf-transcriptome profiles of *phoebe bournei* provide insights into temporal drought stress responses. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1010314

Li, Y., Sun, A., Wu, Q., Zou, X., Chen, F., Cai, R., et al. (2021). Comprehensive genomic survey, structural classification and expression analysis of C(2)H(2)-type zinc finger factor in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 21, 380. doi: 10.1186/s12870-021-03016-3

Lima, T. R. A., Carvalho, E. C. D., Martins, F. R., Oliveira, R. S., Miranda, R. S., Muller, C. S., et al. (2018). Lignin composition is related to xylem embolism resistance and leaf life span in trees in a tropical semiarid climate. *New Phytol.* 219, 1252–1262. doi: 10.1111/nph.15211

Lin, L., Wang, J., Wang, Q., Ji, M., Hong, S., Shang, L., et al. (2022). Transcriptome approach reveals the response mechanism of *Heimia myrtifolia* (Lythraceae, Myrtales) to drought stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.877913

Lu, X., Vitousek, P. M., Mao, Q., Gilliam, F. S., Luo, Y., Zhou, G., et al. (2018). Plant acclimation to long-term high nitrogen deposition in an N-rich tropical forest. *Proc. Natl. Acad. Sci. U.S.A.* 115, 5187–5192. doi: 10.1073/pnas.1720777115

Mansinhos, I., Goncalves, S., Rodriguez-Solana, R., Duarte, H., Ordonez-Diaz, J. L., Moreno-Rojas, J. M., et al. (2022). Response of *Thymus lotocephalus in vitro* cultures to drought stress and role of green extracts in cosmetics. *Antioxidants* 11, 1475. doi: 10.3390/antiox11081475

Marmagne, A., Jasinski, S., Fagard, M., Bill, L., Guerche, P., Masclaux-Daubresse, C., et al. (2020). Post-flowering biotic and abiotic stresses impact nitrogen use efficiency and seed filling in *Arabidopsis thaliana*. *J. Exp. Bot.* 71, 4578–4590. doi: 10.1093/jxb/eraa011

Muhammad Aslam, M., Waseem, M., Jakada, B. H., Okal, E. J., Lei, Z., Saqib, H. S. A., et al. (2022). Mechanisms of abscisic acid-mediated drought stress responses in plants. *Int. J. Mol. Sci.* 23, 1084. doi: 10.3390/ijms23031084

Perochon, A., Dieterle, S., Pouzet, C., Aldon, D., Galaud, J. P., and Ranty, B. (2010). Interaction of a plant pseudo-response regulator with a calmodulin-like protein. *Biochem. Biophys. Res. Commun.* 398, 747–751. doi: 10.1016/j.bbrc.2010.07.016

Razi, K., and Muneer, S. (2021). Drought stress-induced physiological mechanisms, signaling pathways and molecular response of chloroplasts in common vegetable crops. *Crit. Rev. Biotechnol.* 41, 669–691. doi: 10.1080/07388551.2021.1874280

Rellan-Alvarez, R., Lobet, G., and Dinneny, J. R. (2016). Environmental control of root system biology. *Annu. Rev. Plant Biol.* 67, 619–642. doi: 10.1146/annurev-arplant-043015-111848

Rodrigues, J., Inze, D., Nelissen, H., and Saibo, N. J. M. (2019). Source-sink regulation in crops under water deficit. *Trends Plant Sci.* 24, 652–663. doi: 10.1016/j.tplants.2019.04.005

Sakuraba, Y., Kim, Y. S., Han, S. H., Lee, B. D., and Paek, N. C. (2015). The *Arabidopsis* transcription factor *NAC016* promotes drought stress responses by repressing *AREB1* transcription through a trifurcate feed-forward regulatory loop involving NAP. *Plant Cell* 27, 1771–1787. doi: 10.1105/tpc.15.00222

Salomon, R. L., Peters, R. L., Zweifel, R., Sass-Klaassen, U. G. W., Stegehuis, A. I., and Smiljanic, M. (2022). The 2018 European heatwave led to stem dehydration but not to consistent growth reductions in forests. *Nat. Commun.* 13, 28. doi: 10.1038/s41467-021-27579-9

Tang, X. L., Liu, G. Z., Jiang, J., Lei, C. J., Zhang, Y. X., Wang, L. Y., et al. (2020). Effects of growth irradiance on photosynthesis and photorespiration of *Phoebe bournei* leaves. *Funct. Plant Biol.* 47, 1053–1061. doi: 10.1071/Fp20062

Valliyodan, B., and Nguyen, H. T. (2006). Understanding regulatory networks and engineering for enhanced drought tolerance in plants. *Curr. Opin. Plant Biol.* 9, 189–195. doi: 10.1016/j.pbi.2006.01.019

Van Der Does, D., Leon-Reyes, A., Koornneef, A., Van Verk, M. C., Rodenburg, N., Pauwels, L., et al. (2013). Salicylic acid suppresses jasmonic acid signaling downstream of SCFCOI1-JAZ by targeting GCC promoter motifs via transcription factor *ORA59*. *Plant Cell* 25, 744–761. doi: 10.1105/tpc.112.108548

Vaziriyeganeh, M., Khan, S., and Zwiazek, J. J. (2021). Transcriptome and metabolome analyses reveal potential salt tolerance mechanisms contributing to maintenance of water balance by the halophytic grass *Puccinellia nuttalliana*. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.760863

Wakamori, K., and Mineno, H. (2019). Optical flow-based analysis of the relationships between leaf wilting and stem diameter variations in tomato plants. *Plant Phenomics* 2019, 9136298. doi: 10.34133/2019/9136298

Wang, G., Weng, L., Li, M., and Xiao, H. (2017). Response of gene expression and alternative splicing to distinct growth environments in tomato. *Int. J. Mol. Sci.* 18, 475. doi: 10.3390/ijms18030475

Wang, J., Yu, H. H., Weng, X. Y., Xie, W. B., Xu, C. G., Li, X. H., et al. (2014). An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population. *J. Exp. Bot.* 65, 1069–1079. doi: 10.1093/jxb/ert464

Wen, B., Gong, X., Tan, Q., Zhao, W., Chen, X., Li, D., et al. (2022). *MdNAC4* Interacts with *MdAPRR2* to regulate nitrogen deficiency-induced leaf senescence in apple (*Malus domestica*). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.925035

Wu, Q., Yang, L. Y., Liang, H. Y., Yin, L., Chen, D. X., and Shen, P. (2022). Integrated analyses reveal the response of peanut to phosphorus deficiency on phenotype, transcriptome and metabolome. *BMC Plant Biol.* 22, 524. doi: 10.1186/s12870-022-03867-4

Xue, Y. B., Bao, Y. M., Zhang, Z., Zhao, W. M., Xiao, J. F., He, S. M., et al. (2022). Database resources of the National Genomics Data Center, China National Center for bioinformation in 2022. *Nucleic Acids Res.* 50, D27–D38. doi: 10.1093/nar/gkab951

Xue, Y., Yang, Y., Yang, Z., Wang, X., and Guo, Y. (2018). *VAMP711* is required for abscisic acid-mediated inhibition of plasma membrane H(+)-ATPase activity. *Plant Physiol.* 178, 1332–1343. doi: 10.1104/pp.18.00499

Yan, W. P., Cao, S. J., Wu, Y. G., Ye, Z. C., Zhang, C., Yao, G. L., et al. (2022). Integrated analysis of physiological, mRNA sequencing, and miRNA sequencing data reveals a specific mechanism for the response to continuous cropping obstacles in *Pogostemon cablin* roots. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.853110

Yan, K., Mei, H. M., Dong, X. Y., Zhou, S. W., Cui, J. X., and Sun, Y. H. (2022). Dissecting photosynthetic electron transport and photosystems performance in Jerusalem artichoke (*Helianthus tuberosus* L.) under salt stress. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.905100

Yang, Y., Chen, X., Xu, B., Li, Y., Ma, Y., and Wang, G. (2015). Phenotype and transcriptome analysis reveals chloroplast development and pigment biosynthesis together influenced the leaf color formation in mutants of *Anthurium andraeanum* 'Sonate'. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00139

Yang, X. H., Liu, J., Xu, J. F., Duan, S. G., Wang, Q. R., Li, G. C., et al. (2019). Transcriptome profiling reveals effects of drought stress on gene expression in diploid potato genotype P3-198. *Int. J. Mol. Sci.* 20, 852. doi: 10.3390/ijms20040852

Yang, C. P., and Wei, H. R. (2015). Designing microarray and RNA-Seq experiments for greater systems biology discovery in modern plant genomics. *Mol. Plant* 8, 196–206. doi: 10.1016/j.molp.2014.11.012

Yoshida, T., Christmann, A., Yamaguchi-Shinozaki, K., Grill, E., and Fernie, A. R. (2019). Revisiting the basal role of ABA - Roles outside of stress. *Trends Plant Sci.* 24, 625–635. doi: 10.1016/j.tplants.2019.04.008

Yu, T. F., Liu, Y., Fu, J. D., Ma, J., Fang, Z. W., Chen, J., et al. (2021). The NF-Y-PYR module integrates the abscisic acid signal pathway to regulate plant stress tolerance. *Plant Biotechnol. J.* 19, 2589–2605. doi: 10.1111/pbi.13684

Zander, M., La Camera, S., Lamotte, O., Metraux, J. P., and Gatz, C. (2010). *Arabidopsis thaliana* class-II TGA transcription factors are essential activators of jasmonic acid/ethylene-induced defense responses. *Plant J.* 61, 200–210. doi: 10.1111/j.1365-313X.2009.04044.x

Zhang, Q. L., Huang, J. D., Ke, W. Q., Cai, M. L., Chen, G. X., and Peng, C. L. (2021). Responses of *Sphagneticola trilobata*, *Sphagneticola calendulacea* and their hybrid to drought stress. *Int. J. Mol. Sci.* 22, 11288. doi: 10.3390/ijms222011288

Zhang, F. P., Sussmilch, F., Nichols, D. S., Cardoso, A. A., Brodribb, T. J., and Mcadam, S. (2018). Leaves, not roots or floral tissue, are the main site of rapid, external pressure-induced ABA biosynthesis in angiosperms. *J. Exp. Bot.* 69, 1261–1267. doi: 10.1093/jxb/erx480

Zhang, H., Zhao, Y., and Zhu, J. K. (2020). Thriving under stress: how plants balance growth and the stress response. *Dev. Cell* 55, 529–543. doi: 10.1016/j.devcel.2020.10.012

Zhang, J. H., Zhu, Y. J., Pan, Y., Huang, H. H., Li, C. L., Li, G. Z., et al. (2018). Transcriptomic profiling and identification of candidate genes in two *Phoebe bournei* ecotypes with contrasting cold stress responses. *Trees Struct. Funct.* 32, 1315–1333. doi: 10.1007/s00468-018-1713-0

Zhang, Y., Zhu, H., Zhang, Q., Li, M., Yan, M., Wang, R., et al. (2009). Phospholipase dalpha1 and phosphatidic acid regulate NADPH oxidase activity and production of reactive oxygen species in ABA-mediated stomatal closure in *Arabidopsis*. *Plant Cell* 21, 2357–2377. doi: 10.1105/tpc.108.062992

Zhou, M. A., Jiang, Y. F., Liu, X. H., Kong, W. L., Zhang, C. H., Yang, J., et al. (2023). Genome-wide identification and evolution analysis of the CYP76 subfamily in rice (*Oryza sativa*). *Int. J. Mol. Sci.* 24, 8522. doi: 10.3390/ijms24108522

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029

Zlobin, I. E. (2022). Linking the growth patterns of coniferous species with their performance under climate aridization. *Sci. Total Environ.* 831, 154971. doi: 10.1016/j.scitotenv.2022.154971

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 January 2024
DOI 10.3389/fpls.2024.1285094

OPEN ACCESS

EDITED BY
Meng-Zhu Lu,
Zhejiang Agricultural & Forestry University, China

REVIEWED BY
Marie Lillehammer,
Fisheries and Aquaculture Research (Nofima), Norway
Zitong Li,
Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia

*CORRESPONDENCE
Kyeong-Seong Cheon
kscheon16@korea.kr

RECEIVED 29 August 2023
ACCEPTED 05 January 2024
PUBLISHED 23 January 2024

CITATION
Kang H-I, Kim IS, Shim D, Kang K-S and Cheon K-S (2024) Genomic selection for growth characteristics in Korean red pine (*Pinus densiflora* Seibold & Zucc.). *Front. Plant Sci.* 15:1285094. doi: 10.3389/fpls.2024.1285094

Genomic selection for growth characteristics in Korean red pine (*Pinus densiflora* Seibold & Zucc.)

Hye-In Kang[1], In Sik Kim[1], Donghwan Shim[2], Kyu-Suk Kang[3] and Kyeong-Seong Cheon[1*]

[1]Division of Tree Improvement and Biotechnology, Department of Forest Bio-resources, National Institute of Forest Science, Suwon, Republic of Korea, [2]Department of Biological Sciences, Chungnam National University, Daejeon, Republic of Korea, [3]Department of Agriculture, Forestry and Bioresources, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea

Traditionally, selective breeding has been used to improve tree growth. However, traditional selection methods are time-consuming and limit annual genetic gain. Genomic selection (GS) offers an alternative to progeny testing by estimating the genotype-based breeding values of individuals based on genomic information using molecular markers. In the present study, we introduced GS to an open-pollinated breeding population of Korean red pine (*Pinus densiflora*), which is in high demand in South Korea, to shorten the breeding cycle. We compared the prediction accuracies of GS for growth characteristics (diameter at breast height [DBH], height, straightness, and volume) in Korean red pines under various conditions (marker set, model, and training set) and evaluated the selection efficiency of GS compared to traditional selection methods. Training the GS model to include individuals from various environments using genomic best linear unbiased prediction (GBLUP) and markers with a minor allele frequency larger than 0.05 was effective. The optimized model had an accuracy of 0.164–0.498 and a predictive ability of 0.018–0.441. The predictive ability of GBLUP against that of additive best linear unbiased prediction (ABLUP) was 0.86–5.10, and against the square root of heritability was 0.19–0.76, indicating that GS for Korean red pine was as efficient as in previous studies on forest trees. Moreover, the response to GS was higher than that to traditional selection regarding the annual genetic gain. Therefore, we conclude that the trained GS model is more effective than the traditional breeding methods for Korean red pines. We anticipate that the next generation of trees selected by GS will lay the foundation for the accelerated breeding of Korean red pine.

KEYWORDS

Korean red pine, progeny test, genomic selection, accelerated breeding, breeding value, genetic gain

1 Introduction

Korean red pine (*Pinus densiflora* Siebold & Zucc.), a species native to South Korea, belongs to the genus *Pinus* of the family Pinaceae and is widely distributed throughout East Asia, from the Korean Peninsula to Japan and China (Szmidt and Wang, 1993). Considering its high timber value, the reforestation of Korean red pine accounted for approximately 17% of the annual reforestation area in South Korea as of 2020, highlighting the need for research on improving wood productivity by breeding economic traits (KFS, 2021).

Breeding is important for genetically improving trees by applying genetic principles and techniques (White et al., 2007). However, compared with crops and livestock, forest trees have a longer generation period, and the establishment of a breeding population by crossing and nurturing takes a long time. Consequently, the advancement of a generation requires 30–45 years for tree breeding. Currently, the most advanced tree-breeding program for loblolly pines (*Pinus taeda*) has progressed to the fourth generation (Isik and McKeand, 2019). As the response to selection, which is the expected value of improvement according to the progress of one generation, is limited, rapid generational advancement is required to achieve accelerated breeding.

Since the development of next-generation sequencing (NGS) and statistical analysis methods for large-scale data, genomic selection (GS) has been proposed as an alternative to traditional selection methods. GS is a selective breeding method that uses molecular marker information rather than phenotype or pedigree information to estimate the genetic value of each individual as a criterion for selection in the breeding population (Meuwissen et al., 2001). Unlike conventional family selection (FS) and marker-assisted selection, GS estimates breeding value by summing the effects of thousands to hundreds of thousands of markers, making it possible to discern individual distinctions caused by minute marker effects (Goddard et al., 2011). Additionally, GS focuses solely on selection efficiency and does not require any prior information, such as the association between the phenotype and marker, location of the quantitative trait loci on the genome, or the relative influence of the marker on the phenotype (Isik et al., 2016). Accordingly, the time and effort required to find information on the association between a specific marker and the trait can be saved in GS. In addition, GS is particularly advantageous for forest trees because GS does not reply on reference genomes.

Studies on GS in loblolly pine and eucalyptus hybrids marked the first application of this approach to forestry (Resende et al., 2012; Resende et al., 2012a; Resende et al., 2012b). Subsequently, experimental studies on GS in conifers such as *Pinus*, *Picea*, *Pseudotsuga*, and *Cryptomeria* and broadleaf trees such as *Eucalyptus*, *Castanea*, *Fraxinus*, and *Populus* have been conducted (reviewed in Lebedev et al., 2020). Previous studies on forest trees employed population sizes of 25–338 half-sibling or full-sibling families as research subjects, primarily focusing on assessing their growth and wood quality attributes.

Generally, the GS in forest trees proceeds as follows: First, the genotypes and phenotypes of the training group in the breeding population are evaluated. Second, the two datasets are combined to create a prediction model that simultaneously estimate the effects of all marker loci. Third, cross-validation is performed to test the applicability of the developed model. Finally, genetic values are predicted for different subgroups of the breeding population, and individuals for advanced generations are selected based on the estimated genomic values (Grattapaglia, 2014).

In this study, we proposed an efficient GS model for Korean red pine by investigating the impact of various factors, including the single nucleotide polymorphism (SNP) marker set, predictive model, and training dataset, on prediction accuracy in the half-sib population. Additionally, the prediction accuracy of the trained GS model was evaluated through standardization. Moreover, the efficiency of GS was evaluated by comparing the response to GS and conventional selections.

2 Materials and methods

2.1 Study population

The study population was from an open-pollinated progeny trial established for the progeny test of plus trees by the National Institute of Forest Sciences (NIFoS). Seeds obtained from 49 grafted clones (20 and 29 clones of plus trees from Gangwon and Kyeongbuk provinces, respectively) of the clone bank located in Taean were nursed to produce 1-1 seedlings. Subsequently, the seedlings were distributed and planted in 6 regions: Taean (36°46' N, 126°38' W), Chuncheon (37°89' N, 127°63' W), Gongju (36°60' N, 127°10' W), Kyeongju (35°92' N, 129°09' W), Naju (35°06' N, 126°84' W), and Wanju (35°85' N, 127°28' W). A total of 2,643 trees for which both phenotypes and genotypes were investigated from the open-pollinated progeny trial were used for the analysis: 609 from Taean, 726 from Chuncheon, 456 from Gongju, 380 from Kyeonju, 247 from Naju, and 225 from Wanju. These trees belonged to 44 open-pollination families (Supplementary Table 1).

2.2 Phenotyping

The target traits of GS were growth, including diameter at breast height (DBH), height, straightness, and volume. Laser scanning technology, which is capable of consistently generating high-precision outputs as well as environmentally friendly because of its nondestructive properties (Dittmann et al., 2017; Chen et al., 2019), was used to replace the conventional survey. To obtain the phenotype of each individual, point cloud data from six sites in the study population were obtained using a LiDAR device (ScanStation P40, Leica) in 2017 and 2018. For batch data processing, the point-cloud data were separated into trees and ground, and the ground was flattened. The DBH (m) was obtained by dividing the circumference of the tree at a height of 1.2 m from the ground by the circumference ratio (π). Tree height (m) was measured as the distance from the ground to the top of the canopy. The volume (m^3) of each tree was calculated as $(DBH)^2 \times$ (height). To calculate straightness, an imaginary baseline was written connecting the center point of the tree trunk at the root collar and a height of

6 m. The distances from the baseline to the center point of the tree at heights of 0.5, 1.5, 2.5, 3.5, 4.5, and 5.5 m were calculated. The straightness was digitized by taking the negative natural log (-$\log_e$) of the standard deviation of the six distance values. Specifically, the more the stem had a shape similar to the baseline, the greater the calculated straightness. The distribution of the phenotypic data used in this study is shown in Supplementary Figure 2..

2.3 Genotyping

To investigate the genotype of each tree at the six test sites, cambium tissues were collected from tree stems, and DNA was extracted using the Exgene™ Plant SV Kit (Geneall, Seoul). SNP probe information for genotyping was obtained using a 50 K SNP chip for Korean red pines (Cheon et al., 2021). The SNP chip was developed using genotyping-by-sequencing (GBS) information from 46 trees tested in the open-pollinated progeny trial. Axiom™ analysis suite (v5.0.1) was used for SNP calling. First, the sample quality was investigated, and samples with a sample call rate of less than 97% or a DQC value of less than 0.82 based on probe intensity were excluded. The genotypes (AA, BB, AB, and NN) were clustered for each SNP marker based on the intensity ratio of the two allele probes.

2.4 Preparation of GRM

As there should be no missing genotype data to write a genomic realized relationship matrix (GRM), imputation was performed using Beagle (v5.2), which operates even without a reference genome (Browning and Browning, 2009). GRMs were prepared according to the formula described by VanRaden (2008). For comparison with the GRM, a numerator relationship matrix (NRM) consisting of the relationship coefficients of two individuals was also written using pedigree information. The GRMs differed according to the marker set used, but the coefficients of the GRM and NRM commonly showed approximate matches (Supplementary Figure 3).

2.5 Heritability

The heritability of the mixed model was obtained as the variance of NRM, and the GRM was divided by the total variance in the best linear unbiased prediction (BLUP).

$$h^2_{NRM} = \frac{\sigma^2_{NRM}}{\sigma^2_{NRM} + \sigma^2_r}, \quad h^2_{GRM} = \frac{\sigma^2_{GRM}}{\sigma^2_{GRM} + \sigma^2_r}$$

The site and block were set as the fixed effects of the mixed linear model in the combined and one-site analyses, respectively.

Family heritability was estimated using analysis of variance (ANOVA). The mathematical linear model for analyzing the variance component and heritability of each site was as follows:

$$X_{ijk} = \mu + B_i + F_j + BF_{ij} + \epsilon_{ijk}$$

where X_{ijk} is the phenotype of the k-th tree of the j-th family in the i-th block. μ is the overall mean, B_i is the effect of the block, F_j is the effect of family, BF_{ij} is the interaction effect of block and family, and ϵ_{ijk} is the error term. The mathematical linear model for the combined analysis of the six sites is as follows:

$$X_{ijkl} = \mu + S_i + B_{ij} + F_k + SF_{ik} + BF_{ijk} + \epsilon_{ijkl}$$

where X_{ijkl} is the phenotype of the l-th tree of the k-th family in the j-th block of the i-th site, μ is the overall mean, S_i is the effect of the site, B_{ij} is the effect of the block, F_k is the effect of the family, SF_{ik} is the interaction effect of the site and family, BF_{ij} is the interaction effect of the block and family, and ϵ_{ijk} is the error term.

2.6 SNP marker selection

To compare the prediction accuracy according to the SNP calling the quality of markers, markers selected based on loose, moderate, and strict standards were used for GS. The loose standard was the use of all markers, the moderate standard was to meet the default quality threshold of the SNP calling program (call rate (CR)≥97% and Fisher's linear discriminant (FLD)≥3.6), and the strict standard was the addition of the threshold of CR≥99%, FLD≥5, heterozygous strength offset (HetSO)≥0, and homozygote ratio offset (HomRO)≥0. In common, markers with minor allele frequency (MAF) of 0.05 or higher were used.

To study the prediction accuracy according to the number of markers, all 17 K (17,074) markers showing genotype variation and 2 K (2,000), 6 K (6,000), and 10 K (10,000) markers randomly selected markers were used for GS. In addition, markers with MAF of 0.25, 0.05, and 0.0005 or higher were selected and compared to investigate the impact of MAF on prediction accuracy.

2.7 Genomic estimated breeding value prediction

Six genomic predictive models, including genomic BLUP (GBLUP), Bayesian least absolute shrinkage and selection operator (LASSO), Bayesian Ridge Regression, Bayes A, Bayes B, and Bayes C, were used to predict genomic estimated breeding value (GEBV). In addition, additive BLUP (ABLUP), a pedigree-based prediction method using NRM, was compared. BLUPs were performed using remlf90 in the R package BreedR (v0.12.5). In addition, five Bayesian models were applied with 20,000 iterations using the GBLR function of the R package BGLR (v1.0.8) (Pérez and de Los Campos, 2014).

2.8 Genomic selection scenario and evaluating prediction accuracy

To evaluate the prediction accuracy of GS, the data set was split into the training set, of which both genotype and phenotype were

used for training the model, and the test set, of which phenotype were masked, so to be predicted by model. In this study, three genomic selection scenarios were performed according to how training set and test set were defined: within-region prediction, between-region prediction, and combined-region prediction. Three-, five-, ten-, and twenty-fold cross-validations were examined and compared in a within-region prediction scenario. In the between-region prediction scenario, all individuals in the remaining five regions (training set) were trained to predict the GEBV for one region (test set). In addition, in the combined-region prediction scenario, 10-fold cross-validation was performed by randomly dividing the groups regardless of the region. When multiple regions were included in a scenario, the phenotype was corrected by setting the region as a fixed effect in the linear model.

The prediction accuracy was evaluated using accuracy (AC) and predictive ability (PA). Accuracy was the Pearson correlation coefficient of the GEBV of the test set and the EBV calculated by ABLUP using pedigree information and all phenotypes (Supplementary Figure 4), which we assume as true breeding value (TBV) as in traditional genetics and majority of forest tree GS studies (Chen et al., 2018; Li et al., 2019; Beaulieu et al., 2020). The predictive ability was the Pearson correlation coefficient between the GEBV of the test set and the phenotype. In cross-validation, prediction accuracy is the average values of correlation coefficients for each subset. For a meaningful comparison, we also presented the average correlation coefficient for the same subset in between-region prediction.

2.9 Response to selection

To compare the response to the selection of GS and two traditional selections, FS and phenotypic selection (PS), the annual genetic gain of each was calculated. Genetic gains from FS and PS were calculated as follows (Voss-Fels et al., 2019).

$$\Delta G_P = \frac{ih^2\sigma_P}{t}$$

where ΔG_P is the annual genetic gain, i is the selection intensity, h^2 is $h_{GRM}{}^2$ for PS and family heritability by ANOVA (Supplementary Table 5) for FS, σ_P is the square root of phenotypic variance for PS and family phenotypic variation for FS, and t is breeding cycle. The genetic gain from GS was measured as follows (Isik et al., 2017; Voss-Fels et al., 2019).

$$\Delta G_A = \frac{ir\sigma_A}{t}$$

where ΔG_A is the annual genetic gain, i is the selection intensity, r is the accuracy of GS, σ_A is the square root of the additive genetic variance, and t is the breeding cycle. The selection intensity, according to the selection ratio, was calculated assuming a normal phenotypic distribution.

The breeding cycle of the GS was assumed to be 15 years, considering the age of reproduction of Korean red pine. For the FS, the time required for the progeny test (30 years in the study population) was added. Accordingly, a total breeding cycle of 45 years was assumed for the FS. In addition, for PS, the breeding cycle was assumed to be 30 years because of the time required to grow to the phenotyping age.

3 Results

3.1 Heritability

To estimate narrow-sense individual heritability and breeding values in this study, a mixed model was employed instead of the traditionally used ANOVA as the latter is not suitable for unbalanced data with different observation numbers per family and block (Isik et al., 2017). The relationship matrices, GRM and NRM, were used as random effects of the mixed model. The heritability of the population ranged from 0.000 to 0.723, exhibiting substantial variability across test sites and traits (Table 1). Height displayed the highest heritability among all the assessed traits. NRM and GRM heritability estimates showed a similar trend depending on the test sites and traits. Notably, the heritability values derived from the GRM were generally higher than those derived from the NRM.

3.2 Impact of SNP marker set on predictive accuracy

To determine the impact of SNP marker quality on the prediction accuracy of GS, GBLUP analysis was performed using different marker quality standards, and the resulting accuracies and predictive abilities were compared. The number of markers meeting the standards was 6,464 for the loose standard, 1,164 for the moderate standard, and 571 for the strict standard, respectively. Across all the traits and regions, the stricter the applied standard, the lower the accuracy and predictive ability (Supplementary Table 6). Although some data showed differences within the error range, all traits in Chuncheon exhibited significant differences in accuracy across the different marker sets. In the subsequent analysis, a loose standard was applied for the selection of markers to achieve relatively high GS accuracy.

To examine the impact of the number of markers, all 17 K (17,074) markers showing genotype variation and 2 K (2,000), 6 K (6,000), and 10 K (10,000) marker sets randomly selected from among them were used in GBLUP, and prediction accuracy was compared. As a result, the accuracy was 0.09–0.46 for 2 K, 0.18–0.47 for 6 K, 0.24–0.48 for 10 K, and 0.22–0.51 for 17 K marker sets (Figures 1A-D; Supplementary Table 7). Also, the predictive ability was -0.03–0.43 for 2 K, 0.01–0.42 for 6 K, 0.01–0.44 for 10 K, and 0.02–0.46 for 17 K marker sets (Figures 1E-H; Supplementary Table 7). Generally, as the number of markers decreased, the prediction accuracy decreased.

Subsequently, an analysis was conducted to investigate whether a reduction in the number of MAF markers would result in a decrease in prediction accuracy. The prediction accuracy was compared using GBLUP analysis with 2 K (2,248), 6 K (6,464), and 10 K (9,799) markers with MAF of 0.25, 0.05, and 0.0005 or higher, respectively. As a result, the accuracy was 0.14–0.5, and the

TABLE 1 Narrow-sense heritability by the mixed model using NRM and GRM in progeny test sites.

Region [a]	Estimates	DBH	Height	Straightness	Volume
T	h_{NRM}^2	0.103 (0.083)	0.249 (0.124)*	0.078 (0.092)	0.143 (0.092)
	h_{GRM}^2	0.186 (0.106)	0.723 (0.103)**	0.146 (0.105)	0.276 (0.112)*
C	h_{NRM}^2	0.195 (0.101)	0.380 (0.148)*	0.045 (0.060)	0.208 (0.103)*
	h_{GRM}^2	0.420 (0.099)**	0.393 (0.098)**	0.227 (0.095)*	0.463 (0.105)**
G	h_{NRM}^2	0.407 (0.187)*	0.457 (0.200)*	0.245 (0.144)	0.413 (0.190)*
	h_{GRM}^2	0.354 (0.151)*	0.720 (0.138)**	0.356 (0.146)*	0.393 (0.150)**
K	h_{NRM}^2	0.055 (0.115)	0.000 (0.000)	0.327 (0.193)	0.001 (0.003)
	h_{GRM}^2	0.051 (0.105)	0.564 (0.197)**	0.405 (0.171)*	0.004 (0.091)
N	h_{NRM}^2	0.068 (0.155)	0.132 (0.179)	0.366 (0.277)	0.043 (0.151)
	h_{GRM}^2	0.546 (0.228)*	0.642 (0.215)**	0.395 (0.291)	0.519 (0.225)*
W	h_{NRM}^2	0.070 (0.188)	0.108 (0.212)	0.000 (0.003)	0.206 (0.235)
	h_{GRM}^2	0.173 (0.175)	0.105 (0.187)	0.286 (0.245)	0.173 (0.187)
Combined	h_{NRM}^2	0.074 (0.031)*	0.146 (0.049)**	0.113 (0.042)**	0.108 (0.040)**
	h_{GRM}^2	0.066 (0.017)**	0.104 (0.020)**	0.059 (0.017)**	0.082 (0.018)**

Standard error of heritability estimation in parenthesis.
The values marked with asterisk indicate a difference from zero (*, $\alpha<0.05$; **, $\alpha<0.01$).
[a]T, Taean; C, Chuncheon; G, Gongju; K, Kyeongju; N, Naju; W, Wanju.

predictive ability was -0.02–0.46, showing differences within the error range (Supplementary Table 8). Specifically, when markers with low MAF were excluded, the prediction accuracy did not decrease, unlike when random markers were excluded.

3.3 Impact of the predictive model on predictive accuracy

The prediction accuracies of various models were analyzed and compared to identify a suitable predictive model for GS in Korean red pine. In addition, ABLUP was performed to compare the efficiency of GS with pedigree-based selection. As a result, six GS models exhibited an accuracy of 0.15–0.5 and a predictive ability of 0.01–0.44, with no significant differences observed among one another (Figure 2; Supplementary Table 9). Meanwhile, compared with ABLUP (accuracy 0.32–0.72, predictive ability -0.18–0.25) based on pedigree information, the GS models generally showed lower accuracy and higher predictive ability (Figure 2; Supplementary Table 9), which may likely be attributed to the fact that the phenotype is not strongly correlated with the initially assumed TBV (Supplementary Figure 4).

3.4 Impact of training data set on predictive accuracy

To examine whether the size of the training and test sets affected the prediction accuracy of the GS of Korean red pine, the accuracy and predictive ability of GBLUP were compared by varying the number of cross-validation folds. As a result of 3-, 5-, 10-, and 20-fold cross-validations, the accuracy was 0.14–0.51, and the predictive ability was -0.07–0.44, showing differences within the error range regardless of the size or ratio of the training set (Supplementary Table 10). For some regions and traits, such as the volume in Naju, the prediction accuracy increased as the training set size increased.

To determine whether GS can be applied to populations in different environments, within-, between-, and combined-region GS scenarios were compared (Figure 3; Supplementary Table 11). Accuracy did not show consistent results depending on whether the analysis was within or between regions. However, the predictive ability was generally higher in the within-region prediction (0.02–0.41) than in the between-region prediction (0.05–0.24). In the combined-region scenario, the accuracy was 0.38–0.48, which was higher than the average of six regions (0.33–0.38), and the predictive ability was 0.07–0.18, which was lower than the average of the six regions (0.17–0.28). Higher accuracy could be obtained when the GS model was trained by combining multiple environments, consistent with previous study in two spruce species (Lenz et al., 2017; Chen et al., 2018).

3.5 Prediction accuracy evaluation

Synthesizing the study results, the optimized GS model for Korean red pine was the GBLUP model, using the genotypes of 6,464 markers with an MAF of 0.05 or higher. Finally, the prediction accuracy of the GS performed by this model was 0.164–0.498 for accuracy and 0.018–0.441 for predictive ability (Table 2). The prediction accuracy was highest for height in Taean. The mean prediction accuracy for all regions was the highest for height. The population size and prediction accuracy were not correlated.

10.3389/fpls.2024.1285094

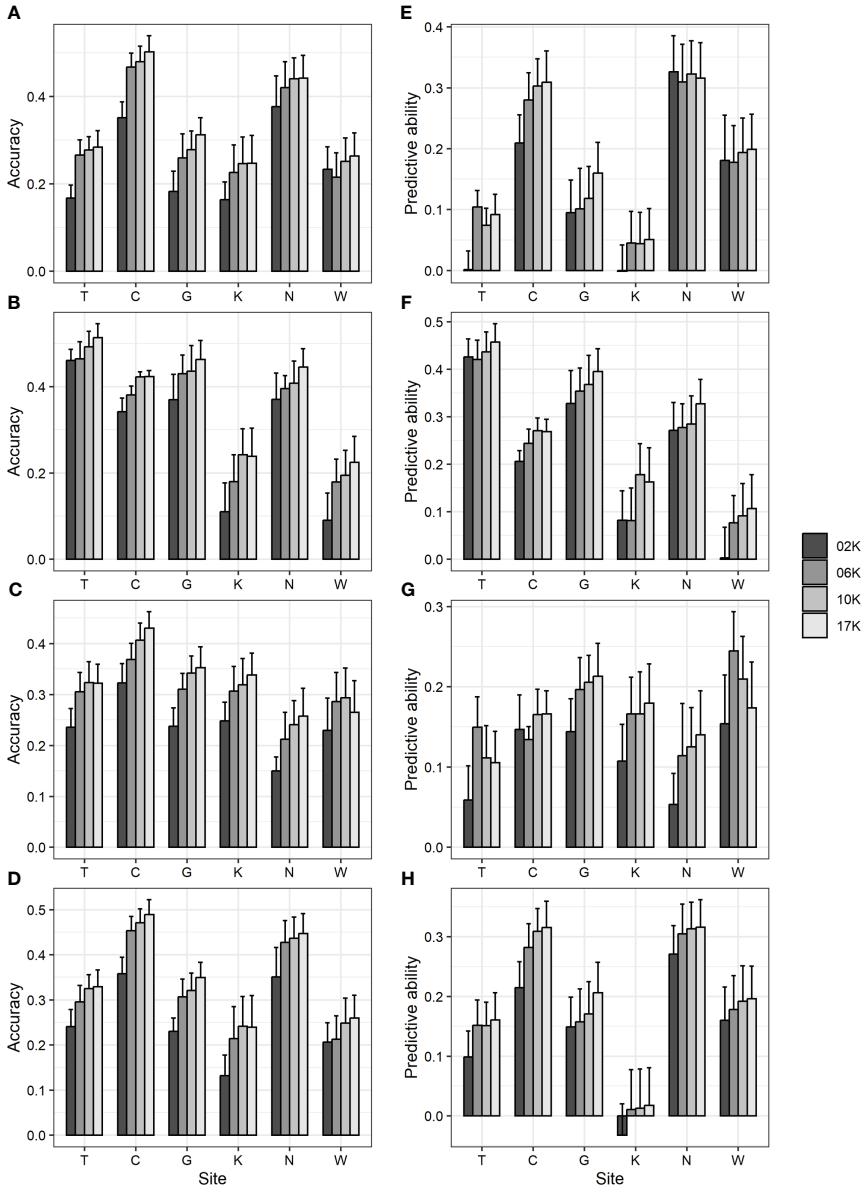


FIGURE 1
GBLUP accuracy and predictive ability using 17 K (17,074) markers showing genotype variation and 2 K (2,000), 6 K (6,000), and 10 K (10,000) markers randomly selected from them. **(A, E)** DBH **(B, F)** height **(C, G)** straightness and **(D, H)** volume. **(A-D)** accuracy **(E-H)** predictive ability. Bar and error bar are mean and standard error of accuracy and predictive ability from 10-fold cross-validation.

Pearson correlation analysis was performed for each region and trait to reveal the correlation between heritability and prediction accuracy of GS in Korean red pine. The correlation coefficient between accuracy and heritability was 0.735 ($p<0.001$), and that between predictive ability and heritability was 0.924 ($p<0.001$), indicating a strong correlation (Figure 4).

3.6 Response to selection

To test the efficiency of GS versus traditional breeding, the genetic gains expected to be obtained using GS, PS, and FS were compared. Regarding the annual genetic gain in the combined region analysis, GS was the highest, followed by FS, and PS was the lowest for height, straightness, and volume (Figure 5). The efficiency of GS was superior to that of the two traditional selection methods for DBH. GS showed an annual genetic gain equivalent of 2.9–3.7 times PS and 1.7–3.2 times FS when 20% selection was conducted. In addition, the annual genetic gain of GS was the highest at 0.09–1.24% volume. In addition, GS was the most efficient of the three selection methods in the within-region analysis (Supplementary Table 12). In the GS within the region, a response to selection of up to 1.9% per year was obtained.

4 Discussion

Breeding woody plants is a time- and cost-intensive process compared with that of crops, primarily because trees have a lengthy

10.3389/fpls.2024.1285094

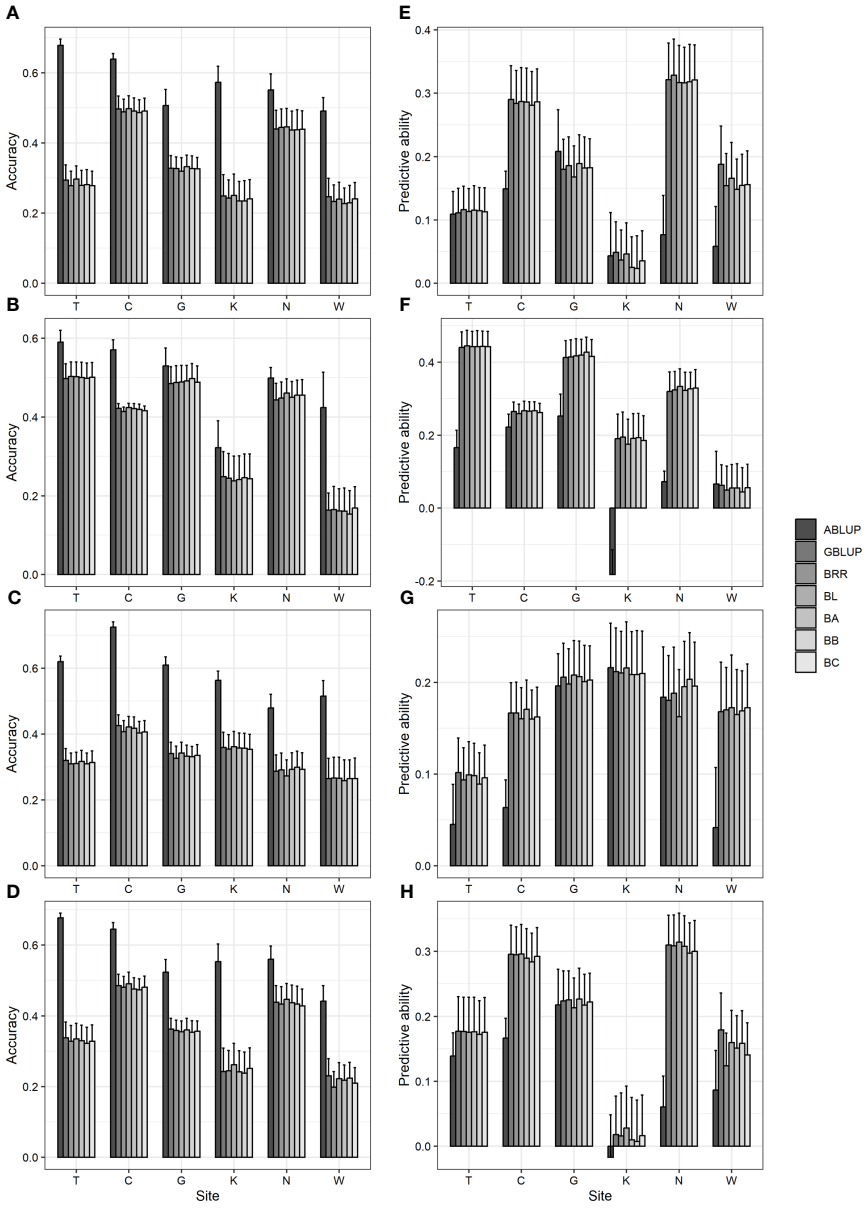


FIGURE 2
Accuracy and predictive ability by ABLUP and genomic selection models, including GBLUP and five Bayesian models. **(A, E)** DBH **(B, F)** height **(C, G)** straightness and **(D, H)** volume. **(A-D)** accuracy **(E-H)** predictive ability. BRR, Bayesian ridge regression; BL, Bayesian LASSO; BA, Bayes A; BB, Bayes B; BC, Bayes C. Bar and error bar are mean and standard error of accuracy and predictive ability from 10-fold cross-validation.

juvenile period, take longer to flower and produce seeds, and are physically larger than crops. Progeny tests for forest trees require a vast area and prolonged observation. In addition, as forest tree utilization continues to diversify and global climate change intensifies, the target traits of tree breeding are rapidly changing. Therefore, accelerated breeding is crucial for forestry.

The major advantage of GS in forest trees is that the selection efficiency can be improved by reducing the generation interval through genomic information-based selection before the phenotypes are expressed (Grattapaglia and Resende, 2011). Moreover, GS can increase selection intensity, resulting in a greater response to selection (Isik, 2014; Grattapaglia, 2017). Therefore, if GS is introduced into the breeding of Korean red pine, we can expect significant improvements by accelerating generations at a relatively low cost through early and intense marker-based selection rather than relying solely on progeny tests.

4.1 Heritability predictive accuracy

Heritability has been reported to have a significant effect on the prediction accuracy of GS. The prediction accuracy of GS increased as heritability increased in loblolly pine (Resende et al., 2012a), and the square root of heritability was strongly correlated with GS accuracy (Lian et al., 2014). However, in large training populations of 1,000 or more, heritability has a relatively small effect on GS

10.3389/fpls.2024.1285094

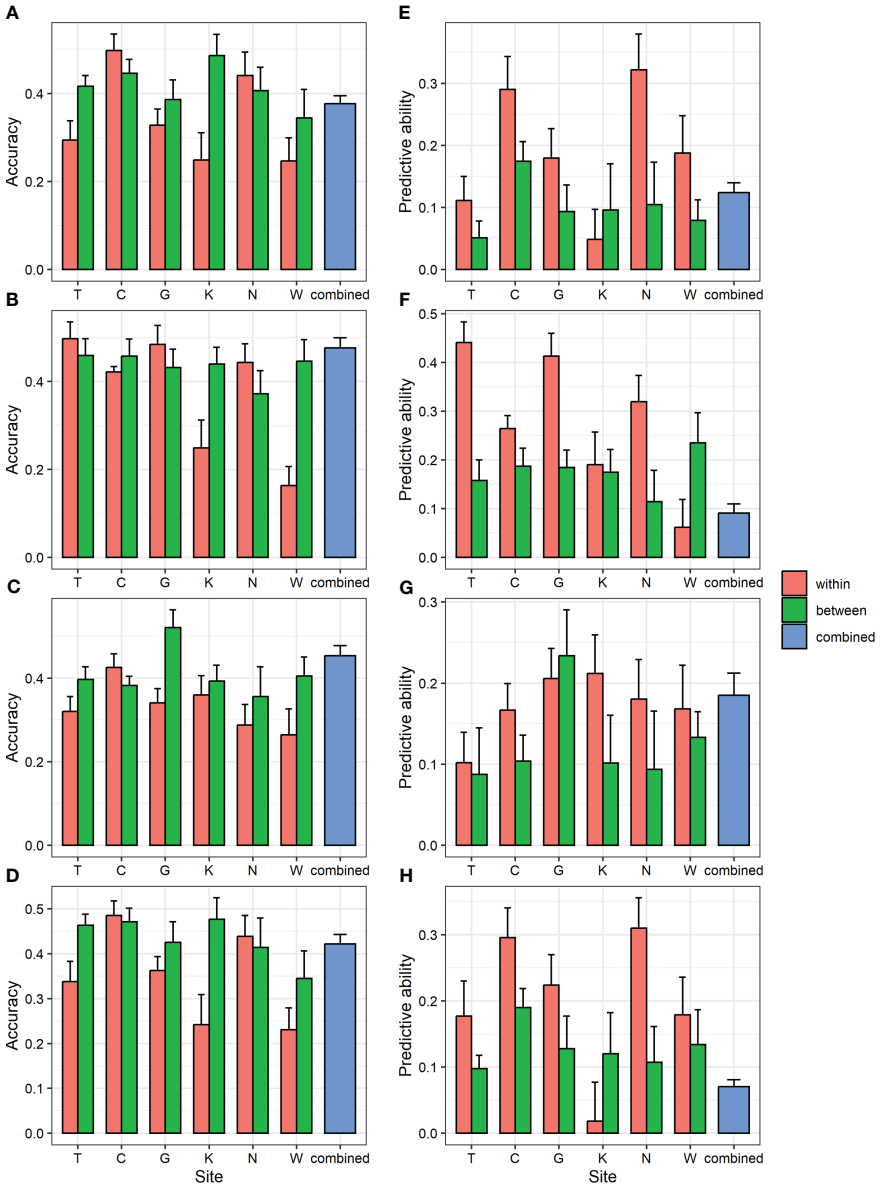


FIGURE 3
GBLUP accuracy and predictive ability of within- and between-region analysis and combined region analysis for four traits. **(A, E)** DBH **(B, F)** height **(C, G)** straightness and **(D, H)** volume. **(A-D)** accuracy **(E-H)** predictive ability. Bar and error bar are mean and standard error of accuracy and predictive ability from 10-fold cross-validation.

accuracy compared to other factors (Grattapaglia and Resende, 2011).

Heritability estimates obtained using the GRM in this study were generally higher than those obtained using the NRM for each region (Table 1). This difference is because the relationship coefficients of the NRM may not accurately reflect the actual relationships among individuals. In the present study, the kinship coefficient of open-pollinated family progenies was assumed to be 0.25, based on the half-sib family assumption (Wright, 1922). However, in the real world, kinship relationships among open-pollinated siblings might be stronger because of various factors, including self-pollination, self-half, half-sib, full-sib, and common ancestry between female and male parents (Askew and El-Kassaby, 1994).

According to the correlation analysis, the prediction accuracy of GS in the Korean red pine was strongly influenced by heritability (Figure 4). Predictive ability showed a stronger correlation with heritability than with accuracy because it was calculated by phenotype, which implied the effects of environment and non-additive genetic variance. Therefore, heritability must be considered when performing GS on Korean red pines.

4.2 Model optimization

The essential stage of GS is training the model to estimate the effects of all markers. This stage includes optimizing the model to achieve the best prediction efficiency. We investigated the impact of

10.3389/fpls.2024.1285094

TABLE 2 Prediction accuracy for four traits using GBLUP of 10-fold cross-validation.

Trait	Prediction accuracy [a]	Region [b]							
		T	C	G	K *	N	W	Mean	Combined
DBH	AC	0.294	0.497	0.328	0.249	0.441	0.267	0.346	0.377
	PA	0.111	0.29	0.18	0.049	0.322	0.188	0.19	0.124
	G/A	1.022	1.948	0.864	1.117	4.197	3.23	–	1.223
	G/H	0.257	0.447	0.303	0.217	0.436	0.452	–	0.483
Height	AC	0.498	0.422	0.485	0.249	0.443	0.164	0.377	0.476
	PA	0.441	0.264	0.413	0.19	0.32	0.062	0.282	0.091
	G/A	2.664	1.189	1.633	–	4.41	0.94	–	1.159
	G/H	0.519	0.421	0.487	0.253	0.399	0.191	–	0.282
Straight-ness	AC	0.32	0.426	0.341	0.36	0.288	0.265	0.333	0.453
	PA	0.102	0.167	0.206	0.212	0.18	0.168	0.173	0.185
	G/A	2.258	2.626	1.048	0.981	0.981	4.038	–	1.558
	G/H	0.267	0.351	0.345	0.333	0.286	0.314	–	0.762
Volume	AC	0.338	0.485	0.363	0.242	0.439	0.231	0.35	0.421
	PA	0.177	0.296	0.224	0.018	0.31	0.179	0.201	0.07
	G/A	1.274	1.775	1.029	–	5.098	2.067	–	0.539
	G/H	0.337	0.435	0.357	0.285	0.43	0.43	–	0.244

[a]AC, accuracy, r(GEBV, EBV_{im}), EBV_{im} is estimated using ABLUP with all phenotypic data; PA, predictive ability, r(GEBV, phenotype); G/A, Predictive ability of GBLUP against that of ABLUP; G/H, Predictive ability of GBLUP against the square root of heritability.
[b]T, Taean; C, Chuncheon; G, Gongju; K, Kyeongju; N, Naju; W, Wanju.
* Values for height and volume were omitted because the predictive ability of ABLUP was negative.

the marker set, predictive model, and training dataset on the prediction accuracy of the model optimization for GS in Korean red pine.

Ensuring the quality of the SNP array data is crucial because it significantly affects the accuracy and precision of subsequent analyses. Contaminated data may lead to false-positive or false-negative results, underscoring the importance of controlling the data quality (Yang et al., 2011). Although previous studies on GS in forest trees have generally set a call rate criterion of 85–95% (Beaulieu et al., 2014; Cappa et al., 2019; Ukrainetz and Mansfield, 2020a), the impact of marker quality on the accuracy of GS has not been extensively explored. This study compared the effects of marker set size and quality on GS accuracy in Korean red pines (Supplementary Table 6). Notably, we found that including markers of slightly lower quality but increasing the number of markers was more effective in improving the accuracy of GS than focusing only on high-quality markers. This may be because the imputation of missing genotypes could compensate for the low call rates of the markers. Previous studies have shown that adding markers with low call rates can improve prediction accuracy when the markers are not saturated in the whole genome (Rutkoski et al., 2013). This suggests that the 500–1000 SNP markers in the array used in this study may not be sufficient to capture the entire Korean red pine genome. Another possible explanation for the higher accuracy of the imputed data is that imputation can capture associations among closely related individuals that might have been missed due to missing data (Weigel et al., 2010; Rutkoski et al., 2013).

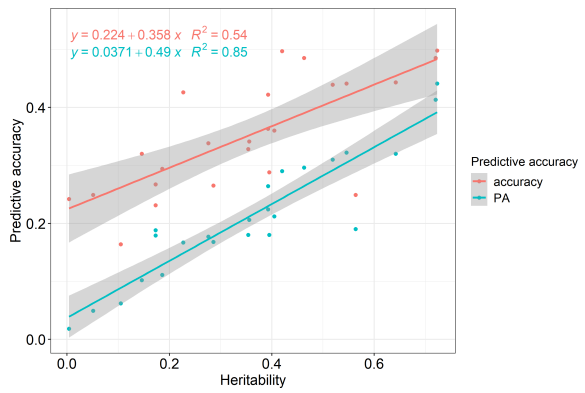


FIGURE 4
Prediction accuracies according to heritability.

The number of markers used in GS is a critical factor that affects the prediction accuracy as well as the computational time required for the analysis. Therefore, identifying the optimal number of markers for GS analysis is important. Our study showed that the MAF had a more significant impact on the prediction accuracy of GS than the number of markers (Supplementary Tables 7, 8). Specifically, for height in Chuncheon and Kyeongju, we observed

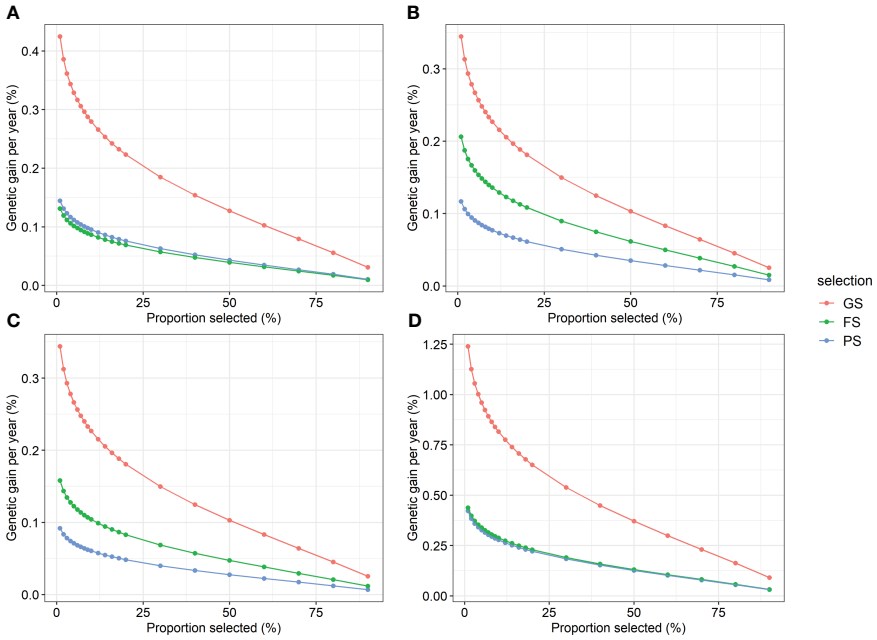


FIGURE 5
Annual genetic gain of genomic selection and two traditional selections by proportion selected for four traits. **(A)** DBH, **(B)** height, **(C)** straightness, and **(D)** volume. GS, genomic selection; FS, family selection; PS, phenotypic selection.

a significant decline in the prediction accuracy and predictive ability when the MAF was below 0.25 (2 K), suggesting that useful markers for prediction were present between the MAF range of 0.05 and 0.25. This MAF range was consistent with the selection of markers based on an MAF of 0.005–0.05 in previous studies on the GS of forest trees (Beaulieu et al., 2014; Cappa et al., 2019; Ukrainetz and Mansfield, 2020a). Therefore, our study supports the conclusion that selecting markers based on an MAF of 0.05 is efficient for GS analysis of Korean red pine, consistent with previous studies.

The GBLUP model was originally proposed as a predictive model for GS, and subsequent developments have led to the application of Bayesian models in various situations. However, most studies on the quantitative traits of forest trees, including spruce hybrids, blue gum (*Eucalyptus globulus*), maritime pine (*Pinus pinaster*), lodgepole pine (*Pinus contorta*), and Japanese cedar (*Cryptomeria japonica*), have not found any significant advantage in a particular model (Ratcliffe et al., 2015; Isik et al., 2016; Durán et al., 2017; Hiraoka et al., 2018; Ukrainetz and Mansfield, 2020b). Factors such as overfitting the training population and computing time should be considered when the accuracies of all models are similar (Heslot et al., 2012). Considering the similar accuracy and predictive ability of the six models tested in this study and that GBLUP was nine times faster than the Bayesian methods, we concluded that GBLUP is a more efficient choice for the GS of Korean red pine.

Generally, the ratio of training set size to test set size did not affect predictive accuracy (Supplementary Table 10), which was consistent with the results for Norway spruce (*Picea abies*) (Chen et al., 2018) and eucalyptus (Tan et al., 2017). This suggests that other factors, such as population structure and environment, were more important than the ratio of the training set when the population size was 200–700 for the GS of Korean red pine.

4.3 Environment effect on GS

The interaction between genotype and environment (G×E) refers to the inconsistency in the expression of traits when individuals are grown in different environments. Typically, there is greater interaction when clones or families change ranks across different environments. Because GS ranks individuals according to the GEBV, considering G×E when developing GS strategies is important (Grattapaglia, 2017). The predictive ability was lower in the between-region scenario than that in the within-region scenario (Figure 3; Supplementary Table 11), indicating a strong G×E in study population. This result was further supported by the generally low type-B genetic correlation ($r_{12} = \sigma_{a12}/\sqrt{\sigma_{a1}^2\sigma_{a2}^2}$) and changing phenotype ranks of family across six sites (Supplementary Table 13, Supplementary Figure 14). Training across a range of environments may be beneficial to increase the probability of including similar environments for new populations in the training dataset. As a direction for further research, an approach worth considering is the utilization of models that take environmental factors into account, as exemplified in some crop studies (Burgueño et al., 2012; Jarquín et al., 2014; Sukumaran et al., 2017).

4.4 The efficiency of GS in Korean red pine

Similar to many previous studies on black spruce (*Picea mariana*), white spruce (*Picea glauca*), and lodgepole pine (Lenz

et al., 2017; Beaulieu et al., 2020; Lenz et al., 2020a; Ukrainetz and Mansfield, 2020b), the predictive ability of GS models was higher than those of FS and PS in Korean red pine, confirming that GS for obtaining GEBV using DNA markers has an advantage in phenotype prediction over traditional selection in forest trees.

The prediction accuracy of GS is influenced by population-specific features, such as heritability, making direct comparisons across different populations unreliable. Instead, standardization using measures, such as the prediction accuracy of ABLUP or heritability, can be used for more accurate comparisons. Within-region analysis showed that the predictive ability of GBLUP compared to ABLUP ranged from 0.864 to 5.098 (Table 2), which was higher than the range of 0.80 to 0.95 reported for Norway spruce (Chen et al., 2018). The predictive ability of GBLUP against the square root of heritability was 0.19 to 0.76 (Table 2), which was lower on average than the result for Norway spruce (higher than 0.69) but similar to the result for white spruce (Beaulieu et al., 2020; Lenz et al., 2020b). Therefore, GS in Korean red pine was as efficient as in previous forest tree studies.

The annual genetic gain comparison revealed that GS was more efficient than the conventional selection methods (Figure 5). Moreover, selection intensity in the GS was enhanced. Because selection can be conducted at the seedling stage in GS, whereas PS and FS are conducted at the age after phenotype expression, the selection intensity could be increased in GS for the same number of selected trees (Grattapaglia, 2017). Therefore, in terms of annual genetic gain, the GS of Korean red pine was judged to be more efficient than the two traditional selection methods in both within- and combined-region scenarios.

4.5 Prospective of GS of Korean red pine

Statistical GS models such as GBLUP and Bayesian models are constrained by various factors, such as their underlying assumption of additive effects, limited capacity to capture non-linear relationships, and the difficulty they encounter when handling extensive datasets (Budhlakoti et al., 2022). Conversely, machine learning-based GS methods provide a more versatile and scalable approach, adept at capturing intricate non-linear relationships, rendering them particularly well-suited for addressing various genetic scenarios (Montesinos-López et al., 2021a; Montesinos-López et al., 2021b). Given the substantial impact of environmental factors and the prevalence of missing genotype data in this study, exploring deep learning-based genomic selection is essential for future research endeavors in GS of Korean red pine.

5 Conclusion

In this study, an efficient GS model for Korean red pine was presented, and its selection efficiency was evaluated. As a result of comparing various marker subsets, predictive models, and scenarios to optimize the GS model in Korean red pine, training the model with markers of MAF of 0.05 or more, using GBLUP as a predictive model, and including as many environments as possible was effective. The GS of the Korean red pine was as efficient as that of the other tree species. In addition, it was found to have a higher response to selection than traditional PS and FS. Thus, GS is an appropriate alternative to the traditional selection of Korean red pines. The results of this study provide evidence that GS is effective for forest trees even in challenging environments characterized by mountainous terrain, low heritability, and widespread distribution across diverse regions.

Data availability statement

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

Author contributions

HK: Formal analysis, Software, Visualization, Writing – original draft. IK: Conceptualization, Supervision, Writing – review & editing. DS: Conceptualization, Investigation, Methodology, Writing – review & editing. KK: Formal analysis, Writing – review & editing. KC: Data curation, Project administration, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Institute of Forest Science of the Republic of Korea. Grant number FG0400-2022-01-2023.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1285094/full#supplementary-material

SUPPLEMENTARY FIGURE 2
Distribution of phenotypes by test site and combined data. **(A)** DBH, **(B)** height, **(C)** straightness, and **(D)** volume. Box colors and red X symbols indicate the median and mean of phenotypes respectively in each site. Alphabets in box indicate the Games-Howell *post-hoc* analysis group. T, Taean; C, Chuncheon; G, Gongju; K, Kyeongju; N, Naju; W, Wanju.

SUPPLEMENTARY FIGURE 3
Heatmaps of coefficient of **(A)** numerator relationship matrix and **(B)** genomic realized relationship matrix ordered by open-pollinated family and **(C)** distribution of GRM coefficients according to their corresponding NRM coefficients. Symbol X indicates the mean of the genomic realized relationship coefficient. GRM was prepared with marker filtering according to the default threshold for marker quality suggested by the SNP calling program, MAF of 0.05, and classifications of high resolution (1,164 SNPs) in this figure.

SUPPLEMENTARY FIGURE 4
Breeding values estimated using ABLUP by site for four traits. **(A)** DBH, **(B)** height, **(C)** straightness, and **(D)** volume. T, Taean; C, Chuncheon; G, Gongju; K, Kyeongju; N, Naju; W, Wanju.

References

Askew, G. R., and El-Kassaby, Y. A. (1994). Estimation of relationship coefficients among progeny derived from wind-pollinated orchard seeds. *Theor. Appl. Genet.* 88 (2), 267–272. doi: 10.1007/BF00225908

Beaulieu, J., Doerksen, T. K., MacKay, J., Rainville, A., and Bousquet, J. (2014). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics* 15 (1), 1–16. doi: 10.1186/1471-2164-15-1048

Beaulieu, J., Nadeau, S., Ding, C., Celedon, J. M., Azaiez, A., Ritland, C., et al. (2020). Genomic selection for resistance to spruce budworm in white spruce and relationships with growth and wood quality traits. *Evolutionary Appl.* 13 (10), 2704–2722. doi: 10.1111/eva.13076

Browning, B. L., and Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84 (2), 210–223. doi: 10.1016/j.ajhg.2009.01.005

Budhlakoti, N., Kushwaha, A. K., Rai, A., Chaturvedi, K. K., Kumar, A., Pradhan, A. K., et al. (2022). Genomic selection: a tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. *Front Genet*, 13, 832153. doi: 10.3389/fgene.2022.832153

Burgueño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52 (2), 707–719. doi: 10.2135/cropsci2011.06.0299

Cappa, E. P., de Lima, B. M., da Silva-Junior, O. B., Garcia, C. C., Mansfield, S. D., and Grattapaglia, D. (2019). Improving genomic prediction of growth and wood traits in Eucalyptus using phenotypes from non-genotyped trees by single-step GBLUP. *Plant Sci.* 284, 9–15. doi: 10.1016/j.plantsci.2019.03.017

Chen, Z. Q., Baison, J., Pan, J., Karlsson, B., Andersson, B., Westin, J., et al. (2018). Accuracy of genomic selection for growth and wood quality traits in two control-pollinated progeny trials using exome capture as the genotyping platform in Norway spruce. *BMC Genomics* 19 (1), 1–16. doi: 10.1186/s12864-018-5256-y

Chen, S., Liu, H., Feng, Z., Shen, C., and Chen, P. (2019). Applicability of personal laser scanning in forestry inventory. *PloS One* 14 (2), e0211392. doi: 10.1371/journal.pone.0211392

Cheon, K. S., Kang, H. I., Park, Y. W., Song, J. H., Kim, I. S., and Shim, D. (2021). Development of SNP chip for Genomic Selection of Korean Red Pine (Pinus densiflora) Trees. *Proc. Korean Soc. Breed. Sci.* 406.

Dittmann, S., Thiessen, E., and Hartung, E. (2017). Applicability of different non-invasive methods for tree mass estimation: A review. *For. Ecol. Manage.* 398, 208–215. doi: 10.1016/j.foreco.2017.05.013

Durán, R., Isik, F., Zapata-Valenzuela, J., Balocchi, C., and Valenzuela, S. (2017). Genomic predictions of breeding values in a cloned Eucalyptus globulus population in Chile. *Tree Genet. Genomes* 13 (4), 1–12. doi: 10.1007/s11295-017-1158-4

Goddard, M. E., Hayes, B. J., and Meuwissen, T. H. (2011). Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128 (6), 409–421. doi: 10.1111/j.1439-0388.2011.00964.x

Grattapaglia, D. (2014). Breeding forest trees by genomic selection: current progress and the way forward. *Genomics Plant Genet. Resour.* 1, 651–682. doi: 10.1007/978-94-007-7572-5_26

Grattapaglia, D. (2017). "Status and perspectives of genomic selection in forest tree breeding," in *Genomic selection for crop improvement* (Cham: Springer), 199–249.

Grattapaglia, D., and Resende, M. D. (2011). Genomic selection in forest tree breeding. *Tree Genet. Genomes* 7 (2), 241–255. doi: 10.1007/s11295-010-0328-4

Heslot, N., Yang, H. P., Sorrells, M. E., and Jannink, J. L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52 (1), 146–160. doi: 10.2135/cropsci2011.06.0297

Hiraoka, Y., Fukatsu, E., Mishima, K., Hirao, T., Teshima, K. M., Tamura, M., et al. (2018). Potential of genome-wide studies in unrelated plus trees of a coniferous species, Cryptomeria japonica (Japanese cedar). *Front. Plant Sci.* 9, 1322. doi: 10.3389/fpls.2018.01322

Isik, F. (2014). Genomic selection in forest tree breeding: the concept and an outlook to the future. *New Forests* 45 (3), 379–401. doi: 10.1007/s11056-014-9422-z

Isik, F., Bartholomé, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., et al. (2016). Genomic selection in maritime pine. *Plant Sci.* 242, 108–119. doi: 10.1016/j.plantsci.2015.08.006

Isik, F., Holland, J., and Maltecca, C. (2017). *Genetic data analysis for plant and animal breeding (Vol. 400)* (Cham, Switzerland: Springer International Publishing).

Isik, F., and McKeand, S. E. (2019). Fourth cycle breeding and testing strategy for Pinus taeda in the NC State University Cooperative Tree Improvement Program. *Tree Genet. Genomes* 15 (5), 1–12. doi: 10.1007/s11295-019-1377-y

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., et al. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* 127, 595–607. doi: 10.1007/s00122-013-2243-1

Korea Forest Service (KFS) (2021). *2021 Statistical Yearbook of Forestry* Vol. 51 (Daejeon, Korea).

Lebedev, V. G., Lebedeva, T. N., Chernodubov, A. I., and Shestibratov, K. A. (2020). Genomic selection for forest tree improvement: Methods, achievements and perspectives. *Forests* 11 (11), 1190. doi: 10.3390/f11111190

Lenz, P., Beaulieu, J., Mansfield, S. D., Clément, S., Desponts, M., and Bousquet, J. (2017). Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (Picea mariana). *BMC Genomics* 18 (1), 1–17. doi: 10.1186/s12864-017-3715-5

Lenz, P., Nadeau, S., Azaiez, A., Gérardi, S., Deslauriers, M., Perron, M., et al. (2020a). Genomic prediction for hastening and improving efficiency of forward selection in conifer polycross mating designs: an example from white spruce. *Heredity* 124 (4), 562–578. doi: 10.1038/s41437-019-0290-3

Lenz, P. R., Nadeau, S., Mottet, M. J., Perron, M., Isabel, N., Beaulieu, J., et al. (2020b). Multi-trait genomic selection for weevil resistance, growth, and wood quality in Norway spruce. *Evolutionary Appl.* 13 (1), 76–94. doi: 10.1111/eva.12823

Li, Y., Klápště, J., Telfer, E., Wilcox, P., Graham, N., Macdonald, L., et al. (2019). Genomic selection for non-key traits in radiata pine when the documented pedigree is corrected using DNA marker information. *BMC Genomics* 20 (1), 1–10. doi: 10.1186/s12864-019-6420-8

Lian, L., Jacobson, A., Zhong, S., and Bernardo, R. (2014). Genomewide prediction accuracy within 969 maize biparental populations. *Crop Sci.* 54 (4), 1514–1522. doi: 10.2135/cropsci2013.12.0856

Meuwissen, T. H., Hayes, B. J., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4), 1819–1829. doi: 10.1093/genetics/157.4.1819

Montesinos-López, O. A., Montesinos-López, A., Hernandez-Suarez, C. M., Barrón-López, J. A., and Crossa, J. (2021a). Deep-learning power and perspectives for genomic selection. *Plant Genome* 14 (3), e20122. doi: 10.1002/tpg2.20122

Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W., Fajardo-Flores, S. B., et al. (2021b). A review of deep learning applications for genomic selection. *BMC Genomics* 22, 1–23. doi: 10.1186/s12864-020-07319-x

Pérez, P., and de Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198 (2), 483–495. doi: 10.1534/genetics.114.164442

Ratcliffe, B., El-Dien, O. G., Klápště, J., Porth, I., Chen, C., Jaquish, B., et al. (2015). A comparison of genomic selection models across time in interior spruce (Picea engelmannii× glauca) using unordered SNP imputation methods. *Heredity* 115 (6), 547–555. doi: 10.1038/hdy.2015.57

Resende, M. F. R. Jr., Muñoz, P., Acosta, J. J., Peter, G. F., Davis, J. M., Grattapaglia, D., et al. (2012b). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytol.* 193 (3), 617–624. doi: 10.1111/j.1469-8137.2011.03895.x

Resende, M. F. R. Jr., Munoz, P., Resende, M. D., Garrick, D. J., Fernando, R. L., Davis, J. M., et al. (2012a). Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.). *Genetics* 190 (4), 1503–1510. doi: 10.1534/genetics.111.137026

Resende, M. D., Resende, M. F. Jr., Sansaloni, C. P., Petroli, C. D., Missiaggia, A. A., Aguiar, A. M., et al. (2012). Genomic selection for growth and wood quality in Eucalyptus:

capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* 194 (1), 116–128. doi: 10.1111/j.1469-8137.2011.04038.x

Rutkoski, J. E., Poland, J., Jannink, J. L., and Sorrells, M. E. (2013). Imputation of unordered markers and the impact on genomic selection accuracy. *G3: Genes Genomes Genet.* 3 (3), 427–439. doi: 10.1534/g3.112.005363

Sukumaran, S., Crossa, J., Jarquin, D., Lopes, M., and Reynolds, M. P. (2017). Genomic prediction with pedigree and genotype× environment interaction in spring wheat grown in South and West Asia, North Africa, and Mexico. *G3: Genes Genomes Genet.* 7 (2), 481–495. doi: 10.1534/g3.116.036251

Szmidt, A. E., and Wang, X. R. (1993). Molecular systematics and genetic differentiation of Pinus sylvestris (L.) and P. densiflora (Sieb. et Zucc.). *Theor. Appl. Genet.* 86 (2), 159–165. doi: 10.1007/BF00222074

Tan, B., Grattapaglia, D., Martins, G. S., Ferreira, K. Z., Sundberg, B., and Ingvarsson, P. K. (2017). Evaluating the accuracy of genomic prediction of growth and wood traits in two Eucalyptus species and their F1 hybrids. *BMC Plant Biol.* 17 (1), 1–15. doi: 10.1186/s12870-017-1059-6

Ukrainetz, N. K., and Mansfield, S. D. (2020a). Prediction accuracy of single-step BLUP for growth and wood quality traits in the lodgepole pine breeding program in British Columbia. *Tree Genet. Genomes* 16 (5), 1–13. doi: 10.1007/s11295-020-01456-w

Ukrainetz, N. K., and Mansfield, S. D. (2020b). Assessing the sensitivities of genomic selection for growth and wood quality traits in lodgepole pine using Bayesian models. *Tree Genet. Genomes* 16 (1), 1–19. doi: 10.1007/s11295-019-1404-z

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91 (11), 4414–4423. doi: 10.3168/jds.2007-0980

Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* 132 (3), 669–686. doi: 10.1007/s00122-018-3270-8

Weigel, K. A., de Los Campos, G., Vazquez, A. I., Rosa, G. J. M., Gianola, D., and Van Tassell, C. P. (2010). Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. *J. Dairy Sci.* 93 (11), 5423–5435. doi: 10.3168/jds.2010-3149

White, T. L., Adams, W. T., and Neale, D. B. (Eds.) (2007). *Forest genetics* (Wallingford, UK: Cabi).

Wright, S. (1922). Coefficients of inbreeding and relationship. *Am. Nat.* 56 (645), 330–338. doi: 10.1086/279872

Yang, H. C., Lin, H. C., Kang, M., Chen, C. H., Lin, C. W., Li, L. H., et al. (2011). SAQC: SNP array quality control. *BMC Bioinf.* 12 (1), 1–14. doi: 10.1186/1471-2105-12-100

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 August 2024
DOI 10.3389/fpls.2024.1339958

OPEN ACCESS

EDITED BY
Juan Du,
Zhejiang University, China

REVIEWED BY
Ram Singh Purty,
Guru Gobind Singh Indraprastha University, India
Madhab Kumar Sen,
Czech University of Life Sciences Prague, Czechia

*CORRESPONDENCE
Takashi Tamura
tktamura@okayama-u.ac.jp

RECEIVED 17 November 2023
ACCEPTED 07 August 2024
PUBLISHED 29 August 2024

CITATION
Ishio S, Kusunoki K, Nemoto M, Kanao T and Tamura T (2024) Illumina-based transcriptomic analysis of the fast-growing leguminous tree *Acacia crassicarpa*: functional gene annotation and identification of novel SSR-markers. *Front. Plant Sci.* 15:1339958. doi: 10.3389/fpls.2024.1339958

Illumina-based transcriptomic analysis of the fast-growing leguminous tree *Acacia crassicarpa*: functional gene annotation and identification of novel SSR-markers

Shougo Ishio[1], Kazutaka Kusunoki[1], Michiko Nemoto[2], Tadayoshi Kanao[2] and Takashi Tamura[3]*

[1]Tsukuba Research Institute, Sumitomo Forestry Co. Ltd., Tsukuba, Japan, [2]Graduate School of Environment, Life, Natural Science and Technology, Okayama University, Okayama, Japan, [3]Institute of Global Human Resource Development, Okayama University, Okayama, Japan

Acacia crassicarpa is a fast-growing leguminous tree that is widely cultivated in tropical areas such as Indonesia, Malaysia, Australia, and southern China. This tree has versatile utility in timber, furniture, and pulp production. Illumina sequencing of *A. crassicarpa* was conducted, and the raw data of 124,410,892 reads were filtered and assembled *de novo* into 93,317 unigenes, with a total of 84,411,793 bases. Blast2GO annotation, Benchmark Universal Single-Copy Ortholog evaluation, and GO-term classification produced a catalogue of unigenes for studying primary metabolism, phytohormone signaling, and transcription factors. Massive transcriptomic analysis has identified microsatellites composed of simple sequence repeat (SSR) loci representing di-, tri-, and tetranucleotide repeat units in the predicted open reading frames. Polymorphism was induced by PCR amplification of microsatellite loci located in several genes encoding auxin response factors and other transcription factors, which successfully distinguished 16 local trees of *A. crassicarpa* tested, representing potentially exploitable molecular markers for efficient tree breeding for plantation and biomass exploitation.

KEYWORDS

***Acacia crassicarpa*, illumina sequencing, polymorphism, auxin response factor, lignin**

1 Introduction

Acacia is a fast-growing leguminous tree that can be harvested within a relatively short period (several years). The genetic traits for such rapid carbon fixation allow for a rapid cycle of harvesting and reforestation of tree plantations. Because of their ability to thrive in degraded soils, even under drought conditions (Pan and You, 1994), these trees are planted

in large areas of Southeast Asia and Southern China (Midgley and Turnbull, 2003). Acacia crassicarpa is a preferred leguminous tree with valuable properties, such as high wood density, excellent biomass yield, low moisture content, and high combustion heat. Because of the intrinsic ability of trees to store abundant plant nutrients K+, Ca2+, and Mg2+ from the soil, their ash can become highly alkaline with a pH greater than 12 (Yusiharni and Gilkes, 2012). Large-scale burning of wood produces residual alkaline ash, which is highly sticky and causes significant mechanical damage to movable floors in biomass power plants (Fuller et al., 2019; Hallgren et al., 1999). These properties of wood species present breeding challenges that must be solved, and accurate genomic information and molecular breeding techniques are required to address these challenges.

Because traditional breeding of tree species has intrinsic limitations owing to their slow growth and naturally long lifecycles, it is highly challenging to select improved varieties that have the desired genetic traits. Molecular breeding of commercial plants requires several basic resources such as genome sequence information and annotation of genes responsible for target traits. A. crassicarpa has a relatively large genome of 1,350 Mbp (Mukherjee and Sharma, 1995), which is approximately double that of the model species Acacia mangium, at 635 Mbp (Blakesley et al., 2002). Genetic modification of A. crassicarpa is difficult without precise sequence data, such as those obtained from expressed sequence tag (EST) libraries. Efforts have been made to construct transcriptomic databases and microsatellite markers for Acacia tree species (McKinnon et al., 2018), and a few pioneering studies have identified a limited number of ESTs. For example, in A. mangium, 8,963 of 10,752 clones of a cDNA library were constructed via conventional molecular cloning and sequencing (Suzuki et al., 2011).

Massively parallel sequencing of cDNA libraries (RNA-Seq) has become a valuable tool for genome analysis, and is rapidly replacing ESTs for structural and functional genome analysis in plants (Pashley et al., 2006). The illumina-based technology permits the investigation of spliced transcripts, including alternative splicing, leading to the large-scale discovery of novel transcripts and the identification of gene boundaries at single-nucleotide resolution.

Plant somatic cells can undergo dedifferentiation to give rise to a pluripotent cell mass called a callus, which can potentially regenerate new organs or the whole plant (Sugimoto et al., 2010). Proliferative plant somatic cells also provide the opportunity to introduce foreign genes to alter their genetic traits. After genetic alteration, it is necessary to induce differentiation of callus cells into an adventitious shoot and then rooting to establish a plant body. The cellular reprogramming process involved several genetic perturbations induced by physical injury of explant and supplementation of auxin and cytokinin. Elucidation of the gene regulatory mechanism, involving several transcription factors, signaling pathway components, epigenetic alteration of chromatin, and activation of biosynthetic pathways for growth and propagation, may give us clue how to reverse the transformed callus back to the differentiated plant body.

In this study, de novo parallel sequencing was performed to characterize the transcriptome of the pluripotent state of A. crassicarpa. To our knowledge, this is the first illumination-based transcriptomic analysis of A. crassicarpa. The cultured cells provided sufficient RNA without severe mechanical shearing or chemical damage. The annotation of gene ontology (GO) terms was established with reference to the model plant Arabidopsis thaliana and the related tree Prosopis alba. Illumina sequencing data also enabled the identification of microsatellites containing simple sequence repeats (SSR). Unigenes expressing a high degree of polymorphism may reveal hidden links between genotypic diversity and physiological function, which may lead to segregation of subspecies to obtain desirable wood properties.

2 Materials and methods

2.1 Plant material and RNA isolation

Callus tissues were obtained from *A. crassicarpa* seedlings. The seeds were purchased from Australian Tree Seed Center (ATSC) of the Commonwealth Scientific and Industrial Research Organization. Acacia seeds were soaked in concentrated sulfuric acid overnight to burn off the surface wax, and the seed husk was then removed by washing with tap water. The seeds were sterilized with bleach (0.05% chlorine), transferred onto solid 1/2MS medium containing 0.5 ppm TDZ and 1.0 ppm IAA, and cultured for 30 days at 32°C until callus development.

For transcriptome sequencing, 0.1 g of tissue pieces were manually homogenized with a pestle and vortexed for 2 min using super beads. The total RNA was fractionated using an automated RNA extraction system (Maxwell RSC48, Maxwell Plant RNA Kit; Promega, Tokyo, Japan). RNA integrity was determined using a NanoDrop spectrophotometer and Agilent Bioanalyzer (Agilent Technologies Japan, Ltd.). For SSR analysis, 16 individual tree of *A. crassicarpa* were harvested from the Seed Production Area, Conn, Australia, were purchased from the ATSC.

2.2 Complementary DNA library construction and sequencing

A cDNA library was constructed from the mRNA using the Illumina TruSeq RNA Sample Preparation Kit (Illumina Inc. CA, USA). The cDNA isolated using AMPure XP beads was subjected to an end-repair process that converted the overhanging nucleotide ends into blunt ends. The blunt-end cDNAs were then incubated with adenylate to attach to the 3′ end; thus, adaptor DNA bearing a 5′-T overhang could capture the cDNA library. DNA was amplified by PCR and the PCR primer cocktail was annealed onto adaptor sequences. The Illumina next-generation sequencing protocol uses a sequencing-by-synthesis approach with four proprietary nucleotides that possess a reversible fluorophore and terminating properties. A series of images, each representing a single-base extension at a specific cluster, were recorded as the sequencing cycle was repeated at specific clusters on the flow cell surface. Raw sequence reads were deposited in the DNA Data Bank of Japan and Sequence Read Archive database (Experiment DRX208625; Analysis Run DRR218312).

2.3 Data analysis and estimation of abundance

Raw sequence data were checked for quality using the Fast QC algorithm (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The trimmomatic algorithm was then employed to trim the sequence data and to detect and remove adapter sequences (Bolger et al., 2014). This process performs in silico normalization of the total reads to reduce the number of reads that are subject to de novo assembly, thus improving the runtime of the assembly required for the following Trinity algorithm. A large volume of raw data to be analyzed by paired-end sequencing was reconstructed de novo. The relative abundance of transcripts was estimated and quantified using the RSEM 1.2.15 machine (https://www.encodeproject.org/software/rsem/) in the Trinity program (Grabherr et al., 2011).

2.4 Sequencing analysis and annotation

For sequence analysis, all unigenes were initially aligned using BLASTx (e-value < 10^{-5}) to protein databases such as NCBI non-redundant proteins (nr) (Altschul et al., 1990), Swiss-Prot (Boeckmann et al., 2003), COG (Tatusov et al., 2000) and KEGG (Kanehisa and Goto, 2000) databases. Unigenes were then aligned using BLASTN (e-value < 10^{-5}) to the NCBI non-redundant nucleotide (nt) nucleic acid database, and proteins with the highest sequence similarity to the given unigenes were retrieved along with their protein functional annotations. Homologous genes in *A. crassicarpa* were identified using a BLASTp search of putative protein sequences, with an e-value threshold of <1e–10 and > 90% identity. Annotation of the unigenes with GO terms was performed using the Blast2GO software based on BLASTX hits against the NCBI nr database (e-value < 10^{-5}). The quality of the unigenes was evaluated *via* BUSCO analysis using the gVolante website (Nishimura et al., 2017), which categorizes unigenes into complete, fragmented, duplicated, and missing genes. WEGO software (Ye et al., 2018) was used to perform GO functional classification of the height of the unigenes and plot the macro-level distribution of the *Acacia* database. The regulation of gene expression by transcription factors (TF) is common and important for cellular response. To characterize the operational TF families underneath the dedifferentiating *A. crassicarpa*, 41,716 amino acid sequences deduced from the unigene sequences were subjected to transcription factor enrichment analysis in Plant TFDB search (Jin et al., 2017).

2.5 EST-SSR analysis for polymorphism

The MIcroSAtellite (MISA) identification tool ver. 2.0 in OmicX (Beier et al., 2017) was used to search for repeated nucleotide motifs. The tool predicted the presence of multiple sets of SSR in the assembled unigenes and the allele sizes of the amplified PCR products of the candidate genes were measured. The sequence search for SSRs was performed by setting the search parameters to identify at least five repeat units of SSR containing a maximum of 10 base pairs. The software requires the input file to be the sequence in which SSRs are to be found in FASTA format, and the output file contains the name of the sequence in which the SSR is detected, the repeat motif of the SSR, the number of repetitions, the start and end of the repeat, and the length of the sequence. The number, frequency, and distribution of repeats in the SSR motifs were recorded.

To obtain experimental evidence, pairs of primers were designed for the region where the SSR sequences were predicted with an intended product size of approximately 50-bp. Fluorescently labelled PCR primers were prepared using U-19 universal reverse primers (5'-gttttcccagtcacgacgt-3') labelled with four fluorescent molecules, 6-FAM, VIC, NED, and PET. Each labelled U-19 primer was tagged with a 2-bp barcode at the 3' end: TG for 6-FAM, AC for VIC, CA for NED, and GT for PET. Non-labelled forward primers had a pigtail (5'-gtttctt-3') at the 5' end. The PCR mixture was prepared at a volume of 20 µL using the Type-it Microsatellite PCR Kit (Qiagen, Germany) containing 50–100 ng of template DNA, 0.2 µM of each labelled U-19 primer, 0.2 µM of each pigtailed forward primer and 0.04 µM of each U-19-fused reverse primer. The PCR conditions were as follows: 95°C for 5 min, followed by 10 cycles of 95°C for 30 s, 58°C for 90 s with a decrease of 0.5°C in each cycle, and 72°C for 30 s, followed by 25 cycles of 95°C for 30 s, 53°C for 90 s, and 72°C for 30 s. The final extension step was performed at 60°C for 30 min. Fragment analysis was conducted on an ABI 3130xl Genetic Analyzer (Applied Biosystems) using a LIZ-600 size standard (Applied Biosystems). Allele sizes were scored using GeneMapper 4.1 (Applied Biosystems, Foster City, CA).

3 Results

3.1 Paired-end illumina sequencing and *de novo* assembly

Total RNA was extracted from cultured cells and mRNA was fractionated using poly T-tailed affinity beads with an automated Maxwell RSC48 System (Promega). Fractionated mRNA (59.5 µg/mL, 0.2 mL) was obtained, and cDNA was prepared and subjected to Illumina HiSeq 2000 sequencing. The raw data comprised 124,410,892 reads containing 12,565,500,092 bp, which were filtered to remove cloudy reads, low-quality reads with ambiguous N bases, reads in which >10% of the bases had Q < 20, and flanking adapter sequences (Table 1). Clean reads were assembled to recover full-length transcripts across a wide range of expression levels (Grabherr et al., 2011). After stringent quality assessment and data filtering, 123,829,386 clean reads were obtained with a total of 12,473,930,687 bases. The Q20 was 97.64%, and the guanine–cytosine (GC) content was 45.9%. Using Trinity, all clean reads were assembled de novo into 93,317 unigenes, with a total of 84,411,793 bases and an N50 length of 1,563 bp. The average unigene length 905 bp was comparable to that reported by recent

10.3389/fpls.2024.1339958

TABLE 1 Summary of RNA-seq Analysis of *A. crassicarpa*.

Items	Number	
Illumina HiSeq 2000 sequencing		
Total raw reads	124,410,892	read
Total raw nucleotide (nt	12,565,500,092	base
De novo Gene Assembling by Trinity		
Total clean reads	123,829,386	read
Total clean nucleotide (nt)	12,473,930,687	base
Q20%[1]	97.64	%
GC[2] percentage	45.9	%
Gene Ontology		
Total unigenes	93,317	read
Total clean nucleotides (nt)	84,411,793	Base in total
N50 length	1,563	bp
Average unigene length (nt)	905	bp
Total annotated Unigenes	46,788	(50.14%)
EST-SSR search by MIcroSAtellite		
Total number of SSR-containing sequences	18,141	
Number of plural SSRs in a single contig[3]	3,875	
Number of tetramer-nucleotide-SSR	135	
Number of pentamer-nucleotide-SSR	14	
Number of hexamer-nucleotide-SSR	14	

[1]Q20% is the percentage of nucleotides with a quality value greater than 20.
[2]GC, guanine–cytosine.
[3]The number of nucleotides in a repeated unit is 4 for tertramer-, 5 for pentamer-, and 6 for hexamer-nucleotide-SS.

studies on avocado, Persea americana (average unigene length, 988 bp; N50 = 1,050) (Vergara-Pulgar et al., 2019); raspberry, Rubus idaeus 'Heritage' (average, 1,168 bp; N50 = 2,046) (Travisany et al., 2019); and Salvia guaranitica L. (average, 1,039 bp; N50 = 1,603) (Ali et al., 2018).

3.2 Quality assessment of assembled unigenes by benchmark universal single-copy ortholog analysis

The assembled unigenes were subjected to BUSCO analysis using the gVolante website (Nishimura et al., 2017) for a completeness assessment. BUSCO performs a like-for-like comparison, which allows categorization of assembled unigenes into complete, fragmented, duplicated, or missing genes. About 59% (853) of the unigenes were categorized as 'complete,' 23% (334) were 'duplicate,' 5% (68) were 'fragmented,' and 13% (185) were missing from the A. thaliana reference.

The term 'complete' refers to a unigene within two standard deviations (2σ) of the BUSCO group's mean aligned sequence length. Unigenes outside of this limit were classified as 'fragmented' transcripts. The 'duplicated' category refers to unigenes that match multiple BUSCO reference genes, fulfilling both the 'expected-score' and the 'expected-length' cut-offs. 'Missing' unigenes are the ones that do not meet the expected score cut-off. The low percentage of fragmented (5%) and missing (13%) unigenes substantiated the quality of the assembled unigenes.

To further investigate the relatedness of the assembled unigenes to closely related tree species, differentially expressed genes were mapped to Acacia ESTs using all-versus-all pairwise comparison. The 93,317 unigenes identified in the present study were aligned against 81,212 genes expressed in the inner bark of Acacia koa (Lawson and Ebrahimi, 2018), 8,963 genes expressed in the secondary xylem and shoot of Acacia mangium (Suzuki et al., 2011), and 2,459 ESTs in the bark of Acacia auriculiformis × A. mangium (Yong et al., 2011). The similarities are shown in the Venn diagram (Figure 1). There were 29,707 overlapping unigenes between A. crassicarpa and the inner bark of A. koa, accounting for 44% and 37% of the assembled unigenes, respectively. The proportion of overlapping unigenes between A. crassicarpa and the secondary xylem and shoot of A. mangium was only 9%, representing a smaller library comprising 8,963 unigenes. The ESTs identified in the inner bark tissues of Acacia auriculiformis × mangium comprised a small number of 2,459 genes, and the A. crassicarpa unigenes identified in our transcriptomic library shared only 7% gene identity overlap. This indicates that the number of genes that next-generation sequencing can decode is an order of magnitude higher than the number of genes that EST sequencing can decode.

3.3 Functional annotation and GO-term assignment

The Blast2GO program (Conesa et al., 2005) was used to annotate the assembled 93,317 unigenes, and Gene ontology (GO) terms were assigned to 46,788 (50.14%) unigenes in reference to the NCBI nr, NCBI nt, Swiss-Prot, Clusters of Orthologous Groups (COG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases for annotation and validation. Base alignment with an e-value of < 10-5 was selected. To perform the GO functional classifications for all the 'complete' unigenes, the WEGO software (Ye et al., 2018) was used to examine the macro-level distribution of gene functions for this species. Of the complete unigenes assembled in our study, GO terms assigned by BLAST2GO were used in WEGO to categorize GO functional classifications. The Unigenes were categorized into three major categories: cellular components (31%), molecular functions (36%), and biological processes (33%). The Unigenes were assigned to cellular components (22.9%), cell parts (22.9%), membranes (17.4%), membrane parts (13.8%), and organelles (16.5%) (Figure 2). Molecular functions of the unigenes included catalytic activity (25%) and binding (21.8%). The biological processes of the unigenes included cellular (26.0%) and metabolic (26.1%) processes.

10.3389/fpls.2024.1339958

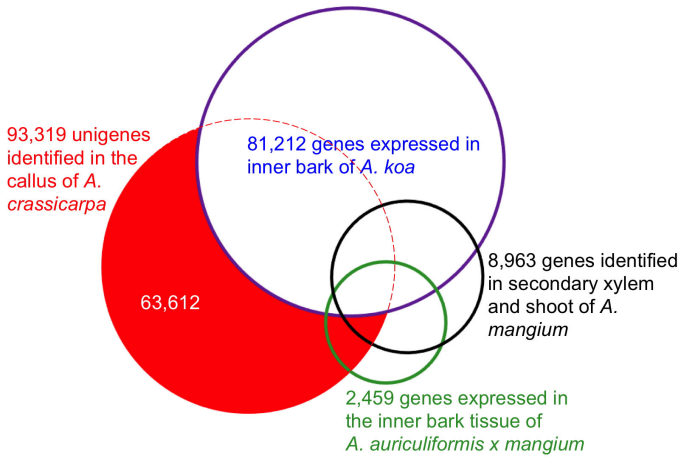


FIGURE 1
Overlap expressed genes in the callus of *A. crassicarpa* compared to the previously published EST database. Putative orthologous genes were defined by blastn search with the e-value cutoff 1xe^-10.

3.4 Metabolic pathway reconstruction by KEGG mapping

The KEGG database was used to systematically assign annotated unigenes to the metabolic pathways. KEGG mapping successfully categorized 16,662 unigenes, which accounted for 35.6% of the total unigenes, in accordance with the hierarchal annotation system KEGG BRITE, into the three major protein families involved in metabolism (2,406 unigenes), genetic information processing (2,666 unigenes), and signaling and cellular process (595 unigenes). The metabolic pathways were reconstructed by mapping the functionally annotated unigenes onto pathways for carbohydrate and energy metabolism, nucleotide and amino acid metabolism, lipids, terpenes, lignin metabolism, and cofactor/vitamin biosynthesis (Figure 3). The number of genes involved in metabolic pathways was the largest for carbohydrate and energy metabolism, accounting for 666 unigenes in 35 pathways, followed by biosynthetic pathways for amino acids and nucleotides, which involved 440 unigenes in 23 pathways. The metabolic pathways for lipid, terpenoid, and lignin metabolism accounted for 277 genes in the 27 pathways. Biosynthesis of cofactors and vitamins accounted for 179 unigenes in the 12 pathways.

Several important genes were involved in the monolignol biogenesis pathway, including phenylalanine/tyrosine ammonia-lyase [EC:4.3.1.25], trans-cinnamate 4-monooxygenase [EC:1.14.14.91], 4-coumarate-CoA ligase [EC:6.2.1.12], shikimate O-hydroxycinnamoyltransferase [EC:2.3.1.133], 5-O-(4-coumaroyl)-D-

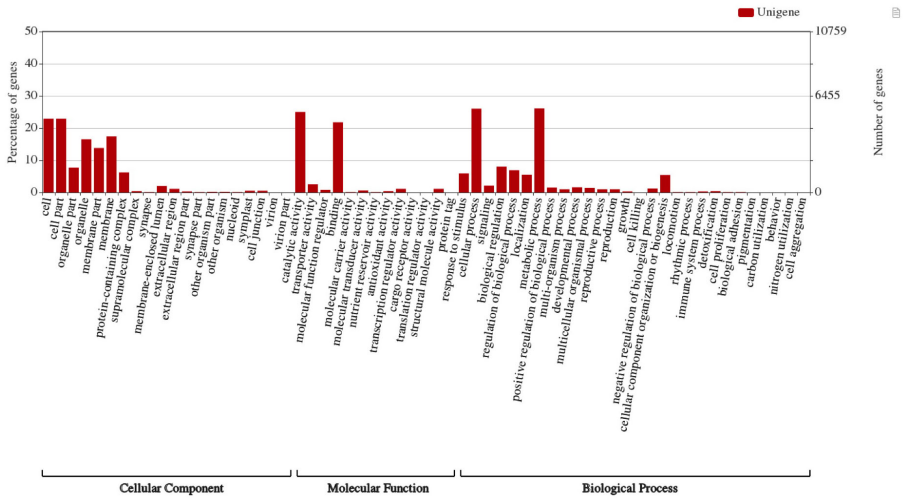


FIGURE 2
Gene ontology categories of assembled unigenes in *A. crassicarpa* callus shown by WEGO web application. GO terms were assigned in three processes regarding cellular component, molecular function and biological process.

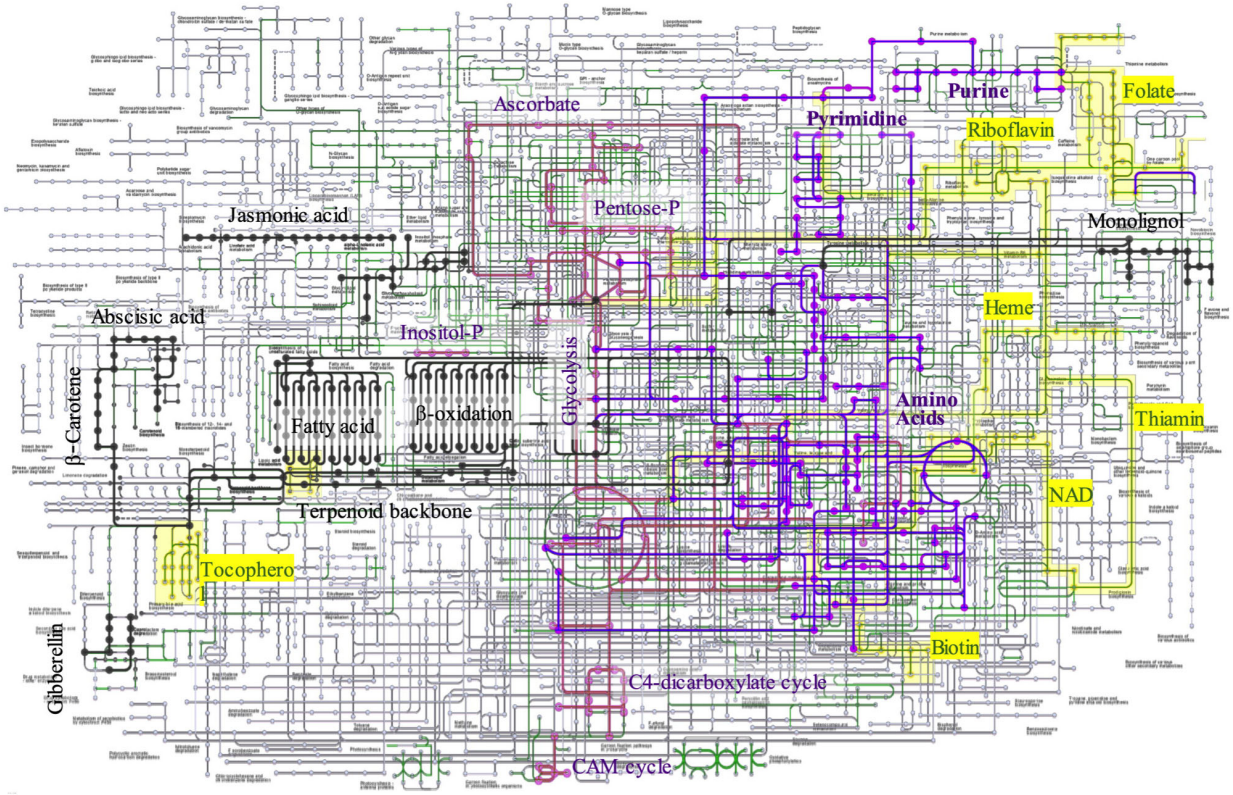


FIGURE 3
Metabolic pathway construction by KEGG mapping. The unigenes annotated to metabolic enzymes were reconstructed to KEGG metabolic map. Carbohydrate and Energy metabolism (pink), Nucleoside and amino acid metabolites (purple), lipid, terpen, and lignin (black), cofactors (yellow), and others (green).

quinate 3'-monooxygenase [EC:1.14.14.96], caffeoyl-CoA O-methyltransferase [EC:2.1.1.104], cinnamoyl-CoA reductase [EC:1.2.1.44], ferulate-5-hydroxylase [EC:1.14.-.-], and caffeic acid 3-O-methyltransferase/acetylserotonin O-methyltransferase [EC:2.1.1.68 2.1.1.4]. The monolignol biosynthetic pathway is well characterized, but the coordination and regulation of cell wall formation are not well understood. Several transcription factors may be involved in the formation of lignin and wood formation (Zhang et al., 2021).

3.5 Identification of putative transcription factors

Plant Transcription Factor Database (PlantTFDB 4.0) predicted 756 transcription factors that matched the homologues of A. thaliana (E-value < 10-5) (Table 2) and classified them into 40 families. The most abundant family was the bHLH13-like protein (78 unigenes), followed by the ethylene-responsive element (ERF) (62 unigenes), MYB family (55 unigenes), WRKY DNA-binding proteins (54 unigenes), bZIP family (53 unigenes), and lysine-specific histone demethylase (53 unigenes), as well as others that were lower in number. It has been reported that major transcription factors, ERF (Ikeuchi et al., 2022), MYB (Sakamoto et al., 2022), bHLH and WRKY (Xu et al., 2023a), play pivotal roles in reprogramming in response to physical injury and hormone signaling.

3.6 Unigenes involved in phytohormone signaling

Acacia crassicarpa expressed genes involved in all known plant hormone signaling pathways (Figure 3). Unigenes encoding components of auxin (AUX), gibberellin (GA), and jasmonic acid (JA) signaling were annotated. Most of the unigenes involved in cytokinin (CK) and abscisic acid (ABA) signaling were also annotated, except for the B-ARR transcription factor in CK signaling, and EIN2 and BSU1 were involved in phosphor-relaying components. Signaling components for ethylene (ET), brassinosteroids (BR), and salicylic acid (SA) were all expressed in the undifferentiated cells of A. crassicarpa.

3.7 SSR analysis for polymorphism

The assembled unigene sequences of A. crassicarpa were subjected to MISA to identify SSR representing di-, tri-, and tetra-nucleotide repeats. Dinucleotide SSRs (AG, AT, and AC) were more frequent than trinucleotide SSRs (Figure 4). Among the trinucleotide repeat motifs, the dominant repeat motif was (AAG)n, followed by (AGG)n, (AAT)n, (ATC)n, and the other trinucleotide repeats. Dinucleotide repeats occur more frequently than tri- or tetranucleotide repeat units. Among the dinucleotide repeat motifs, the most frequent repeat detected in this study was AG/CT, followed by AT/TA, and AC/GT. The least frequent

10.3389/fpls.2024.1339958

TABLE 2 Annotation of putative transcription factors.

Description	Unigenes
Transcription factor bHLH13-like	78
Ethylene Responsive Element	62
MYB family protein	55
WRKY DNA-binding protein	54
bZIP family protein	53
Lysine-specific histone demethylase (C2H2)	53
MYB-related family protein	48
CCCH-type zinc finger protein	33
Trihelix	33
FAR1 family protein	30
GRAS family protein	30
G2-like Family Protein	28
Nuclear Factor Y	26
Prosopis alba B3 domain-containing protein	19
AP2-like ethylene-responsive transcription factor	16
DOF family protein	15
Homeodomain-like transcriptional regulator	13
GATA transcription factor	12
M-type_MADS	12
TCP Family Protein	10
BEACH domain-containing protein B	8
SBP family protein	7
Prosopis alba two-component response regulator ARR2-like	6
DBB family protein	6
DNA-binding storekeeper protein-related transcriptional regulator	6
Prosopis alba protein BASIC PENTACYSTEINE6-like	5
CO-like	5
DP-E2F-like 1	5
CPP	4
BES1/BZR1 homolog protein 2	3
CAMTA	3
LSD family protein	3
VOZ protein	3
WOX Family protein	3
Ethylene-Insensitive Protein	2
Pathogenesis-related homeodomain protein-like	2
RAV Family Protein	2
Prosopis alba growth-regulating factor 3-like	1
Prosopis alba protein Effector of Transcription 2-like	1
ZF-HD	1
	756

dinucleotide was GC/CG. This finding is consistent with those of previous studies conducted on Stevia rebaudiana (Kaur et al., 2015), the traditional Chinese medicinal plant Epimedium sagittatum (Zeng et al., 2010) and coffee (Aggarwal et al., 2007).

Primers were designed to detect polymorphisms by targeting unigenes that contained triplet repeats in their ORFs (Table 3). Among the 16 DNA samples from local A. crassicarpa trees, three multiple signals of 258, 287, and 290-bp were observed for the SSR marker designed for the unigene ICSL01027786 (293bp), which was annotated as auxin response factor 3 (ARF3). Three signals were also produced from the unigene ICSL01000311 (266 bp), which was annotated as ARF4 (Table 4). Other unigenes annotated as transcription factors, such as ICSL01001643 (254bp), ICSL01023085 (175 bp), and ICSL01008756 (216 bp), also produced multiplicity in SSR analysis. Unigene ICSL01001643, annotated as a putative ethylene-responsive transcription factor (Charfeddine et al., 2019), produced two PCR bands with lengths of 214 and 277 bp. Unigene ICSL01023085 encodes a proline-rich PRCC (Weterman et al., 2000) and exhibits polymorphism with four PCR bands of 197, 200, 249, and 257 bp. Unigene ICSL01008756 was annotated as tubby-like F-box protein 5 (Song et al., 2019) (Xu et al., 2016);, and its polymorphisms appeared as PCR products with lengths of 235, 238, and 241 bp. Moderate multiplicity giving signals of 179 and 182-bp was observed for unigene ICSL01034178 (154 bp), which was annotated as a pollen tube guidance gene (Li et al., 2015). The polymorphic information content (PIC), which represents the ability to detect polymorphisms, was calculated using Cervus (Kalinowski, 2007), and was high (PIC > 0.4) for ARF3, ARF4, PRCC, and Tubby-like F-box protein 5, and moderate (PIC < 0.4) for ERTF and PTG.

4 Discussion

4.1 Informative foundation for the molecular breeding of *A. crassicarpa*

Molecular genetics provides valuable information for developing programs for efficient tree breeding. Several commercially important traits of trees, such as fast growth, few branches, high strength, and resistance to pathogens, are likely to be controlled by multiple layers of gene regulation. Although these traits are difficult to control by environmental management such as optimized fertilization, several genes are involved in the commercially important traits of various plants. The present study reports a transcriptomic analysis utilizing next-generation sequencing, which offers an effective approach to genomic resources

10.3389/fpls.2024.1339958

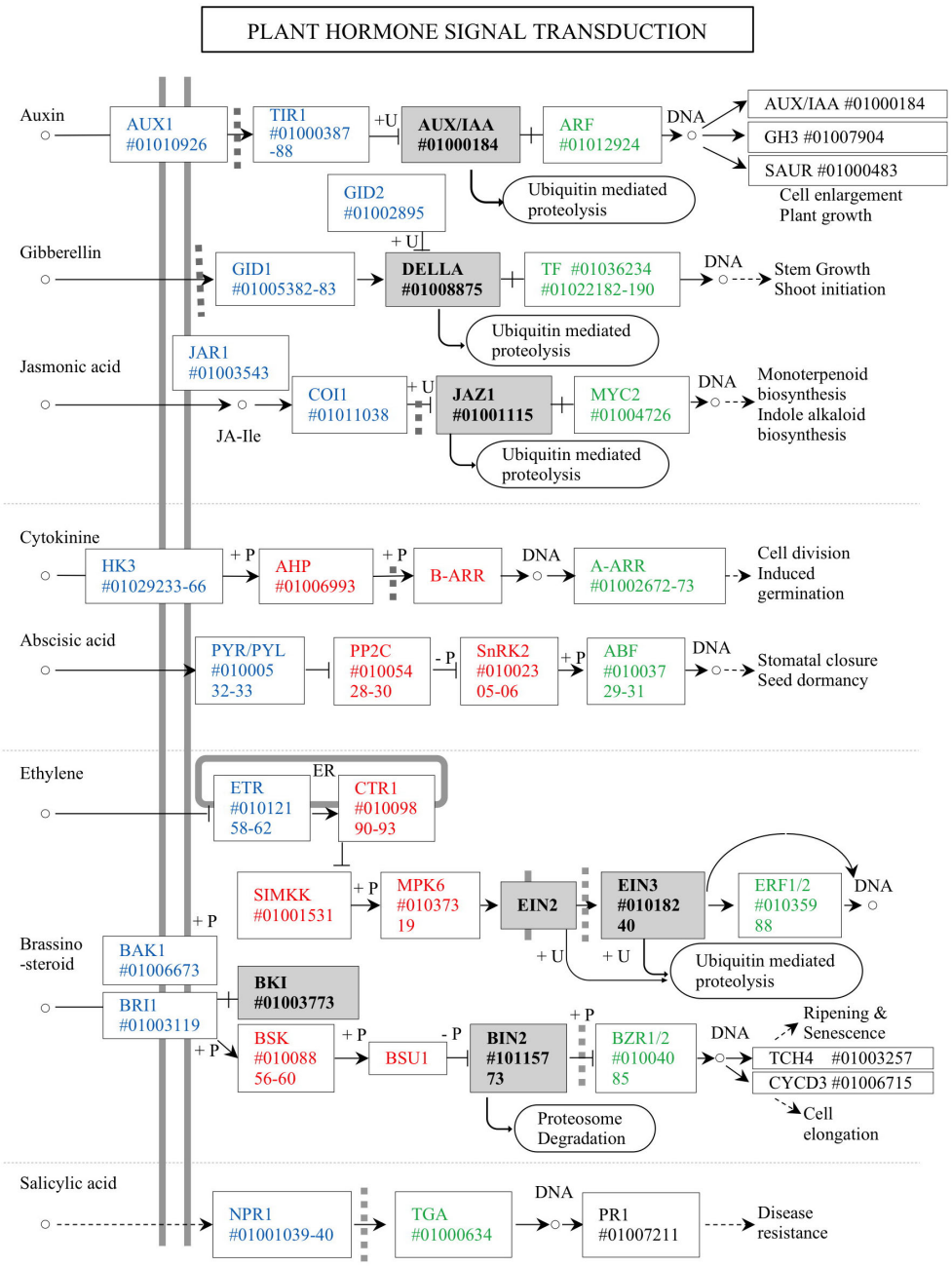


FIGURE 4
Unigenes annotated for the plant hormone signaling in the *A. crassicarpa* transcriptome. Signal transduction components are classified as ligand receptors (blue), Phospho-relay components (red), transcription factor regulators (gray), and transcription factors (green). The unigene identities are specified as ICSL numbers starting with a # symbol.

at a reasonable cost in a relatively short period, even for non-model tree species.

4.2 Annotation of unigenes for metabolic enzymes and transcription factors

Annotation of genes encoding metabolic enzymes and mapping on the KEGG metabolic pathway provides an atlas that describes the metabolic flow in the pluripotent cells of A. crassicarpa. Potential targets, such as those involved in monolignol biosynthesis, were among the genes involved in the primary metabolic pathways. This genetic information may offer a candidate for selective breeding and genome editing programs to increase the pulp yield and reduce pulping costs. Annotation of transcription factors also provides another set of candidates to be disrupted by genome-editing technology, thereby modifying the texture of lignin content and hormone responses, including pathogen resistance.

TABLE 3 PCR primers for SSR marker amplification.

Target gene ID		Primers
c26532_gl_il	F	gtttcttACGATGAAACACCGTCGTCT
	R	gttttcccagtcacgacgt CA CGACAGAACACGTGAAATGG
c11193_gl_i1	F	gtttcttCCCAGCACACATTCAGAAGA
	R	gttttcccagtcacgacgt GI CAAAGCACAAGACGTGAAGO
c13147_g2_i1	F	gtttctTCTTATTCAACCCCTGAGGC
	R	gttttcccagtcacgacgt CA GAGAAAGACATTGACGGGGA
c25518_g1_i1	F	gtttctTCCCTCCCCAAACCTAAATC
	R	gttttcccagtcacgacgt TG CGTTTTCTGCTTTACGCTCC
c20030_g1_i1	F	gtttcTTGGATAGTCAGCTTCCGCT
	R	gttttcccagtcacgacgt CA CCCTTCTTGGATGGCAGATA
c27691_gl_i1	F	gtttcttGCCTTGCTTCGATCTTGTTC
	R	gttttcccagtcacgacgt GT ACTTGGATCGGCTAAAGGGT

The forward primer (F) has a pigtail motif (5'-gtttctt-3') at the 5' end. The reverse primer (R) has a U-19 universal reverse sequence (5'-gttttcccagtcacgacgt-3') that is labelled with four kinds of fluorescent molecules (6-FAM, VIC, NED, and PET). The Labelled U-19 primer was also tagged with 2 bp barcodes at the 3′ end (TG for 6-FAM, AC for VIC, CA for NED, and GT for PET).

The cellular reprogramming process involved several genetic perturbations induced by physical injury of explant (Deng et al., 2022) (Park et al., 2023b), and supplementation of auxin and cytokinin (Asghar et al., 2023). Elucidation of the gene regulatory mechanism, involving several transcription factors (Kumar et al., 2021, Dai et al., 2020, Xu et al., 2023b), signaling pathway components (Park et al., 2023a), and epigenetic alteration of chromatin (Chen et al., 2023) may give us clue how to reverse the transformed callus back to the differentiated plant body.

4.3 Plant hormone signaling as the genetic switch

Phytohormones can provide an effective means of operating genetic switches from outside the hard bark with reproducible and quantitative responses. We considered the nature of these switches in terms of their reversibility, dose responsiveness, and mechanisms that enable crosstalk with other hormones. The three phytohormones auxin (AUX), gibberellin (GA), and jasmonic acid (JA) exploit ubiquitin-dependent proteolysis, which eliminates the specific regulator proteins that negatively control transcription factors. The elimination of negative regulators allows transcription factors to activate hormone-responsive gene expression. The switchover to the signal-responding state continues until lifted restrictions are regenerated through the biosynthesis of negative regulators. Therefore, hormone regulation via AUX, GA, and JA could provide a binary on/off switch.

Cytokinin (CK) and abscisic acid (ABA) signaling is transmitted through a phosphorylation relay between receptors and regulators. Phosphorelay-dependent regulation involves a series of phosphorylation events that lead to transcription factor activation. Phosphorylation, typically catalyzed by protein kinases, can occur in multiple steps, creating a phosphorylation gradient. The extent of phosphorylation determines the activity level of the transcription factor, and consequently, the strength of the signaling response. Accordingly, this gradient regulation allows fine-tuning of the response, enabling different levels of gene expression based on the concentration or duration of the hormone signal.

Ethylene (ET) and brassinosteroid (BR) signaling constitute the phosphorylation relay upstream; however, proteolytic removal of the transcriptional regulator is involved downstream. Such a hybrid mechanism can allow a quantitative response to intermediate signaling components, which may mediate crosstalk with other hormone signaling (Yasumura et al., 2015; Wang et al., 2020; Zhao et al., 2023), while the proteolytic switchover to gene expression provides a determinative response to hormone stimuli by the accumulated signaling of ET (Chen et al., 2022) or that of BR (Kour et al., 2021). In the systemic immune response induced by salicylic acid, the response to the SA receptor NPR1, which undergoes transformation from an oligomeric state in the cytosol to monomers upon binding to SA, activates pathogenesis-related genes in the nucleus (Palmer et al., 2019). The straightforward transmission of SA signaling allows a quick response to pathogen invasion.

4.4 SSR marker development

Individual trees may exhibit slight but discernible differences in growth rate, lignin content, stress tolerance, and pest resistance. The genetic traits underlying these characteristics may be associated with molecular genetic markers, which provides a scientific foundation for describing these tree characteristics. Short sequence repeats (SSR) are nucleotide regions that represent tandem repeat units located in coding and non-coding regions of the genome (Gupta et al., 1996).

None of the 16 subspecies of A. crassicarpa was identical in terms of the PCR products amplified by pairs of primers flanking the repeated trinucleotide sequences. The highest similarity in the EST-SSR products was observed between subspecies 05 and 11, which differed only in the PCR products from unigene ICSL01034178, which was annotated as a pollen tube guidance protein. The highest difference in polymorphism was observed between subspecies 03 and 04, which exhibited diverse multiplicities, except for ICSL01008756 and ICSL01034178. Polymorphisms in these transcription genes may exert a significant effect on the set of genes expressed under these regulators, which may be relevant to the physiological diversity of the 16 subspecies of A. crassicarpa.

Although the predominant trinucleotide repeats contained adenine-leading motifs such as AAG, AGG, AAT, and ATC (Figure 5), polymorphisms were produced when targeting repeats started with G or C, such as CTT, CAT, and GCA (Table 4). The

10.3389/fpls.2024.1339958

TABLE 4 Polymorphism represented as the length distribution of PCR products focused on SSR regions.

Contig	ICSL01027786			ICSL01000311			ICSL01001643		ICSL01023085				ICSL01008756			ICSL01034178	
Repeat	CTT x 5			CAT x 5			TGC x 5		GAC x 6				GCT x 5			CTT x 5	
Annotation	ARF3			ARF4			ERTF		Pro-rich PRCC				Tubby-like F-box protein 5			PTG	
Tree 01	258	287			286	292	214	277	197		249		235	238		179	
Tree 02	258		290			292	214	277	197		249			238			182
Tree 03		287				292		277	197			257		238		179	182
Tree 04	258	287		280		292	214	277		200	249			238		179	182
Tree 05	258	287			286	292	214	277	197		249			238			182
Tree 06	258		290			292	214	277	197		249			238			182
Tree 07	258	287		280	286		214	277			249	257	235		241	179	
Tree 08	258	287			286	292	214	277	197		249		235	238		179	
Tree 09	258	287			286	292	214	277		200	249			238		179	182
Tree 10	258		290		286	292	214	277	197		249		235	238		179	182
Tree 11	258	287			286	292	214	277	197		249			238		179	
Tree 12	258	287	290			292	214		197		249		235			179	
Tree 13	258		290			292	214	277		200	249		235		241	179	
Tree 14	258	287	290		286	292	214	277	197		249			238		179	
Tree 15	258	287	290	280	286		214	277		200	249			238		179	182
Tree 16	258	287			286	292	214	277		200	249		235		241		182
H_o	0.938			0.688			0.938		1				0375			0.313	
He	0.655			0.558			0.514		0.675				0.514			0.498	
PIC	0.561			0.464			0.374		0.593				0.436			0.366	
F(Null)	-0.223			-0.1104			-0.306		-0.2309				0.1934			0.2138	
Position	CDS			CDS			CDS		CDS				CDS			CDS	

The microsatellite locus was amplified by PCR using a primer pair flanking the in-frame SSR region. Polymorphisms were observed in contigs encoding auxin response factors (ARFs), ethylene response transcription factor (ERTF), and pollen tube guidance protein (PTG). The length of the PCR products presented in the table was measured by size calibration using the LIZ-600 standard marker. Ho, observed heterozygosity; He, expected heterozygosity; PIC, polymorphic information content; F (null), estimated null allele frequency; CDS, Coding sequence for protein.

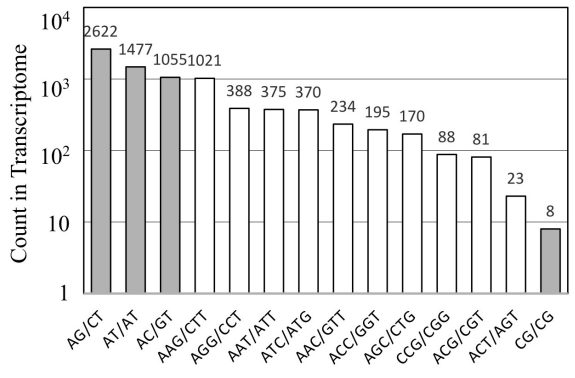


FIGURE 5
The distribution of SSRs identified from the *A. crassicarpa* transcriptome. Occurrence of repeat sequences with di- and tri-nucleotide units are designated in gray and white, respectively.

high occurrence of repeated units does not necessarily guarantee successful targeting by SSR fingerprinting.

SSR fingerprinting can be produced by PCR amplification of SSR on assembled transcriptomic sequences (transcriptome-SSRs), ESTs (EST-SSRs), or genomic DNA sequences (genomic SSRs). Transcriptome SSR sequences are more accessible than genomic SSR markers because they can focus on gene-rich regions and are associated with transcription. Transcriptome-SSR markers contribute to the segregation of closely related subspecies and variants, and these markers may be used in breeding programs to improve the quality and biomass of trees as a fiber resource and to achieve the desired properties, such as the ability to lower the salt content in combustion ash when used as biomass fuel (Thiel et al., 2003; Chung et al., 2014).

In our future studies, gene co-expression networks may be constructed using time-resolved transcriptomic approaches as

illustrated for maize (Du et al., 2019) soybean (Park et al., 2023a) and cotton (Fan et al., 2022), which will enable the clustering of gene expression at each phase of the cellular reprogramming process from the differentiated plant body to the pluripotent state. The molecular genetic sequence information established in the present study will assist in the molecular breeding of *Acacia* trees with SSR markers and functional annotation of genes and assignment to metabolic map pathways.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ddbj.nig.ac.jp/, SAMD00213802, https://www.ddbj.nig.ac.jp/, PRJDB9506, https://www.ddbj.nig.ac.jp/, DRX208625, https://www.ddbj.nig.ac.jp/, DRR218312.

Author contributions

SI: Conceptualization, Writing – review & editing. KK: Writing – review & editing, Data curation, Investigation. MN: Investigation, Writing – review & editing, Methodology. TK: Investigation, Methodology, Writing – review & editing. TT: Investigation, Conceptualization, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study received funding from Sumitomo Forestry Co. Ltd, and funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. The study was also supported by the Adaptable and Seamless Technology Transfer Program through Target-driven R&D (A-STEP) from the Japan Science and Technology Agency (JST) Grant Number JPMJTM20QX21445979. All authors declare no other competing interests.

Conflict of interest

Author SI and KK are employed by Sumitomo Forestry Co. Ltd.

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

Publisher's note

References

Aggarwal, R. K., Hendre, P. S., Varshney, R. K., Bhat, P. R., Krishnakumar, V., and Singh, L. (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor. Appl. Genet.* 114, 359–372. doi: 10.1007/s00122-006-0440-x

Ali, M., Hussain, R. M., Rehman, N. U., She, G. B., Li, P. H., Wan, X. C., et al. (2018). *De novo* transcriptome sequencing and metabolite profiling analyses reveal the complex metabolic genes involved in the terpenoid biosynthesis in Blue Anise Sage (Salvia guaranitica L.). *DNA Res.* 25, 597–617. doi: 10.1093/dnares/dsy028

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/s0022-2836(05)80360-2

Asghar, S., Ghori, N., Hyat, F., Li, Y., and Chen, C. (2023). Use of auxin and cytokinin for somatic embryogenesis in plant: a story from competence towards completion. *Plant Growth Regul.* 99, 413-428. doi: 10.1007/s10725-022-00923-9

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinf. (Oxford England)* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Blakesley, D., Allen, A., Pellny, T. K., and Roberts, A. V. (2002). Natural and induced polyploidy in *Acacia dealbata* Link. and *Acacia mangium* Willd. *Ann. Bot.* 90, 391–398. doi: 10.1093/aob/mcf202

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31, 365–370. doi: 10.1093/nar/gkg095

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Charfeddine, M., Charfeddine, S., Ghazala, I., Bouaziz, D., and Bouzid, R. G. (2019). Investigation of the response to salinity of transgenic potato plants overexpressing the transcription factor StERF94. *J. Biosci.* 44, 141. doi: 10.1007/s12038-019-9959-2

Chen, H., Bullock, D. A., Alonso, J. M., and Stepanova, A. N. (2022). To fight or to grow: the balancing role of ethylene in plant abiotic stress responses. *Plants-Basel* 11, 33. doi: 10.3390/plants11010033

Chen, Y., Hung, F., and Sugimoto, K. (2023). [Epigenomic reprogramming in plant regeneration: Locate before you modify]. *Curr. Opin. Plant Biol.* 75, 102415. doi: 10.1016/j.pbi.2023.102415

Chung, J. W., Kim, T. S., Sundan, S., Lee, G. A., Park, J. H., Cho, G. T., et al. (2014). New cDNA-SSR markers in the narrow-leaved vetch (Vicia sativa subsp nigra) using 454 pyrosequencing. *Mol. Breed* 33, 749–754. doi: 10.1007/s11032-013-9980-3

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

Dai, X., Liu, N., Wang, L., Li, J., Zheng, X., Xiang, F., et al. (2020). MYB94 and MYB96 additively inhibit callus formation via directly repressing LBD29 expression in *Arabidopsis thaliana*. *Plant Sci.* 293, 110323. doi: 10.1016/j.plantsci.2019.110323

Deng, J., Sun, W., Zhang, B., Sun, S., Xia, L., Miao, Y., et al. (2022). GhTCE1-GhTCEE1 dimers regulate transcriptional reprogramming during wound-induced callus formation in cotton. *Plant Cell.* 34, 4554-4568. doi: 10.1093/plcell/koac252

Du, X., Fang, T., Liu, Y., Huang, L., Zang, M., Wang, G., et al. (2019). Transcriptome profiling predicts new genes to promote maize callus formation and transformation. *Front. Plant Sci.* 10, 1633. doi: 10.3389/fpls.2019.01633

Fan, Y., Tang, Z., Wei, J., Yu, X., Guo, H., Li, T., et al. (2022). Dynamic transcriptome analysis reveals complex regulatory pathway underlying induction and dose effect by different exogenous auxin IAA and 2,4-D during in vitro embryogenic redifferentiation in cotton. *Front. Plant Sci.* 13, 931105. doi: 10.3389/fpls.2022.931105

Fuller, A., Omidiji, Y., Viefhaus, T., Maier, J., and Scheffknecht, G. (2019). The impact of an additive on fly ash formation/transformation from wood dust combustion in a lab-scale pulverized fuel reactor. *Renewable Energy* 136, 732–745. doi: 10.1016/j.renene.2019.01.013

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Gupta, P. K., Balyan, I. S., Sharma, P. C., and Ramesh, B. (1996). Microsatellites in plants: A new class of molecular markers. *Curr. Sci.* 70, 45–54. doi: 10.2307/24097472

Hallgren, A. L., Engvall, K., and Skrifvars, B. J. (1999). Ash-induced operational difficulties in fluidised bed firing of biofuels and waste. *Biomass: Growth Opportunity Green Energy Value-Added Products* 1, 1365–1370. doi: 10.1016/j.fuel.2006.06.020

Ikeuchi, M., Iwase, A., Ito, T., Tanaka, H., Favero, D. S., Kawamura, A., et al. (2022). Wound-inducible WUSCHEL-RELATED HOMEOBOX 13 is required for callus growth and organ reconnection. *Plant Physiol.* 188, 425–441. doi: 10.1093/plphys/kiab510

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kaur, R., Sharma, N., and Raina, R. (2015). Identification and functional annotation of expressed sequence tags based SSR markers of Stevia rebaudiana. *Turk J. Agric. For.* 39, 439–450. doi: 10.3906/tar-1406-144

Kour, J., Kohli, S. K., Khanna, K., Bakshi, P., Sharma, P., Singh, A. D., et al. (2021). Brassinosteroid signaling, crosstalk and, physiological functions in plants under heavy metal stress. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.608061

Kumar, S., Ruggles, A., Logan, S., Mazarakis, A., Tyson, T., Bates, M., et al. (2021). Comparative transcriptomics of non-embryogenic and embryogenic callus in semi-recalcitrant and non-recalcitrant upland cotton lines. *PLANTS-BASEL.* 10, 1775. doi: 10.3390/plants10091775

Lawson, S. S., and Ebrahimi, A. (2018). Development and validation of Acacia koa and A. koaia nuclear SSRs using Illumina sequencing. *Silvae Genetica* 67, 20–25. doi: 10.2478/sg-2018-0003

Li, H. J., Zhu, S. S., Zhang, M. X., Wang, T., Liang, L., Xue, Y., et al. (2015). Arabidopsis CBP1 is a novel regulator of transcription initiation in central cell-mediated pollen tube guidance. *Plant Cell* 27, 2880–2893. doi: 10.1105/tpc.15.00370

McKinnon, G. E., Larcombe, M. J., Griffin, A. R., and Vaillancourt, R. E. (2018). Development of microsatellites using next-generation sequencing for Acacia crassicarpa. *J. Trop. For. Sci.* 30, 252–258. doi: 10.26525/jtfs2018.30.2.252258

Midgley, S., and Turnbull, J. (2003). Domestication and use of Autralian Acacias: case studies of five important species. *Aust. Syst. Bot.* 16, 89–102. doi: 10.1071/SB01038

Mukherjee, S., and Sharma, A. K. (1995). In situ nuclear DNA variation in Australian species of Acacia. *Cytobios*. 75, 33-36.

Nishimura, O., Hara, Y., and Kuraku, S. (2017). gVolante for standardizing completeness assessment of genome and transcriptome assemblies. *Bioinformatics* 33, 3635–3637. doi: 10.1093/bioinformatics/btx445

Palmer, I. A., Chen, H., Chen, J., Chang, M., Li, M., Liu, F. Q., et al. (2019). Novel salicylic acid analogs induce a potent defense response in arabidopsis. *Int. J. Mol. Sci.* 20, 3356. doi: 10.3390/ijms20133356

Pan, Z., and You, Y. (1994). Introduction and provenance test of *Acacia crassicarpa*. *For. Res.* 7, 498–505.

Park, J., Choi, Y., Jeong, M., Jeong, Y., Han, J., and Choi, H. (2023a). Uncovering transcriptional reprogramming during callus development in soybean: insights and implications. *Front. Plant Sci.* 14, 1239917. doi: 10.3389/fpls.2023.1239917

Park, J., Park, K., Park, S., Ko, S., Moon, K., Koo, H., et al. (2023b). WUSCHEL controls genotype-dependent shoot regeneration capacity in potato. *Plant Physiol.* 193, 661-676. doi: 10.1093/plphys/kiad345

Pashley, C. H., Ellis, J. R., McCauley, D. E., and Burke, J. M. (2006). EST databases as a source for molecular markers: lessons from Helianthus. *J. Hered* 97, 381–388. doi: 10.1093/jhered/esl013

Sakamoto, Y., Kawamura, A., Suzuki, T., Segami, S., Maeshima, M., Polyn, S., et al. (2022). Transcriptional activation of auxin biosynthesis drives developmental reprogramming of differentiated cells. *Plant Cell.* 34, 4348-4365. doi: 10.1093/plcell/koac218

Song, Z. B., Wu, X. F., Gao, Y. L., Cui, X., Jiao, F. C., Chen, X. J., et al. (2019). Genome-wide analysis of the HAK potassium transporter gene family reveals asymmetrical evolution in tobacco (Nicotiana tabacum). *Genome* 62, 267–278. doi: 10.1139/gen-2018-0187

Sugimoto, K., Jiao, Y. L., and Meyerowitz, E. M. (2010). *Arabidopsis* regeneration from multiple tissues occurs via a root development pathway. *Dev. Cell* 18, 463–471. doi: 10.1016/j.devcel.2010.02.004

Suzuki, S., Suda, K., Sakurai, N., Ogata, Y., Hattori, T., Suzuki, H., et al. (2011). Analysis of expressed sequence tags in developing secondary xylem and shoot of Acacia mangium. *J. Wood Sci.* 57, 40–46. doi: 10.1007/s10086-010-1141-2

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

Thiel, T., Michalek, W., Varshney, R. K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor. Appl. Genet.* 106, 411–422. doi: 10.1007/s00122-002-1031-0

Travisany, D., Ayala-Raso, A., Di Genova, A., Monsalve, L., Bernales, M., Martinez, J. P., et al. (2019). RNA-Seq analysis and transcriptome assembly of raspberry fruit (Rubus idaeus "Heritage") revealed several candidate genes involved in fruit development and ripening. *Sci. Hortic.* 254, 26–34. doi: 10.1016/j.scienta.2019.04.018

Vergara-Pulgar, C., Rothkegel, K., González-Agüero, M., Pedreschi, R., Campos-Vargas, R., Defilippi, B. G., et al. (2019). *De novo* assembly of Persea americana cv. 'Hass' transcriptome during fruit development. *BMC Genomics* 20, 108. doi: 10.1186/s12864-019-5486-7

Wang, Q., Yu, F. F., and Xie, Q. (2020). Balancing growth and adaptation to stress: Crosstalk between brassinosteroid and abscisic acid signaling. *Plant Cell Environ.* 43, 2325–2335. doi: 10.1111/pce.13846

Weterman, M. A. J., von Groningen, J. J. M., Jansen, A., and van Kessel, A. G. (2000). Nuclear localization and transactivating capacities of the papillary renal cell carcinoma-associated TFE3 and PRCC (fusion) proteins. *Oncogene* 19, 69–74. doi: 10.1038/sj.onc.1203255

Xu, C., Chang, P., Guo, S., Yang, X., Liu, X., Sui, B., et al. (2023). Transcriptional activation by WRKY23 and derepression by removal of bHLH041 coordinately establish callus pluripotency in Arabidopsis regeneration. *Plant Cell*. 7, 1693. doi: 10.1093/plcell/koad255

Xu, J. N., Xing, S. S., Zhang, Z. F., Chen, X. S., and Wang, X. Y. (2016). Genome-wide identification and expression analysis of the tubby-like protein family in the malus domestica genome. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01693

Yasumura, Y., Pierik, R., Kelly, S., Sakuta, M., Voesenek, L., and Harberd, N. P. (2015). An ancestral role for CONSTITUTIVE TRIPLE RESPONSE1 proteins in both ethylene and abscisic acid signaling. *Plant Physiol.* 169, 283. doi: 10.1104/pp.15.00233

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., et al. (2018). WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* 46, W71–W75. doi: 10.1093/nar/gky400

Yong, S. Y. C., Choong, C. Y., Cheong, P. L., Pang, S. L., Amalina, R. N., Harikrishna, J. A., et al. (2011). Analysis of ESTs generated from inner bark tissue of an Acacia auriculiformis x Acacia mangium hybrid. *Tree Genet. Genomes* 7, 143–152. doi: 10.1007/s11295-010-0321-y

Yusiharni, E., and Gilkes, R. (2012). Minerals in the ash of Australian native plants. *Geoderma* 189, 369–380. doi: 10.1016/j.geoderma.2012.06.035

Zeng, S., Xiao, G., Guo, J., Fei, Z., Xu, Y., Roe, B. A., et al. (2010). Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics* 11, 94. doi: 10.1186/1471-2164-11-94

Zhang, B. C., Gao, Y. H., Zhang, L. J., and Zhou, Y. H. (2021). The plant cell wall: Biosynthesis, construction, and functions. *J. Integr. Plant Biol.* 63, 251–272. doi: 10.1111/jipb.13055

Zhao, X. Y., Zhang, T. R., Bai, L., Zhao, S. S., Guo, Y., and Li, Z. (2023). CKL2 mediates the crosstalk between abscisic acid and brassinosteroid signaling to promote swift growth recovery after stress in Arabidopsis. *J. Integr. Plant Biol.* 65, 64–81. doi: 10.1111/jipb.13397

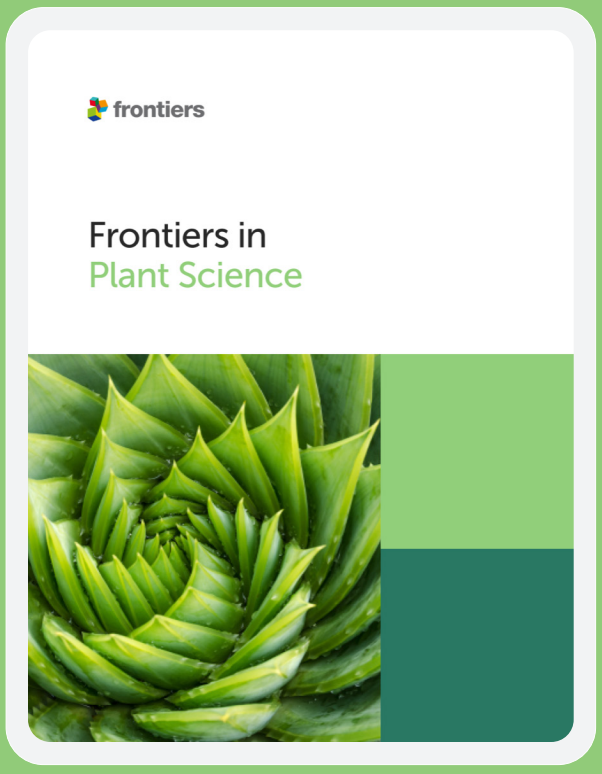
Frontiers in
Plant Science

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