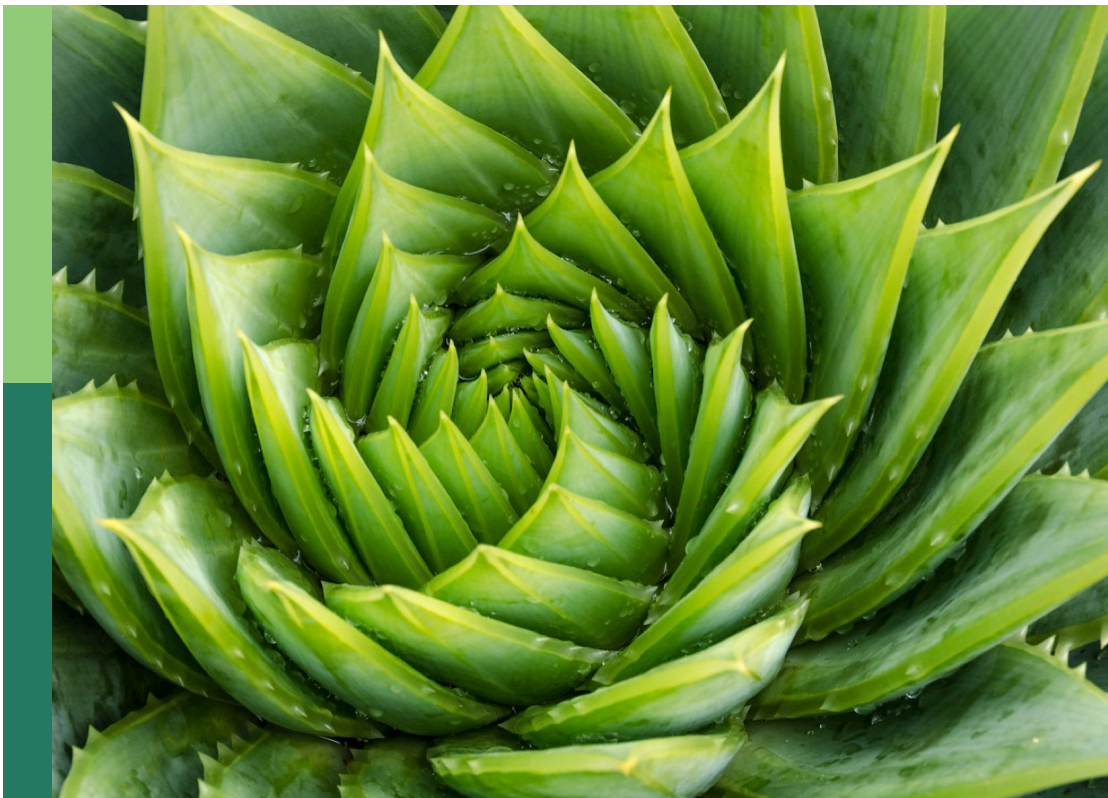
frontiers | Research Topics

# Advances on genomics and genetics of horticultural crops and their contribution to breeding efforts, volume II

**Edited by**
Eleni Tani, Aliki Xanthopoulou and Christos Bazakos

**Published in**
Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

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

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

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

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

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

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

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

ISSN 1664-8714
ISBN 978-2-8325-4569-0
DOI 10.3389/978-2-8325-4569-0

## 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

# Advances on genomics and genetics of horticultural crops and their contribution to breeding efforts, volume II

**Topic editors**

Eleni Tani — Agricultural University of Athens, Greece

Aliki Xanthopoulou — Institute of Plant Breeding and Genetic Resources, Hellenic Agricultural Organisation (HAO), Greece

Christos Bazakos — Max Planck Institute for Plant Breeding Research, Germany

**Citation**

Tani, E., Xanthopoulou, A., Bazakos, C., eds. (2024). *Advances on genomics and genetics of horticultural crops and their contribution to breeding efforts, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4569-0

# Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 27 February 2024
DOI 10.3389/fpls.2024.1385217

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Jihong Hu,
Northwest A&F University, China

*CORRESPONDENCE
Christos Bazakos
bazakos@mpipz.mpg.de

RECEIVED 12 February 2024
ACCEPTED 19 February 2024
PUBLISHED 27 February 2024

CITATION
Tani E, Xanthopoulou A and Bazakos C (2024) Editorial: Advances on genomics and genetics of horticultural crops and their contribution to breeding efforts - volume II.
*Front. Plant Sci.* 15:1385217.
doi: 10.3389/fpls.2024.1385217

COPYRIGHT
© 2024 Tani, Xanthopoulou and Bazakos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Editorial: Advances on genomics and genetics of horticultural crops and their contribution to breeding efforts - volume II

Eleni Tani[1], Aliki Xanthopoulou[2,3] and Christos Bazakos[2,3,4]*

[1]Laboratory of Plant Breeding and Biometry, Agricultural University of Athens, Athens, Greece, [2]Institute of Plant Breeding and Genetic Resources, ELGO-DIMITRA, Thessaloniki, Greece, [3]Joint Laboratory of Horticulture, ELGO-DIMITRA, Thessaloniki, Greece, [4]Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany

KEYWORDS

**omics approaches, phylogenetics, breeding, genome editing, bioinformatics**

Editorial on the Research Topic
Advances on genomics and genetics of horticultural crops and their contribution to breeding efforts - volume II

## 1 Introduction

Horticultural crops harbor a multitude of beneficial attributes for human use. Research focus has shifted to horticultural crops with the goal of not only boosting production and quality but also resilience and sustainability (Lastochkina et al., 2022). The Research Topic of these manuscripts reports the recent advancements of -omic technologies, genome wide association studies as well as genome editing and their contribution in modern horticultural breeding.

Zhang et al., conducted a preparatory study of the mechanism of selenium accumulation and transportation in tea cultivars. Cai et al., used transcriptome analysis in order to study how fern gametophytes respond to different light conditions. Under the same framework, Scandola et al. through proteome and metabolome analysis, classified kale cultivars into two groups based on their amino acid and sugar content, as well as carbon and nitrogen metabolism, mRNA splicing, protein translation and light harvesting.

Hussain et al., conducted a genome-wide identification analysis of NRAMP gene family in *Kandelia obovate* and enhanced our understanding of the roles of *Kandelia obovata* NRAMPs in copper stress conditions. In the same aspect, Zhang et al., applied genome-wide identification analysis in order to explore PEBP gene family potential effect in flower development in pineapple. Another genome-wide identification study in *Ananas comosus* identified that GRF proteins are involved in the regulation of floral organ development and the response to gibberellin (Yi et al.).

Tripodi et al. described the creation of the first-ever SPET panel in lettuce and its use in population structure and genomic diversity analysis. By focusing on the cabbage CENH3 gene linked to haploid induction, Stajič and Kunej aimed to improve the chances for PEG-mediated protoplast transformation to increase editing efficiency using CRISPR/Cas9 in a non-model horticultural plant. The review paper by Rosa-Martinez et al. documented the

research studies through both conventional and molecular breeding to increase the content of phenolic acids and flavonoids in three very important horticultural crops, namely tomato, eggplant and pepper. To provide insights into the evolutionary dynamics and phylogenetic relationships within Hordeum genus, Yuan et al. performed the sequencing, assembly and annotation of the chloroplast (cp) genomes of three wild perennial Hordeum species. In a similar manner, Lu and Li utilized a hybrid assembly strategy to assemble and annotate the mitogenome of *Syzygium samarangense* and enriched our understanding of Wax apple genetic features providing important information for its molecular breeding.

## 2 Molecular analyses through -omic approaches

Ferns, an ancient lineage of land plants, play a pivotal role in the alternation of generations (metagenesis) by nourishing the young sporophyte (Krieg and Chambers, 2022). One of the most important exogenous factors affecting plant growth is light. However, when a plant is exposed to excessive photo conditions, it can cause photostress (Takahashi and Badger, 2011). Cai et al., studied *Adiantum flabellulatum* gametophytes response to light and offered insights into the survival strategies of fern gametophytes in low-light environments. By employing transcriptome sequencing and gene expression analyses, the research has identified key genes and pathways involved in light sensitivity and photoprotection. These findings provide better understanding of the broader ecological dynamics of ferns and their evolutionary adaptations to shaded habitats.

Plants are important organic selenium sources for the human body. They uptake inorganic selenium and convert it into absorbable organic selenium (Ren et al., 2022). Tea plant (*Camellia sinensis* (L.) O. Kuntze) has a high ability to enrich Se (Ren et al., 2022). The manuscript of Zheng et al., provides a comprehensive analysis of selenium accumulation and transportation in different tea cultivars. Using transcriptome analysis, it unravels certain molecular mechanisms behind the varying abilities of tea plants to accumulate selenium. The study highlights key genes and pathways involved in this process, including those related to flavonoid biosynthesis and glutathione metabolism. The findings offer insights for agricultural practices and the health benefits of tea.

Kale contains a wealth of antioxidants, vitamins, minerals and fiber (Becerra-Moreno et al., 2014). The richness of its composition places it in the category of superfoods (Samec et al., 2019). Scandola et al., evaluated proteomics and metabolomics of diverse kale cultivars and marked a significant leap in the comprehension of kale's nutritional dynamics through a systems-level analysis. By examining nine kale cultivars under controlled LED light conditions, the research unveils how the diel molecular activities influence kale's growth, development, and nutritional content. With the use of plant phenotyping, proteomics, and metabolomics kale cultivars were categorized into two distinct groups based on their amino acid and sugar profiles. This study deepens our understanding of kale as a nutritious crop and paving the way for the adaptation of optimized cultivation strategies.

## 3 Genome-wide identification of plant gene families

Natural resistance-associated macrophage proteins (NRAMPs) are a class of metal transporters involved in metal uptake, transport and detoxification (Zhang et al., 2020; Tian et al., 2021). The study of Hussain et al., presents a detailed analysis of the NRAMP gene family in the mangrove species *Kandelia obovata*. It explores the genome-wide identification, structure, and expression of NRAMP genes, especially in response to varying levels of copper stress. This research enhances our knowledge of *Kandelia obovata* resilience mechanisms in harsh conditions, potentially aiding in the development of metal-tolerant crops.

The pineapple (*Ananas comosus* (L.) Merr.) is one of the most cultivated tropical fruits. Its flowering time exerts influence in yield and time of harvesting (Jin et al., 2021; Kim et al., 2022). PEBP gene family is involved in flower development as well as in the regulation of the flowering time (Jin et al., 2021; Kim et al., 2022). The manuscript of Zhang et al., delves into the PEBP gene family in pineapple (*Ananas comosus*). It identifies and categorizes 11 PEBP family members, exploring their phylogenetic relationships, chromosomal localization, and gene structure. Moreover, this research examines the expression patterns of these genes during pineapple flowering and flower development, highlighting the specific roles of certain genes, especially in response to ethylene treatment. These findings enrich our knowledge in understanding flowering mechanisms in pineapple.

Growth-regulating factor (GRF) proteins have been shown to impact plant growth and developmental processes (Kim et al., 2012; Lee et al., 2017; Zhang et al., 2021). Yi et al., focused on the GRF (Growth-Regulating Factor) gene family in pineapple. The study identified eight GRF transcription factor genes, categorized them into five subfamilies and explored their phylogenetic relationships, chromosomal locations, and expression profiles. Furthermore, their potential roles in floral organ development and response to gibberellin were revealed, contributing to penetration of pineapple's growth, development, and stress responses. These results can be the springboard for further research on the regulatory mechanisms of GRF transcription factors during pineapple growth and development.

Single primer enrichment technology (SPET), provides the opportunity to carry out targeted genotyping of known polymorphisms and to find new random polymorphic loci and hundreds of unique SNPs (Scaglione et al., 2019). Tripodi et al. documented the development and use of the first SPET panel in lettuce and its utilization in population structure and genetic diversity analysis. A total of 81,531 SNPs were examined in a heterogeneous collection of 160 *Lactuca sativa* and *Lactuca serriola* accessions originating from different regions around the globe. The efficacy of SPET for GWAS was confirmed when associations were found in

chromosomal areas that had previously been documented to host candidate genes for four major agricultural characteristics of lettuce. These results demonstrated the efficacy of SPET in enabling a more thorough characterization of lettuce collections.

## 4 Genome editing in cabbage

Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated systems (CRISPR/Cas9) are unquestionably becoming an essential tool in plant breeding (Zhu et al., 2020). The work presented by Stajič and Kunej targeted the cabbage CENH3 gene linked to haploid induction to optimize the conditions for PEG-mediated protoplast transformation to improve editing efficiency using CRISPR/Cas9 in a non-model horticultural plant. To effectively transfer the CRISPR/Cas9 vector into cabbage protoplasts, important parameters influencing transformation effectiveness, including PEG4000 concentration, incubation duration, and plasmid quantity, were assessed. These results could be applied not only to successful genome editing of cabbage and other brassicas, but also to studies involving transitory transformation techniques in protoplasts (i.e gene function analysis and subcellular localization).

## 5 Breeding for secondary metabolites content

The review paper by Rosa-Martinez et al. emphasizes on the importance of two of the largest classes of secondary metabolites, namely phenolic acids and flavonoids due their crucial functions regarding plant biology (i.e post-harvest life), plant resilience and health benefits (Cosme et al., 2020). More specifically, it provides detailed information on the progress throughout the years concerning breeding for higher accumulation of phenolic acids and flavonoids in some of the most popular vegetables worldwide, pepper, tomato, and eggplant, where a great variation in their accumulation profile has been observed (Rosa-Martınez et al., 2021). A wide range of genetic and genomic technologies has led to the discovery of structural enzymes, annotated genes and QTLs involved in the biosynthesis and accumulation of flavonoids and phenolic acids. The usefulness of combining phenotypic variability and molecular tools through conventional breeding and genetic engineering procedures has been documented with detailed information. Finally, possible negative effects of breeding for higher levels of phenolics are also mentioned, i.e research that has linked a higher phenolic content to a lower fruit organoleptic quality.

## 6 Non-nuclear genome assemblies and annotations

The assembly and annotation of non-nuclear genomes (mitochondrial and chloroplastic) are important for advancing our understanding of plant biology, phylogenetic evolution and genetic diversity while providing valuable information for the molecular breeding and genetic improvement (Kersten et al., 2016; Xue et al., 2019; Wang et al., 2022; Zhang et al., 2023b).

Wax apple (*Syzygium samarangense*) is a commercially important fruit belonging to one of the world's most species-rich tree genera (Govaerts et al., 2008). Utilizing a hybrid assembly strategy, Lu and Li, successfully assembled and annotated a 530,242 bp circular mitogenome of *S. samarangense* cv Black Diamond (Zheng, 2011), revealing 61 unique genes and a plethora of genetic structures, including simple sequence, tandem and interspersed repeats. The in-depth exploration of the mitogenome's evolutionary trajectory, marked by significant genomic reorganization and the loss of key protein-coding genes (Lu and Li, 2024). Furthermore, the identification of 591 RNA editing sites, particularly those leading to the gain or loss of start and stop codons, underscores the complexity of genetic regulation within this species. By conducting comprehensive analyses, including phylogeny, collinearity, and RNA editing validation, this study enriches our understanding of *S. samarangense* genetic features and provides important information for its molecular breeding (Lu and Li).

The sequencing, assembly and annotation of the chloroplast (cp) genomes of three wild perennial Hordeum species: *H. bogdanii*, *H. brevisubulatum*, and *H. violaceum*, revealed significant sequence variations and identified a series of genetic markers and hotspot regions, providing insights into the evolutionary dynamics and phylogenetic relationships within Hordeum (Yuan et al.). The comparative genomic study between wild and cultivated annual species, enriched the cp genome database and provided a valuable resource for the development of molecular markers for phylogenetic analysis and conservation efforts offering a comprehensive framework for exploring phylogenetic evolution and population genetics in the genus Hordeum (Yuan et al.).

## Author contributions

ET: Writing – original draft, Writing – review & editing. AX: Writing – original draft, Writing – review & editing. CB: Writing – original draft, Writing – review & editing.

## Conflict of interest

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

## Publisher's note

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

# References


Becerra-Moreno, A., Alanis-Garza, P. A., Mora-Nieves, J. L., Mora-Mora, J. P., and Jacobo-Velazquez, D. A. (2014). Kale: an excellent source of vitamin c, pro-vitamin a, lutein and glucosinolates. *Cyta-J Food* 12, 298–303. doi: 10.1080/19476337.2013.850743

Cosme, P., Rodrıguez, A. B., Espino, J., and Garrido, M. (2020). Plant phenolics: Bioavailability as a key determinant of their potential health-promoting applications. *Antioxidants* 9, 1263. doi: 10.3390/antiox9121263

Govaerts, R. H. A., Sobral, M., Ashton, P., Barrie, F. R., Holst, B. K., Landrum, L., et al. (2008). *World checklist of myrtaceae* (London: Royal Botanic Gardens).

Jin, S., Nasim, Z., Susila, H., and Ahn, J. H. (2021). Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. *Semin. Cell Dev. Biol.* 109, 20–30. doi: 10.1016/j.semcdb.2020.05.007

Kersten, B., Faivre Rampant, P., Mader, M., Le Paslier, M. C., Bounon, R., Berard, A., et al. (2016). Genome sequences of Populus tremula chloroplast and mitochondrion: Implications for holistic poplar breeding. *PloS One* 11, e0147209. doi: 10.1371/journal.pone.0147209

Kim, G., Rim, Y., Cho, H., and Hyun, T. K. (2022). Identification and functional characterization of FLOWERING LOCUS T in Platycodon grandiflorus. *Plants (Basel)* 11, 325. doi: 10.3390/plants11030325

Kim, J. -S., Mizoi, J., Kidokoro, S., Maruyama, K., Nakajima, J., Nakashima, K., et al. (2012). Arabidopsis GROWTH-REGULATING FACTOR7 functions as a transcriptional repressor of abscisic acid– and osmotic stress–responsive genes, including DREB2A. *Plant Cell* 24, 3393–3405. doi: 10.1105/tpc.112.100933

Krieg, C. P., and Chambers, S. M. (2022). The ecology and physiology of fern gametophytes: A methodological synthesis. *Appl. Plant Sci.* 10, (2). doi: 10.1002/aps3.11464

Lastochkina, O., Aliniaeifard, S., SeifiKalhor, M., Bosacchi, M., Maslennikova, D., and Lubyanova, A. (2022). Novel approaches for sustainable horticultural crop production: advances and prospects. *Horticulturae* 8, 910. doi: 10.3390/horticulturae8100910

Lee, S. -J., Lee, B. H., Jung, J. -H., Park, S. K., Song, J. T., and Kim, J. H. (2017). Growth-regulating factor and grf-interacting factor specify meristematic cells of gynoecia and anthers. *Plant Physiol.* 176, 717–729. doi: 10.1104/pp.17.00960

Lu, G, and Li, Q. Complete mitochondrial genome of *Syzygium samarangense* reveals genomic recombination, gene transfer, and RNA editing events. *Front. Plant Sci.* (2024) 14, 1301164. doi: 10.3389/fpls.2023.1301164

Ren, H., Li, X., Guo, L., Wang, L., Hao, X., and Zeng, J. (2022). Integrative transcriptome and proteome analysis reveals the absorption and metabolism of selenium in tea plants (Camellia sinensis (L.) O. Kuntze). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.848349

Rosa-Martınez, E., García-Martínez, M. D., Adalid-Martínez, A. M., Pereira-Dias,, L., Casanova, C., Soler, E., et al. (2021). Fruit composition profile of pepper, tomato and eggplant varieties grown under uniform conditions. *Food Res. Int.* 147, 110531. doi: 10.1016/j.foodres.2021.110531

Samec, D., Urlic, B., and Salopek-Sondi, B. (2019). Kale (Brassica oleracea var. acephala) as a superfood: review of the scientific evidence behind the statement. *Crit. Rev. Food Sci.* 59, 2411–2422. doi: 10.1080/10408398.2018.1454400

Scaglione, D., Pinosio, S., Marroni, F., Centa, E., Di Fornasiero, A., Magris, G., et al. (2019). Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. *Ann. Bot.* 124, 543–551. doi: 10.1093/aob/mcz054

Takahashi, S., and Badger, M. R. (2011). Photoprotection in plants: a new light on photosystem II damage. *Trends Plant Sci.* 16, 53–60. doi: 10.1016/j.tplants.2010.10.001

Tian, W., He, G., Qin, L., Li, D., Meng, L., Huang, Y., et al. (2021). Genome-wide analysis of the NRAMP gene family in potato (Solanum tuberosum): Identification, expression analysis and response to five heavy metals stress. *Ecotoxicol. Environ. Saf.* 208, 111661. doi: 10.1016/j.ecoenv.2020.111661

Wang, N., Li, C., Kuang, L., Wu, X., Xie, K., Zhu, A., et al. (2022). Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2206076119. doi: 10.1073/pnas.2206076119

Xue, S., Shi, T., Luo, W., Ni, X, Iqbal, S, Ni, Z., et al. (2019). Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca* and *P. salicina*. *Hortic. Res.* 6, 89. doi: 10.1038/s41438-019-0171-1

Zhang, B., Tong, Y., Luo, K., Zhai, Z., Liu, X., Shi, Z., et al. (2021). Identification of GROWTH-REGULATING FACTOR transcription factors in lettuce (Lactuca sativa) genome and functional analysis of LsaGRF5 in leaf size regulation. *BMC Plant Biol.* 21, 1–13. doi: 10.1186/s12870-021-03261-6

Zhang, J., Zhang, M., Song, H., Zhao, J., Shabala, S., Tian, S., et al. (2020). A novel plasma membrane-based NRAMP transporter contributes to Cd and Zn hyperaccumulation in Sedum alfredii Hance. *Environ. Exp. Bot.* 176, 104121. doi: 10.1016/j.envexpbot.2020.104121

Zhang, S., Wang, J., He, W., Kan, S., Liao, X., Jordan, D. R., et al. (2023b). Variation in mitogenome structural conformation in wild and cultivated lineages of sorghum corresponds with domestication history and plastome evolution. *BMC Plant Biol.* 23, 91. doi: 10.1186/s12870-023-04104-2

Zheng, D. (2011). *Cultivation techniques for high quality and high yield of wax apple* (Beijing: China: Agricultural Science and Technology Press).

Zhu, H., Li, C., and Gao, C. (2020). Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* 21, 661–677. doi: 10.1038/s41580-020-00288-9

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 21 March 2023
DOI 10.3389/fpls.2023.1135237

Check for updates

OPEN ACCESS

EDITED BY
Aliki Xanthopoulou,
Hellenic Agricultural Organisation (HAO), Greece

REVIEWED BY
Pilar Soengas,
Spanish National Research Council (CSIC), Spain
Andrea Moglia,
University of Turin, Italy

*CORRESPONDENCE
Leandro Pereira-Dias
leapedia@etsiamn.upv.es

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

RECEIVED 31 December 2022
ACCEPTED 07 March 2023
PUBLISHED 21 March 2023

CITATION
Rosa-Martínez E, Bovy A, Plazas M, Tikunov Y, Prohens J and Pereira-Dias L (2023) Genetics and breeding of phenolic content in tomato, eggplant and pepper fruits.
*Front. Plant Sci.* 14:1135237.
doi: 10.3389/fpls.2023.1135237

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

# Genetics and breeding of phenolic content in tomato, eggplant and pepper fruits

Elena Rosa-Martínez[1], Arnaud Bovy[2], Mariola Plazas[1], Yury Tikunov[2], Jaime Prohens[1] and Leandro Pereira-Dias[1,3]*

[1]Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Valencia, Spain, [2]Plant Breeding, Wageningen University & Research, Wageningen, Netherlands, [3]Faculdade de Ciências, Universidade do Porto, Porto, Portugal

Phenolic acids and flavonoids are large groups of secondary metabolites ubiquitous in the plant kingdom. They are currently in the spotlight due to the numerous health benefits associated with their consumption, as well as for their vital roles in plant biological processes and in plant-environment interaction. Tomato, eggplant and pepper are in the top ten most consumed vegetables in the world, and their fruit accumulation profiles have been extensively characterized, showing substantial differences. A broad array of genetic and genomic tools has helped to identify QTLs and candidate genes associated with the fruit biosynthesis of phenolic acids and flavonoids. The aim of this review was to synthesize the available information making it easily available for researchers and breeders. The phenylpropanoid pathway is tightly regulated by structural genes, which are conserved across species, along with a complex network of regulatory elements like transcription factors, especially of MYB family, and cellular transporters. Moreover, phenolic compounds accumulate in tissue-specific and developmental-dependent ways, as different paths of the metabolic pathway are activated/deactivated along with fruit development. We retrieved 104 annotated putative orthologues encoding for key enzymes of the phenylpropanoid pathway in tomato (37), eggplant (29) and pepper (38) and compiled 267 QTLs (217 for tomato, 16 for eggplant and 34 for pepper) linked to fruit phenolic acids, flavonoids and total phenolics content. Combining molecular tools and genetic variability, through both conventional and genetic engineering strategies, is a feasible approach to improve phenolics content in tomato, eggplant and pepper. Finally, although the phenylpropanoid biosynthetic pathway has been well-studied in the Solanaceae, more research is needed on the identification of the candidate genes behind many QTLs, as well as their interactions with other QTLs and genes.

KEYWORDS

**phenylpropanoid pathway, QTLs, structural and regulatory genes, flavonoids, phenolic acids, breeding strategies, transcription factors, polyphenols**

# 1 Introduction

The Solanaceae family includes some of the world most important crops (Olmstead et al., 2008). Among them, tomato (*Solanum lycopersicum* L.), eggplant (*S. melongena* L.) and pepper (*Capsicum annuum* L.) are widely cultivated worldwide for their edible berries, providing a wealth of nutritional and health-promoting compounds to the human diet.

Over the last decades, we have witnessed remarkable developments in plant breeding, particularly regarding yield, pest and disease resistance, and fruit external quality. However, traits like flavor and nutritional quality have been neglected for the sake of higher yields and longer shelf-life (Tieman et al., 2017; Zhu et al., 2018). With the increasing amount of data linking a lower risk of disease incidence to a regular consumption of vegetables, consumers are shifting towards healthier foods. In this context, plant secondary metabolites have gained a lot of attention due to their bioactive role in the human body (Kaushik et al., 2016; Pott et al., 2019).

Phenolic compounds are a large and diverse group of secondary metabolites ubiquitous in the plant kingdom. They share an aromatic ring backbone with one or several hydroxyl groups or other substitutes, such as sugar molecules or organic acids (Vogt, 2010). Depending on the number of aromatic rings and on the structural elements that bind these rings together, phenolic compounds may fall into one of several classes: phenolic acids, flavonoids, tannins, stilbenes, lignans, coumarins, chromones and xanthones (Vogt, 2010; Pott et al., 2019). Herein we focused on two major groups, phenolic acids and flavonoids, as they are the most relevant in fruits of tomato, eggplant and pepper. Within the flavonoids we have intentionally left out anthocyanins, a major group of polyphenolic pigments with numerous implications in both plant metabolism and human health, hence an important target for breeders, due to the fact that they have been comprehensively discussed in a recent review regarding their biological function and genetic regulation in the Solanaceae fruits (Liu et al., 2018).

Phenolic acids share a benzene ring and a carboxyl group as a backbone and can be classified into derivatives of benzoic acid or cinnamic acid, depending on whether the aromatic ring has a carboxylic group (C6-C1 structure) or a propanoic acid (C6-C3 structure) attached to it, respectively (Figure 1). Different levels of hydroxylation and methoxylation of each basic structure will result in different compounds with different antioxidant capacity (Scarano et al., 2020). Hence, within the hydroxybenzoic acid derivatives, the most representative acids are gallic, ellagic, protochatechuic, *p*-hydroxybenzoic, vanillic and syringic. On the other hand, the major hydroxycinnamic acids are caffeic, ferulic, sinapic and *p*-coumaric acids, and their derivatives (Valanciene et al., 2020). Rather than in their free form, these compounds are usually found esterified with other organic acids (e.g., quinic, tartaric) or carbohydrates. The most relevant group of these conjugates are the esters of hydroxycinnamic acids with quinic acid, commonly called chlorogenic acids (Clifford et al., 2017). Chlorogenic acids, in particular 5-O-caffeoylquinic acid, along with other caffeic acid derivatives, constitute the most abundant type of phenolic acids in tomato, eggplant and pepper (Whitaker and Stommel, 2003; Marín et al., 2004; Slimestad and Verheul, 2009; Luthria et al., 2010; García-Valverde et al., 2013). In general, hydroxycinnamic acids are more abundant than hydroxybenzoic derivatives in plants, including the Solanaceae (Valanciene et al., 2020).

Flavonoids are the most diverse class of phenolics (Ku et al., 2020). All flavonoids share a C6-C3-C6 carbon backbone formed by a benzo-γ-pyrone structure (ring A) and a phenyl ring (ring B) bound by an oxygen-containing γ-pyrone ring (ring C), as shown in Figure 1. Based on the position at which the B ring is attached to the C ring, as well as the oxidation and saturation degree of the heterocyclic C ring, flavonoid molecules are classified into flavones, flavonols, flavanones, anthocyanidins, flavanols or isoflavones (Vogt, 2010; Ku et al., 2020). Structural variations and the ability to be modified enzymatically are the main reasons contributing to their unique properties and broad functional diversity (Vogt, 2010; Ku et al., 2020). Furthermore, flavonoids antioxidant capacity varies depending on the number and position of hydroxyl groups in the catechol B ring and their position on the pyran C ring, and with the presence of other functional groups on the molecule and their arrangement around the nuclear structure (Dias et al., 2021).

Phenolic acids and flavonoids have been extensively characterized in tomato, eggplant and pepper fruits of cultivated

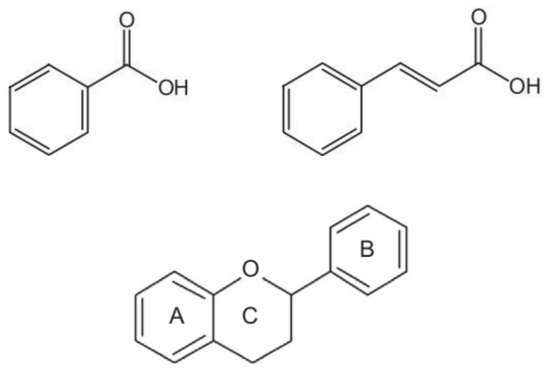



FIGURE 1
Basic structure of benzoic acids (top left), which has a carboxylic group attached to the aromatic ring (C6-C1 backbone), cinnamic acids (top right), which has a propanoic acid attached to the aromatic ring (C6-C3 backbone), and flavonoids (bottom), consisting of a benzo-γ-pyrone ring **(A)** and a phenyl ring **(B)** bound by an oxygen-containing γ-pyrone ring **(C)**.

varieties and wild relatives. Although these vegetables belong to the same family, their accumulation profile is substantially different (Rosa-Martínez et al., 2021).

## 1.1 Why breeding for higher levels of phenolic acids and flavonoids in tomato, eggplant and pepper?

Numerous health benefits have been linked to a phenolic-rich diet, namely related to both the antioxidant properties of phenolic compounds, which translate into their direct ability of reducing oxidant species, scavenging free radicals and chelating metal ions, and to their ability to modulate intracellular signaling molecules/pathways (Williams et al., 2004; Galano et al., 2016; Cosme et al., 2020). Thus, several studies have reported phenolic compounds to show anti-inflammatory, anti-microbial, anti-diabetic, anti-tumoral, and cardioprotective properties *in vitro* and in human trials. For instance, a major circulating human metabolite of quercetin has been proved to decrease the transcription of genes encoding pro-inflammatory interleukins and enzymes involved in oxidative stress responses, thus reducing the effects of atherosclerosis (Derlindati et al., 2012). Gallic acid had a potent effect on Herpes simplex virus type 1 and parainfluenza type 3 (Özçelik et al., 2011). The flavanones naringin and naringenin have been reported to enhance the expression of the insulin receptor GLUT4 and adiponectin in type II diabetic rats (Ahmed et al., 2017). Caffeic acid has demonstrated anti-proliferative and apoptotic effects on human melanoma, and colon, breast and liver cancer cells (Weng and Yen, 2012; Pelinson et al., 2019; Santana-Gálvez et al., 2020). Luteolin has been proven to prevent ischemia-reperfusion injury by reducing necrosis and apoptosis in rat cardiomyocytes and to induce arterial relaxation (Luo et al., 2017). Lastly, human cohorts have shown that chlorogenic acid (25-400 mg $d^{-1}$) may significantly reduce blood pressure in mild hypertensive adults (Mubarak et al., 2012; Kajikawa et al., 2019).

Furthermore, phenolics are important to an array of plant biological processes and to plant-environment interaction, such as protection against solar radiation, pathogens, herbivores and mechanical damage, attracting pollinators and seed dispersers, and abiotic stress signaling. Flavonoids are essential for male fertility in many species, so that they are found as part of the pollen grain wall structure (Wang et al., 2020; Dong and Lin, 2021). Down-regulation of the flavonoid biosynthesis has been shown to produce defective seeds due to impaired pollen-tube growth in tomato (Schijlen et al., 2007). Flavonols participate in the pollen's response to heat stress through scavenging of reactive oxygen species (ROS), thus maintaining its reproductive success (Muhlemann et al., 2018; Rutley et al., 2021). Flavonoids are abundantly present in the tomato fruit cuticle and epidermal cells. Low expression of a flavonoid biosynthesis key enzyme resulted in lower water permeability, lower stiffness and increased deformation of the cuticle (España et al., 2014; Heredia et al., 2015). Zhang et al. (2013) observed that anthocyanins-pigmented tomatoes had extended shelf life compared to anthocyanin-free ones, likely due to their antioxidant and ROS signaling properties. Levels of quercetin-O-rhamnoside-O-hexoside and chlorogenic acid remarkably increased in response to chilling injury, and induced stress tolerance in pepper (López-Velázquez et al., 2020). In *Arabidopsis thaliana*, flavonols have been reported to act as positional signals, integrating hormonal and ROS pathways to regulate root growth direction and rate in response to light (Silva-Navas et al., 2016; Tohge and Fernie, 2016). Also, the accumulation of anthocyanins in eggplant peel is induced by light exposure (Li et al., 2018a). In this respect, the photoprotective role against UV-B radiation of some phenolic compounds, such as anthocyanins, flavones, caffeic acid and chlorogenic acid have been widely demonstrated in several species (Sytar et al., 2018; Righini et al., 2019). Likewise, flavonoids and phenolic acids may act as a physical or chemical barrier to prevent invasion, or as a direct weapon against microbes and insects (Ramaroson et al., 2022). In tomato leaves, increased accumulation of phenolic acids was detected in response to *Pseudomonas syringae* infection (Dadáková et al., 2020), while higher accumulation of naringenin was associated with the resistance against *Botrytis cinerea* (Szymański et al., 2020).

All in all, considering that tomato, eggplant, and pepper are amongst the ten most consumed vegetables in the world, enhancing their phenolics content would lead to a promotion of the consumers health, while at the same time generating more resilient crops. In addition, the fact that some flavonoids lead to fruit pigmentation enables the development of new attractive colors, such as purple, generally associated with healthier produce, leading to diversification of the vegetable market. Finally, an increase in flavonoids content would improve postharvest quality and shelf-life. Consequently, bringing benefits to all the players in the production chain from farmers to consumers.

## 1.2 Which phenolic acids and flavonoids are found in the fruits of these crops?

In general, tomato accumulates more flavonoids than phenolic acids. The latter are mainly represented by caffeic acid and chlorogenic acid (Luthria et al., 2006; Hallmann and Rembialkowska, 2012; Martí et al., 2018), although gallic, ferulic, *p*-coumaric acids, and other derivatives, have been also identified in variable amounts in both the peel and pericarp (Rigano et al., 2014; Alarcón-Flores et al., 2016; Martí et al., 2018). Regarding flavonoids, tomato is known to be a good dietary source of naringenin chalcone and the flavonols rutin (quercetin-rutinoside), kaempferol-rutinoside and quercetin-trisaccharide (Muir et al., 2001; Slimestad et al., 2008), along with many other conjugated forms (Moco et al., 2006; Iijima et al., 2008; Alarcón-Flores et al., 2016). Within the tomato cultivated germplasm, flavonoids are accumulated almost exclusively in the fruit peel, while only trace amounts can be found in the fruit pericarp (Muir et al., 2001; Willits et al., 2005). Naringenin chalcone is the most abundant flavonoid in tomato, while rutin and kaempferol are accumulated at much lower amounts (Muir et al., 2001; Slimestad et al., 2008). Naringenin chalcone is also the pigment responsible for the transient yellow coloration of tomato fruit peel at breaker stage, during the transition from mature green to red ripe fruits (Adato et al., 2009; Ballester

et al., 2010). It is worth mentioning that although cultivated tomato fruits usually do not accumulate anthocyanins, nowadays, we can find purple tomato fruits due to the introgression of wild genes (Povero et al., 2011; Colanero et al., 2020).

Eggplant is one of the vegetables with the highest antioxidant capacity. Morales-Soto et al. (2014) ranked it among the top 5 fruits and vegetables out of 44 evaluated, in terms of total antioxidant capacity. Eggplant antioxidant capacity is empirically attributed to the phenolic acid profile in the fruit flesh, and especially to its high chlorogenic acid content, but also to the presence of anthocyanins in the purple fruit peel (Docimo et al., 2016; Rosa-Martínez et al., 2021). Some have also reported the presence of flavonol derivatives in the eggplant fruit peel. In this way, Singh et al. (2017) identified 11 flavonols, including different glycosides of quercetin, kaempferol and myricetin. Docimo et al. (2016) and Sulli et al. (2021) also quantified rutin in the fruit peel of a RIL population. However, only trace amounts were detected. The first comprehensive characterization of the eggplant phenolic acid profile was reported by Whitaker and Stommel (2003). The authors identified and quantified 14 major hydroxycinnamic acid conjugates in the fruits of seven commercial cultivars. All the compounds identified were esters of caffeic acid with quinic acid and derivatives. Since then, others have evaluated the profile of phenolic acids in fruits of cultivated and wild eggplant (Singh et al., 2009; Mennella et al., 2010; Sunseri et al., 2010; Mennella et al., 2012; Prohens et al., 2012; Mori et al., 2013; Plazas et al., 2014; Stommel et al., 2015). All demonstrated that chlorogenic acid is the most abundant phenolic in eggplant fruits. Indeed, within the cultivated pool, it typically represents 80-95% of the total phenolic acids content present in the fruit flesh.

Pepper has also been reported to have high antioxidant capacity, mainly due to its outstanding content in vitamin C (Morales-Soto et al., 2014). Nevertheless, phenolics of all classes have been identified and quantified in different varieties of pepper. Recently, Lemos et al. (2019) reviewed the data regarding pepper phenolic content and compiled it in a publication, making it readily accessible. Among hydroxybenzoic acids, the major compounds reported in pepper were gallic and vanillic acids (Mokhtar et al., 2015; Mudrić et al., 2017; Moreno-Ramírez et al., 2018) and the most abundant hydroxycinnamic acid was generally chlorogenic acid (Hallmann and Rembialkowska, 2012; Fratianni et al., 2020), although *p*-coumaric, caffeic, ferulic, and sinapic acids, as well as several of their glycosides were also quantifiable (Marín et al., 2004; Mudrić et al., 2017; Fratianni et al., 2020). Regarding flavonoids, the most represented subfamilies in pepper fruits are flavonols and flavones (Liu et al., 2020). Within the first group, quercetin and its glycosides are the most common, followed by myricetin and kaempferol, while within flavones, luteolin and its derivatives are the most representative, followed by apigenin and its derivatives (Jeong et al., 2011; Wahyuni et al., 2011; Hallmann and Rembialkowska, 2012; Chen and Kang, 2013; Morales-Soto et al., 2013; Mokhtar et al., 2015; Fratianni et al., 2020; Ribes-Moya et al., 2020). In addition, cultivated pepper fruits may accumulate anthocyanins during the unripe stages, although they often degrade during the ripening process (Aza-González et al., 2013).

## 1.3 Genetic and genomic resources impact in phenolic acids and flavonoids studies

Recent advances in omics have enhanced the knowledge on the genetics of complex traits like fruit quality. Notwithstanding, not all crops have the same resources available. Tomato has been a model organism for basic and applied plant research for many years now, especially as a model for other Solanaceae. Consequently, numerous genetic and genomic tools have been developed for tomato since the late 1980's, such as isogenic mutant libraries (Menda et al., 2004) and several intra and interspecific mapping populations (Eshed and Zamir, 1995; Monforte and Tanksley, 2000). The tomato reference genome sequence was first published in 2012 for the *S. lycopersicum* 'Heinz 1706' cultivar (The Tomato Genome Consortium, 2012) and has since then been re-sequenced, corrected and re-annotated to the current SL4.0 version with the ITAG4.0, available at Sol Genomics Network database (https://solgenomics.net/). Furthermore, more than 900 high-quality genome sequences of cultivated tomato and its wild relatives, including a recently published pan-genome using 725 representative tomato accessions (Gao et al., 2019), have since been released.

In eggplant, the development of genomic tools started considerably later. However, in the last years, the available genomic information of eggplant has increased dramatically. Several factors have made this possible: 1) the advent of the next-generation sequencing technologies (Gebhardt, 2016); 2) the well-established synteny between the eggplant and tomato genomes (Doganlar et al., 2002; Wu et al., 2009; Gramazio et al., 2014; Rinaldi et al., 2016; Barchi et al., 2019); and, 3) the availability of intra and interspecific mapping populations with associated genetic linkage maps (Doganlar et al., 2002; Gramazio et al., 2014). The first point led to increasingly affordable whole-genome sequencing and in those terms, Hirakawa et al. (2014) published the first draft genome using 'Nakate-Shinkuro', a common Asian eggplant cultivar. Moreover, in the last years several high-quality eggplant genome assemblies with substantial increases of annotated genes have been released (Wei et al., 2020; Barchi et al., 2021; Li et al., 2021). This includes the first eggplant pan-genome, assembled with sequences from 26 accessions belonging to *S. melongena*, *S. incanum* and *S. insanum* (Barchi et al., 2021).

Despite its economic relevance, pepper remains a surprisingly understudied crop. For many years, pepper research relied almost entirely on $F_2$ mapping populations resulting from crosses between contrasting germplasm (Lee, 2019). The first interspecific genetic linkage map was published in 1988 using RFLPs (Tanksley et al., 1988). Since then, several intra and interspecific maps have been constructed to dissect traits of interest (Lee, 2019). Thus, these first maps were of paramount importance to shed light into the synteny between pepper and tomato genomes (Rinaldi et al., 2016) and enable marker-assisted selection and QTL dissection (Wahyuni et al., 2014; Lee, 2019). Recent technological advances have made it possible for researchers to have publicly-available high-quality genome sequences. The first pepper genomes were published in 2014 with the complete sequences of *C. annuum* cv. Serrano Criollo de Morelos 334 (CM334), *C. chinense* PI159236, *C. annuum* cv.

Zunla-1, and *C. annuum* var. *glabriusculum* (Kim et al., 2014; Qin et al., 2014). Since then, new genomes and improved versions of the CM334 accession have been released (Kim et al., 2017; Hulse-Kemp et al., 2018; Acquadro et al., 2020), including the first pepper pan-genome (Ou et al., 2018). Such powerful genomic resources have been successfully applied in many genome-wide association studies (Colonna et al., 2019; Siddique et al., 2019; Lee et al., 2020). However, pepper unusually large genome (~3.5Gb) and degree of repetitiveness (~80%), compared to other Solanaceae, have hampered the study of complex traits.

These tools have helped to identify QTLs and candidate genes associated with the biosynthesis of phenolic acids and flavonoids in tomato, eggplant and pepper. The aim of the following sections is to integrate and summarize the available literature, making it easily available for researchers and breeders.

## 2 Genetic basis of phenolic compounds biosynthesis and accumulation

First, a search throughout the most recent genome versions of tomato, eggplant and pepper, at the Sol Genomics platform, for annotated putative orthologue genes encoding key enzymes involved in the phenylpropanoid pathway, was conducted. For this purpose, a blastp was performed using the sequence of each verified protein from *S. lycopersicum*, or *Arabidopsis thaliana* when the former was not available, which were retrieved from the NCBI protein database (https://www.ncbi.nlm.nih.gov/protein/), and selected hits with an e-value threshold of $1{\cdot}e^{-10}$, score >200 and identity >70%. Secondly, a literature review was carried out to compile all relevant QTLs, and associated candidate genes, related to the phenolic compounds' biosynthesis and accumulation in Solanaceae fruits.

### 2.1 Biosynthesis of phenolic acids and flavonoids: Key annotated genes and structural enzymes of the phenylpropanoid pathway

All phenolic compounds ultimately stem from the shikimic acid pathway, which starts with phosphoenolpyruvate and D-erythrose-4-phosphate forming the C6 core aromatic ring with one carboxyl and three hydroxyl substituents (Pott et al., 2019).

Regarding the synthesis of hydroxybenzoic acids, two pathways have been described in plants. On one hand, the shikimate/chorismate pathway starts with two products of shikimic acid transformation: 3-dihydroshikimic acid (3-DHS) and chorismic acid. In this way, 3-DHS serves as a precursor of protocatechuic acid and gallic acid, which in turn is transformed into ellagic acid and other derivatives; while from chorismic acid *via* the intermediate isochorismic acid, salicylic acid and a wide range of dihydroxybenzoic acids (DHBA) are synthesized (Wildermuth, 2006). Alternatively, hydroxybenzoic acids can also be synthesized from phenylalanine metabolism, in this case by means of the C3 side chain shortening of the hydroxycinnamic C6-C3 structure. In this case, the benzoic acid would be synthesized from cinnamic acid, salicylic acid from *o*-coumaric acid, 4-hydroxybenzoic acid from *p*-coumaric acid, protocatechuic acid from caffeic acid, vanillic acid from ferulic acid and syringic acid from sinapic acid. A network of multiple paths has been proposed within that second pathway, although the contribution of each one remains unclear. For further information about the synthesis pathways of benzoic derived acids we'd like to refer to Widhalm and Dudareva (2015). Not all the enzymes involved in the synthesis of hydroxybenzoic acids have been identified. Nevertheless, there are several key enzymes, extensively studied in tomato, eggplant and pepper, which could be targeted to increase the production of phenolic acids (Figure 2; Supplementary Data S1).

The synthesis of phenylalanine from chorismic acid is considered the branching point that links the shikimate pathway to the phenylpropanoid pathway. The synthesis of *p*-coumaroyl-CoA from phenylalanine in a three-step reaction is considered the starting point of the synthesis of both hydroxycinnamic acids and flavonoids (Vogt, 2010). These initial steps involve three key enzymes: phenylalanine ammonia-lyase (PAL), which catalyzes the non-oxidative deamination of phenylalanine to cinnamic acid, cinnamate 4-hydroxylase (C4H), which catalyzes the subsequent formation of *p*-coumaric acid, and 4-coumaroyl-CoA ligase (4CL), which is involved in the synthesis of the next branching element along the pathway, *p*-coumaroyl-CoA (Marchiosi et al., 2020) (Figure 2; Supplementary Data S1).

From cinnamic acid, *p*-coumaric acid or *p*-coumaroyl-CoA the other major hydroxycinnamic acids (caffeic, ferulic and sinapic acids) are synthesized through a multi-step chain reaction, either by direct transformation or by their CoA-conjugated intermediates. The pool of enzymes catalyzing this series of reactions is also well characterized (Marchiosi et al., 2020). Again, three key enzymes are responsible for the biosynthesis of the following hydroxycinnamic acids: 4-coumarate 3-hydroxylase (C3H), which catalyzes the hydroxylation of *p*-coumaric acid to caffeic acid; caffeate O-methyltransferase (COMT), which is involved in both the subsequent formation of ferulic acid and, together with ferulate 5-hydroxylase (F5H), in the synthesis of sinapic acid. 4CL also catalyzes the synthesis of each hydroxycinnamic acid-CoA conjugate (Figure 2; Supplementary Data S1).

Starting from *p*-coumaroyl-CoA, two synthesis pathways have been proposed for chlorogenic acid (CGA). Again, three key enzymes catalyze these steps (Gramazio et al., 2014). In one branch, *p*-coumaroyl-CoA is converted to *p*-coumaroyl quinic acid by the enzyme hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyl transferase (HCT); then, the latter compound is hydroxylated to form CGA with the action of *p*-coumaroyl ester 3′-hydroxylase (C3'H). In a second branch, *p*-coumaroyl shikimic acid is synthesized from *p*-coumaroyl-CoA *via* HCT, followed by two-step transformation to caffeoyl-CoA *via* the intermediate caffeoyl shikimic acid and catalyzed by C3'H and HCT. Finally, caffeoyl-CoA is trans-esterified with quinic acid to form CGA *via* the enzyme hydroxycinnamoyl-CoA:quinate hydroxycinnamoyl transferase (HQT) (Figure 2; Supplementary Data S1).

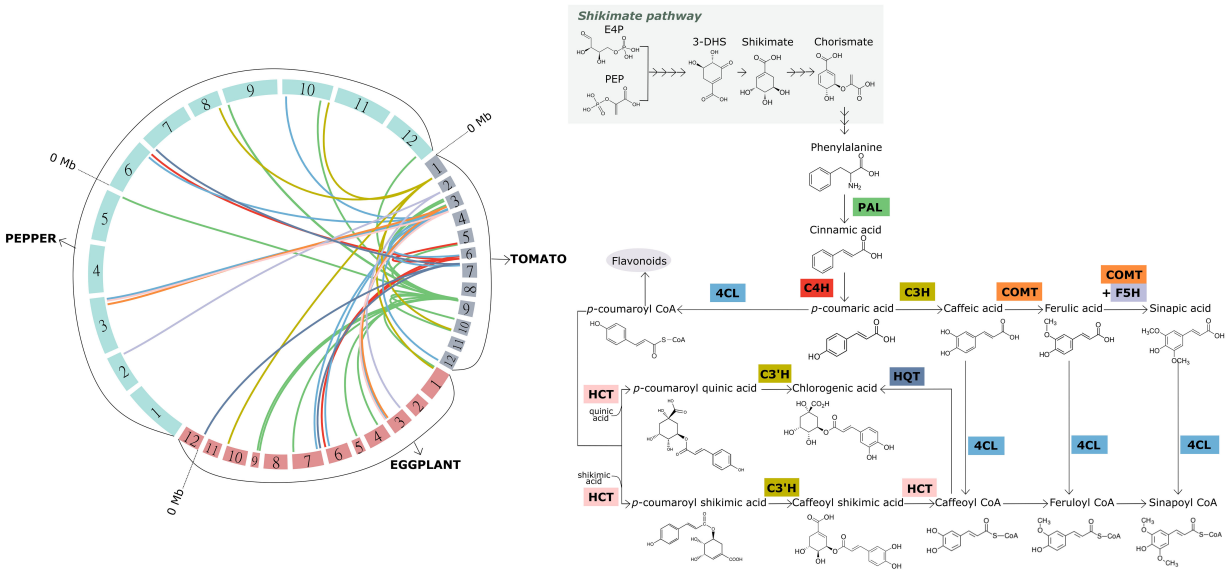



FIGURE 2
Macrosynteny among the tomato, eggplant and pepper genomes for candidate genes encoding the key enzymes involved in the synthesis of hydroxycinnamic acids (left), and overview of the aforementioned synthesis pathway (right). The 12 chromosomes of each of the three genomes represented, which are arranged to form the circle on the left, are colored differently according to the species. The chromosomes are scaled according to their size (Mb). Links with different colors inside the circle were used to connect the physical position of homologous genes within and among the genomes. Same color links represent genes encoding the same key enzyme of the pathway. The acronym of each key enzyme is indicated along the synthesis pathway (right), and framed in the same color as the corresponding link. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; C3H, 4-coumarate 3-hydroxylase; C3'H, p-coumaroyl ester 3'-hydroxylase; HCT, hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyl transferase; HQT, hydroxycinnamoyl-CoA:quinate hydroxycinnamoyl transferase; COMT, caffeate O-methyltransferase; F5H, ferulate 5-hydroxylase (F5H). The genes used for creating the links resulted from a blastp against the current protein databases of tomato, eggplant and pepper at the Sol Genomics platform (http://www.solgenomics.net). The hits with the lowest e-value, higher sequence size and higher sequence identity (>70%) were considered.

Efforts have been made in tomato towards the characterization of the CGA pathway genes through transgenesis, gene expression and enzymatic assays. In this way, gene silencing of an HQT-encoding gene resulted in a 98% reduction of CGA in tomato transgenic lines (Niggeweg et al., 2004), while no further soluble phenolics were affected, demonstrating that the encoded enzyme constitutes the primary route for the synthesis of CGA and controls the flux of the pathway (Niggeweg et al., 2004). More recently, several isomers of dicaffeoylquinic acids (di-CQAs), namely 1,3-, 3,4-, 1,5-, 3,5- and 4,5-di-CQA, have been identified in tomato fruits and their biosynthetic pathway has been elucidated (Moglia et al., 2014). The authors partially purified the enzyme responsible for the chlorogenate:chlorogenate transferase (CCT) activity, which was suggested to be HQT, and this role was confirmed through expression of *SlHQT* in *E. coli*. The authors comprehensively characterized the CCT activity and structural features of the HQT enzyme, as well as its subcellular localization, so that they determined the dual catalytic activity of the enzyme, which would act in the formation of CGA in the cytosol using CoA thioester as acyl donor, and as CCT in the formation of di-CQAs in the vacuole at lower pH and in the presence of high concentrations of CGA, using CGA as acyl donor (Moglia et al., 2014).

In eggplant, an interspecific anchored linkage map, based on a first backcross generation ($BC_1$) of *S. melongena* 'AN-S-26' × *S. incanum* 'MM577', was developed by Gramazio et al. (2014) and exploited for locating genes involved in the CGA synthesis pathway. Based on synteny with the Tomato EXPEN-2000 genetic linkage map, the authors mapped candidate genes encoding each of the six key enzymes involved in the CGA synthesis pathway in eggplant (*PAL, C4H, 4CL, HCT, C3'H, HQT*). They were located on linkage groups E09, E06, E03, E03, E01 and E07, respectively, and showed collinearity with the corresponding genes of the tomato genetic map (Gramazio et al., 2014). Once the sequences of key genes in the chlorogenic acid synthesis pathway were known, gene expression analyses were also performed in eggplant to characterize them and to understand their tissue-specific accumulation. Thus, Docimo et al. (2016) found the higher contents of CGA in fruits, compared to other tissues, to be correlated with elevated transcript abundance of the structural genes *PAL, C4H, 4CL* and *HQT*. They also isolated putative orthologs of the two CGA biosynthetic genes, *PAL* and *HQT*, from an Occidental *S. melongena* variety and demonstrated that both differed from homologs of Asiatic varieties. In addition, using a gene expression panel composed of 15 diverse *S. melongena* landraces and eight accessions of five related species of *Solanum* subgenus *Leptostemonum*, Meyer et al. (2019) characterized the genes encoding *HCT* and *HQT*. Their results suggested that in eggplant *SmHCT* was implicated in the synthesis of mono-caffeoylquinic acid (CQA) (CGA and its isomers 3-CQA, 4-CQA, 5-*cis*-CQA), while *SmHQT* was only catalyzing di-CQA formation. The biosynthesis of flavonoids has been extensively studied, hence, most of the key elements have been identified, especially in tomato (Tohge et al., 2017; Sacco et al., 2019; Tohge et al., 2020).

Briefly, naringenin chalcone (flavanone) is synthetized from *p*-coumaroyl-CoA and malonyl-CoA by chalcone synthase (CHS) and represents the main intermediate of flavonol biosynthesis. Naringenin chalcone is catalyzed by chalcone isomerase (CHI) and converted to the flavanone naringenin, which is a branching point for the formation of several groups of flavonoids. For example, naringenin can be converted to flavonols through the action of four key enzymes. First, flavanone-3-hydroxylase (F3H) converts naringenin into dihydrokaempferol (DHK), which is then converted into dihydroquercetin (DHQ) through the action of flavanone-3'-hydroxylase (F3'H). Finally, dihydroquercetin is catalyzed by flavonol synthase (FLS) in order to produce the flavonol quercetin. Likewise, the flavonol kaempferol is immediately synthesized from dihydrokaempferol through FLS catalysis, skipping hydroxylation at the C-3′ position (Figure 3; Supplementary Data S1).

Gene expression data of ripening tomato peel showed high levels of transcripts of most of the enzymes involved in the flavonoid biosynthesis (CHS, F3H and FLS) except for CHI, whose levels were below the detection limit (Muir et al., 2001; Bovy et al., 2002; Willits et al., 2005). This confirmed that the complete flavonoid pathway is present in the fruit peel but the regulation mechanism of CHI expression is rate-limiting the flavonol biosynthesis, probably due to a mutation in the CHI promotor sequence. In contrast, cultivated tomato flesh showed no detectable levels of any of the above-mentioned transcripts, indicating that the flavonoid biosynthesis pathway is inactive in this tissue (Muir et al., 2001; Bovy et al., 2002; Willits et al., 2005).

Furthermore, dihydrokaempferol can be hydroxylated at the 5' position of the B-ring in addition to the 3' position by the flavonoid-3',5'-hydroxylase (F3'5'H), converting DHK into dihydromyricetin (DHM), which is in turn transformed into the flavonol myricetin by FLS. DHM also acts as the first substrate leading to the anthocyanidin delphinidin, a purple-colored pigment. Both DHQ and DHK can also be converted into cyanidin and pelargonidin, respectively, the other two main anthocyanidins present in Solanaceae fruits. For the synthesis of these three anthocyanidins, DHM, DHQ and DHK, the enzymes dihydroflavonol 4-reductase (DFR) and subsequently anthocyanidin synthase (ANS) are required instead of flavonol synthase (Figure 3; Supplementary Data S1). Moreover, both cyanidin and its intermediate leucocyanidin can act as precursors for the formation of another subclass of flavonoids, the flavan-3-ols, mainly represented by catechin and epicatechin. However, the substrate specificity of DFR for dihydromyricetin leads to the occurrence of only the anthocyanidin delphinidin in these three crops. Delphinidin serves as precursor for the formation of anthocyanins, of which delphinidin-3-(*p*-coumaroyl-rutinoside)-5-glucoside is the most predominant in tomato, eggplant and pepper (Figure 3). The Solanaceae anthocyanin biosynthetic pathway has been thoroughly analyzed in a recent review, hence, it is not discussed herein (Liu et al., 2018).

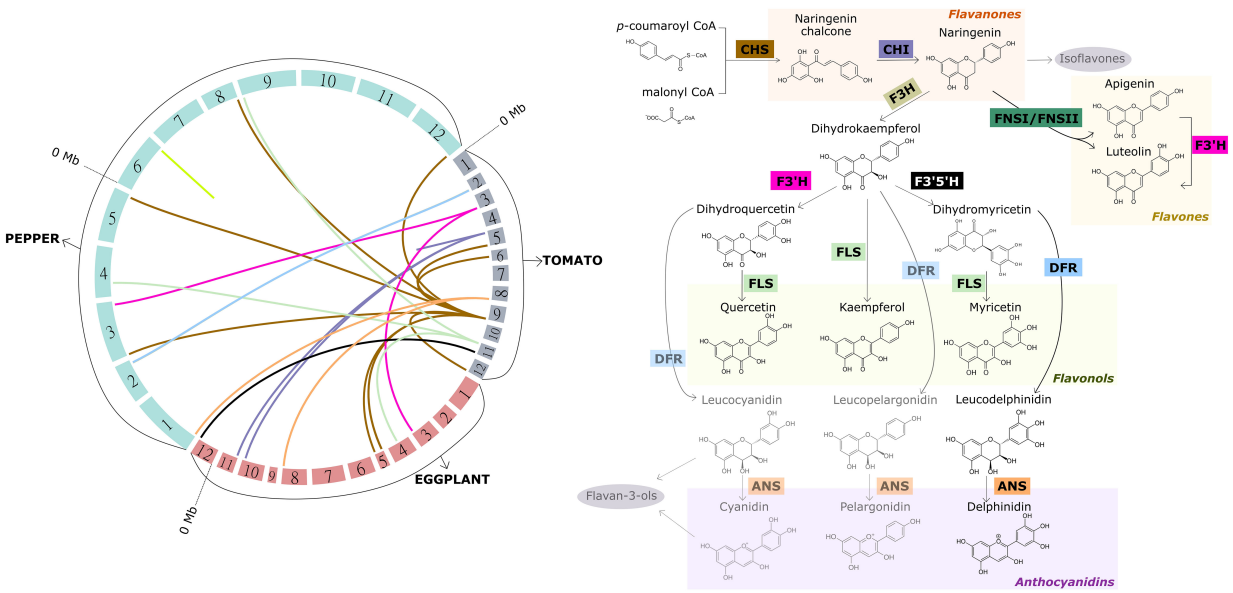



FIGURE 3
Macrosynteny among the tomato, eggplant and pepper genomes for candidate genes encoding the key enzymes involved in the synthesis of major flavonoids in fruits of these crops other than anthocyanins (left), and overview of the aforementioned synthesis pathway (right). The 12 chromosomes of each of the three genomes represented, which are arranged to form the circle on the left, are colored differently according to the species. The chromosomes are scaled according to their size (Mb). Links with different colors inside the circle were used to connect the physical position of homologous genes within and among the genomes. Same color links represent genes encoding the same key enzyme of the pathway. The acronym of each key enzyme is indicated along the synthesis pathway (right), and framed in the same color as the corresponding link. CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone-3-hydroxylase; F3'H, flavanone-3'-hydroxylase; F3'5'H, flavonoid-3',5'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; FNSI/FNSII, flavone synthase I/II. The genes used for creating the links resulted from a blastp against the current protein databases of tomato, eggplant and pepper at the Sol Genomics platform (http://www.solgenomics.net). The hits with the lowest e-value, higher sequence size and higher sequence identity (>70%) were considered.

Using four eggplant genotypes with contrasting peel pigmentation, their hybrids and a white-fruited line ('L131'), Lo Scalzo et al. (2021) studied the biochemical and gene expression changes in the peel during fruit development. This work demonstrated that the activation/deactivation of different branches of the phenylpropanoid pathway, occurs during ripening, resulting in a drastic drop of anthocyanins and CGA concentrations as ripening progresses from commercial to physiological maturity, with distinct phenolic compounds accumulating in the eggplant peel instead, mainly naringenin chalcone and naringenin 7-O-glucoside, along with other minor flavanones and flavonols, hence the change to a yellow coloration. Furthermore, the expression profiling of nine key genes (*SmelCHS_ch00, SmelCHS_ch05, SmelCHI_ch10, SmelCHI_ch05, SmelGT_ch01, SmelGT_ch05, SmelGT_ch10, SmelDFR_ch00 and SmelFLS_ch04)*, encoding enzymes involved in different paths of the phenylpropanoid biosynthesis, revealed changes consistent with the compositional results: *CHS* genes are expressed in eggplant fruit peel at all stages of maturity; a drop in the expression of *CHI* genes occurs at commercial and subsequent stages of maturity; a dramatic increase of glycosyltransferase (*GT*) genes expression during the last stages of physiological maturity may indicate their role in the glycosylation of naringenin and naringenin chalcone; and the *FLS* gene was strongly expressed at physiological maturity, as opposed to the *DFR* gene, being a strong candidate underlying the switch between the two branches of the phenylpropanoid pathway (Lo Scalzo et al., 2021).

A similar trend was reported for four pepper cultivars bearing different fruit colors. Liu et al. (2020) performed a targeted metabolome and transcriptome integrative study at two developmental stages. Results confirmed that flavonoid synthesis was developmental-dependent, as higher expression levels of structural genes of the phenylpropanoid pathway were observed at the immature stage for all four genotypes. Comparison of the metabolic profile with expression levels of structural genes, linked a higher expression of *PAL*, *C4H* and *4CL* to a higher accumulation of flavonols, flavones and anthocyanins in pepper fruits. A higher expression of *CHI* resulted in higher accumulation of naringenin, however, as reported by the authors, this was not translated into higher levels of downstream flavonoids due to the low expression of *F3H* and *F3'5'H* genes at green stage. This indicates the key role of these genes in regulating the flux to the lower branches of the phenylpropanoid pathway (Liu et al., 2020).

Naringenin can also act as a precursor for the other two main subclasses of flavonoids: flavones and isoflavones. The latter are mainly found in legumes, nuts and cereals and less relevant in tomato, eggplant and pepper, thus they are not discussed herein. Naringenin can be converted to the flavones apigenin and luteolin through the action of flavone synthases FNS-I and FNS-II, which have been characterized as a 2-oxoglutarate-dependent dioxygenase and a cytochrome P450 monooxygenase from the CYP93 family, respectively. Both enzymes catalyze the addition of a double bond between C-2 and C-3 in the heterocycle of naringenin (Scossa et al., 2019) (Figure 3; Supplementary Data S1).

It is worth mentioning that despite the structural genes being well conserved across these species, single or multiple mutations, deletions and other alterations to their sequences have been shown to alter significantly the metabolic composition and accumulation pattern in Solanaceae fruits (Chioti et al., 2022).

## 2.2 QTLs and candidate genes controlling phenolics accumulation

Many QTLs encompass the generalist trait of total phenolics content and total antioxidant capacity. Although several classes of compounds may be included in this category, we believe that it is relevant to include them in this review. Because we cannot classify them into specific categories, we kept the generalist designation. Table 1 summarizes all the QTLs found in the literature associated with phenolic compounds in tomato, eggplant and pepper. For more specific information on the compiled QTLs refer to Supplementary Data S2.

*Tomato*

Several interspecific tomato mapping populations have been used to study the genetics of phenolics, yielding an array of QTLs and candidate genes. The most studied mapping population is probably the one of *S. pennellii* (LA0716) in the genetic background of *S. lycopersicum* cv. 'M82' (LA3475) (Eshed and Zamir, 1995). Using this population, Alseekh et al. (2015) identified 67 and 30 robust ($P < 0.01$) metabolic QTLs (mQTLs) controlling hydroxycinnamates and flavonoids accumulation in the fruit pericarp, respectively (Table 1). The heritability of a pool of 23 UPLC-FTMS-identified hydroxycinnamates was high for 11 (e.g., homovanillic acid hexose II, caffeoyl-hexose II), intermediate for four, and low for eight. Likewise, the identified flavonoids showed high heritability, namely naringenin chalcone, with only two displaying intermediate and one low heritability. Furthermore, hydroxycinnamate QTLs showed a prevalent dominant negative mode of inheritance, with heterozygote lines showing lower levels of many compounds (e.g., coumaric acid hexose I and II, homovanillic acid hexose II), whereas flavonoids were reported to be split between dominant and additive negative QTLs. Few traits showed recessive mode of inheritance (Alseekh et al., 2015). The high frequency of negative mQTLs, as pointed out by the authors, suggests transcriptional regulation. In this way, 17 different families of transcription factors (TFs), detected by qRT-PCR, correlated to the expression of mQTLs, including MYB and MADS families (Table 1; Supplementary Data S2). The analysis of sequence polymorphisms of TF-encoding genes showed several variations between wild and cultivated sequences with potential impact on translation, stability, expression level, and tissue specificity (Alseekh et al., 2015). The regulation of the phenylpropanoid pathway is discussed in more detail in section 2.3.

Furthermore, Alseekh et al. (2015) identified 38 robust metabolic mQTLs controlling total phenolics content in the fruit pericarp scattered across the 12 chromosomes (Table 1; Supplementary Data S2). Likewise, Rousseaux et al. (2005) reported nine QTLs (*phe*3-2, *phe*3-3, *phe*3-4, *phe*5-4, *phe*6-2, *phe*7-2, *phe*7-4, *phe*8-2 and *phe*9-1) controlling total phenolics content in this population, although only one was found in both years of the experiment (*phe*6-2). The stable QTL increased

TABLE 1 Number of QTLs associated with phenolic acids, flavonoids (other than anthocyanins) and total phenolics content for tomato, eggplant and pepper.

| Species | Mapping population | Metabolite (s) | QTLs | Chromosome (s) | Candidate gene(s) (encoding protein) | Reference |
|---|---|---|---|---|---|---|
| **Tomato** | ILs *S. pennellii* (LA0716) x *S. lycopersicum* cv. 'M82' | Phenolic acids | 67 | [1-12] | Solyc10g085730, Solyc10g085860, Solyc10g085870, Solyc10g085880, Solyc10g086240 (UGT1, UDP-glycosyltransferase 1), Solyc10g086180 (PAL, phenylalanine ammonia lyase), Solyc12g099130, Solyc12g099120, Solyc12g099140, Solyc12g099620, Solyc12g099850 (MYB) | (Alseekh et al., 2015) |
| | | Flavonoids | 30 | [1-12] | Solyc06g083450 (O-methyltransferase), Solyc06g083470, Solyc06g083480, Solyc06g083490 (3-ketoacyl-acyl carrier protein reductases/tropan reductase), Solyc06g084240 (ent-copalyl diphosphate synthase) | (Alseekh et al., 2015) |
| | | Total phenolics | 38 | [1-12] | – | (Alseekh et al., 2015) |
| | | | 9 | 3, [5-9] | – | (Rousseaux et al., 2005) |
| | | | 1 | 7 | – | (Sacco et al., 2013) |
| | | | 1 | 7 | – | (Di Matteo et al., 2010) |
| | | | 1 | 7 | Solyc10g081260 (MATE, Multidrug resistance protein), Solyc06g005080 (VSP, Vacuolar protein sorting-associated protein 18), Solyc07g056420 and Solyc07g056510 (GST, Glutathione S-transferase), Solyc05g051200 (ERF1, Ethylene Responsive Factor 1) | (Di Matteo et al., 2013) |
| | BILs and ILs *S. pennellii* (LA0716) x *S. lycopersicum* cv. 'M82' | Phenolic acids | 3* | 1, 2, 10, 11 | – | (Szymański et al., 2020) |
| | | Flavonoids | 8* | 1, 5, 9 | Solyc05g010310.3 and Solyc05g010320.3 (CHI1, naringenin chalcone isomerase 1) | (Szymański et al., 2020) |
| | ILs *S. chmielewskii* (LA1840) x *S. lycopersicum* cv. 'Moneyberg' | Phenolic acids | 11 | 3, 4, 7, 10, 12 | Solyc07g005760 (HQT, Hydroxycinnamoyl CoA shikimate/quinate hydroxycinnamoyltransferase) | (Ballester et al., 2016) |
| | | Flavonoids | 23 | 2, [4-9], 12 | Solyc05g010320 (CHI1, naringenin chalcone isomerase 1) | (Ballester et al., 2016) |
| | ILs *S. habrochaites* (LA1777) x *S. lycopersicum* cv. 'E6203' | Flavonoids | 1 | 5 | Solyc05g010320 (CHI1, naringenin chalcone isomerase 1) | (Hanson et al., 2014) |
| | $BC_2F_2$ *S. habrochaites* (LA1223) x *S. lycopersicum* 'TA1166' | Flavonoids | 4 | 2, 3, 5, 11 | – | (Ökmen et al., 2011) |
| | | Total phenolics | 5 | 1, 6, 7, 9, 12 | – | (Ökmen et al., 2011) |
| | 96 cultivated genotypes (GWAS) | Total phenolics | 2 | 8, 11 | Solyc08g082350.2.1 (unknown), Solyc11g010170.1.1 (LanC-like protein2), Solyc11g010200 and Solyc11g010470 (14-3-3), Solyc11g 009100 (ABC-2 Transporter), Solyc11g 010380 (MATE) | (Ruggieri et al., 2014) |
| | RILs, ILs, Sub-ILs *S. pimpinellifolium* (TO-937) x *S. lycopersicum* cv. 'Moneymaker' | Total phenolics | 10 | 1, 4, 5, 7, 8, 11, 12 | Solyc01g079620 (SlMYB12), Solyc01g079240 (LACS1, LONG CHAIN ACYL SYNTHETASE 1), Solyc01g056340 (DET1, DE-ETIOLATED 1), Solyc04g014370 (PAS2, PASTICCINO2), Solyc05g008250 (GL1, GLABRA1), Solyc08g067260 (FDH, FIDDLEHEAD), Solyc08g067410 (KCS11-like 1) | (Barraj Barraj et al., 2021) |
| **Eggplant** | $F_2$ population from '305E40' x '67/3' | Phenolic acids | 2 | 4, 6 | – | (Toppino et al., 2016) |

*(Continued)*

TABLE 1 Continued

| Species | Mapping population | Metabolite (s) | QTLs | Chromosome (s) | Candidate gene(s) (encoding protein) | Reference |
|---|---|---|---|---|---|---|
| | ILs *S. melongena* 'AN-S-26' × *S. incanum* 'MM577' | Phenolic acids | 1 | 7 | SMEL_007g290860.1 (PAL1, phenylalanine ammonia-lyase 1), SMEL_007g288660.1.01 (peroxidase), SMEL_007g288680.1.01 (peroxidase), SMEL_007g288690.1.01 (peroxidase) | (Rosa-Martínez et al., 2022a) |
| | | | 4 | 3, 5, 7, 12 | SMEL_003g185030.1 (4CL2, 4-coumarate-CoA ligase 2), SMEL_007g275930.1 (4CL5, 4-coumarate-CoA ligase 5), SMEL_007g290860.1 (PAL6, phenylalanine ammonia-lyase 6) | (Rosa-Martínez et al., 2022b) |
| | | Total phenolics | 2 | 3, 12 | SMEL_003g185030.1 (4CL2, 4-coumarate-CoA ligase 2) | (Rosa-Martínez et al., 2022b) |
| | $BC_2$ and $BC_3$ *S. melongena* 'MEL3' × *S. elaeagnifolium* 'ELE2' | Phenolic acids | 2 | 1, 5 | – | (Villanueva et al., 2021) |
| | | Total phenolics | 2 | 1, 6 | SMEL_006g261420.1.01 (CCR1, cinnamoyl-CoA reductase 1), SMEL_006g261630.1.01 (4CL2, 4-coumarate-CoA ligase 2) | (Villanueva et al., 2021) |
| | $F_6$RIL *S. melongena* '305E40' x *S. melongena* '67/3' | Flavonoids | 3 | 7, 10 | SMEL_007g288740.1, SMEL_007g288700.1, SMEL_007g288720.1, SMEL_007g288660.1, SMEL_007g288690.1, SMEL_007g288730.1, SMEL_007g288920.1, SMEL_010g353210.1, SMEL_010g352840.1, SMEL_010g352880 (PYL4 abscisic acid receptor), SMEL_010g352930.1, SMEL_010g353090.1, SMEL_010g353110.1, SMEL_010g353170.1 and SMEL_010g353190.1 (kaempferol 3-O-beta-D-sophoroside), SMEL_010g352880 (rutin) | (Sulli et al., 2021) |
| **Pepper** | $F_2$ *C. annuum* 'AC1979' × *C. chinense* 'No. 4661' | Phenolic acids | 1 | 11 | – | (Wahyuni et al., 2014) |
| | | Flavonoids | 32 | [1-4], [6-10] | *CaMYB12* (naringenin chalcone, FLS, CHI, CHS) | (Wahyuni et al., 2014) |
| | NILs, $F_2$ *C. annuum* 'Long Sweet' (P12) × *C. annuum* 'AC2212' (P24) | Flavonoids | 1 | 5 | Ca05g18430 (*CaMYB12-like*, MYB-like DNA-binding domain protein) | (Wu et al., 2022) |

* indicates mQTL hotspots.
Information regarding the mapping population used, chromosome, candidate genes within the QTLs regions and reference is also provided.

phenolics concentration by 40%, on average, compared to 'M82'. Both *phe*6-2 and *phe*7-4 showed a positive effect, whereas the other QTLs (often carrying the *S. pennellii* allele) had a negative impact on phenolics levels. The authors suggested that a complex regulatory system was expected based on the $F_1$ intermediate values (Rousseaux et al., 2005) (Table 1; Supplementary Data S2). Di Matteo et al. (2010) and Sacco et al. (2013) explored the expression, location and inheritance model of the QTL identified in IL7-3 (LA4066), responsible for higher levels of phenolics, located at 39-64 cM in chromosome 7, and confirmed its recessive mode of inheritance. Finally, Di Matteo et al. (2013), using transcriptomics and TILLING, proposed that the higher phenolics content of IL7-3 is due to a more efficient compartmentalization in the vacuoles of these metabolites (specially flavonoids), as a result of higher expression of *MATE* (Solyc10g081260), *VSP* (Solyc06g005080), and *GST* (Solyc07g056420 and Solyc07g056510) genes, ultimately regulated by *ERF1* (Solyc05g051200) located at chromosome 5 (Table 1; Supplementary Data S2).

More recently, a population of backcross inbred lines (BILs) was developed from the *S. pennellii* ILs to generate a many-fold higher mapping resolution (Ofner et al., 2016). A total of 504 of these BILs, along with 76 ILs with *S. pennellii* introgressions were evaluated by Szymański et al. (2020) using a multi-omics approach, which resulted in the identification of several flavonoid-related mQTLs, as well as eQTLs associated with the phenylpropanoid metabolism. As an outcome worth exploring, the authors provided an online repository of the large dataset generated (https://szymanskilab.shinyapps.io/kilbil/) to make it easily accessible. Most notably, a large mQTL in chromosome 5 (bin 424) was found to be a hotspot for metabolic changes of several metabolites, including at least 22 flavonoids that were downregulated at the red fruit stage, as well as rutin and four kaempferol glycosides that were upregulated. We summarized in Table 1 the most relevant mQTL hotspots for phenolic acids, flavonoids, and derivatives, which were highlighted by the authors (Szymański et al., 2020) (Table 1; Supplementary Data S2). Furthermore, this study revealed the key role of specific flavonoids in fruit response against *Botrytis cinerea*. In this way, the authors reported the association of a higher accumulation of naringenin-7-O-glucoside with susceptibility, while a higher accumulation of its aglycone naringenin was associated with the resistant phenotype. (Szymański et al., 2020).

An IL population derived from a cross between *S. chmielewskii* (LA1840) and *S. lycopersicum* cv. 'Moneyberg' has also provided important insights into the phenolic acids and flavonoids accumulation in tomato fruits. Through LC-PDA-QTOF-MS coupled with physical mapping of the introgressions, Ballester et al. (2016) identified 126 different compounds from different classes, including phenolic acids and flavonoids, 56 of which were significantly increased/decreased compared to the cultivated parent. Thus, an introgression in chromosome 7 (IL7d) was linked to significant increases of di- and tri-caffeoylquinic acid and chlorogenic acid. Interestingly, within this region is located the gene *HQT*, responsible for trans-esterification of caffeoyl-CoA with quinic acid in order to form CGA. IL4d was linked to an increase of tri-caffeoylquinic acid, chlorogenic acid and 4-caffeoylquinic acid. Contrastingly, an introgression in chromosome 12 (IL12d) showed almost a three-fold decrease in di- and tri-caffeoylquinic acid compared to the cultivated accession (Ballester et al., 2016). Likewise, IL5b was found responsible for the significant increases of kaempferol and quercetin glycosides in the fruit peel. This region proved to harbor 511 genes, although only 17 were upregulated three-fold or higher, one encoding CHI (Solyc05g010320), which is responsible for directing the flux of the pathway towards the synthesis of flavonol glycosides (Muir et al., 2001). Similarly, IL4d showed potential to increase kaempferol and quercetin glycosides, although in a more modest way (between 3- and 5-fold compared to controls). Other introgressions showed a positive effect on one or two metabolites, like IL2b for quercetin feruloyl deoxyhexose-dihexose or IL6e for quercetin coumaroyl hexose-deoxyhexose-hexoside. Contrarily ILs 8a, 9d and 12d showed a significant negative impact on the concentration of quercetin and naringenin glycosides (Ballester et al., 2016) (Table 1; Supplementary Data S2).

Seven ILs derived from a cross between *S. habrochaites* (LA1777) and *S. lycopersicum* cv. 'E6203' (Monforte and Tanksley, 2000) were selected to be further analyzed regarding their chemical profiles (Hanson et al., 2014). Introgression line LA3984, encompassing a small fragment of *S. habrochaites* in chromosome 5, showed high levels of rutin in ripe fruits. The QTL proved to be harboring 38 genes, including the annotated chalcone-flavonone isomerase gene (Solyc05g010320), which was proposed as the gene responsible for high rutin levels in the *S. chmielewskii* IL population (Ballester et al., 2016). In addition, a second chalcone-flavonone isomerase gene was located upstream the QTL (Table 1; Supplementary Data S2). The introgression of this *S. habrochaites* QTL increased the expression of chalcone isomerase by restoring the flavonol synthesis pathway, resulting in elevated rutin content (Hanson et al., 2014).

Using the $BC_2F_2$ population from *S. habrochaites* (LA1223) × *S. lycopersicum* 'TA1166', Ökmen et al. (2011) reported five QTLs, scattered across chromosomes 1, 6, 7, 9 and 12, controlling the total phenolics fraction of tomato fruits (Table 1; Supplementary Data S2). All QTLs showed a positive effect and two coincided with Rousseaux et al. (2005) findings, at chromosome 7 and 9. For these QTLs, the wild allele increased the phenolics content between 8 and 17%. The QTL located at chromosome 7 showed the biggest increase in phenolics content (Ökmen et al., 2011). Moreover, authors reported four QTLs controlling flavonoid accumulation in tomato fruits, one with a positive effect, located at chromosome 11, and three with negative effect, located at chromosomes 2, 3 and 5. Despite the negative effect of most of the wild alleles, the QTL located at chromosome 11 increased the flavonoid content by 24% (Table 1; Supplementary Data S2).

A GWAS was carried out by Ruggieri et al. (2014) to detect associations between SNPs and fruit nutritional and apparent quality traits in a collection of 96 cultivated tomato genotypes, using the SNP-based SolCAP array. Two polymorphic markers were significantly associated with total phenolics content, located at chromosomes 8 (solcap_snp_sl_100367) and 11 (solcap_snp_sl_34253) (Table 1; Supplementary Data S2). In both cases, the minor alleles had a positive effect on the trait. The first marker was associated with gene Solyc08g082350.2.1, encoding for a protein of unknown function, while the second was associated with gene Solyc11g010170.1.1, which encodes a LanC-like protein 2, which is involved in the modification and transport of peptides in bacteria, although no clear role has been attributed in plants. In addition, four candidate genes were in linkage-disequilibrium with the marker at the beginning of chromosome 11, two *14-3-3* (Solyc11g010200 and Solyc11g010470), one *ABC-2 Transporter* (Solyc11g 009100) and one *MATE* (Solyc11g 010380) (Ruggieri et al., 2014). Interestingly, the *MATE* had been reported to be involved in vacuolar compartmentalization of phenolic compounds by Di Matteo et al. (2013) (Table 1; Supplementary Data S2).

More recently, the study of RILs, ILs and SubILs derived from a cross between *S. pimpinellifolium* (TO-937) and *S. lycopersicum* cv. 'Moneymaker' identified nine QTLs (*ph*1.1, *ph*4.1, *ph*5.1, *ph*5.2, *ph*7.1, *ph*7.2, *ph*8.1, *ph*12.1 and *ph*12.2) involved in the accumulation of cuticle phenolics across six different chromosomes (Barraj Barraj et al., 2021) (Table 1; Supplementary Data S2). Two wild alleles, *ph*4.1 and *ph*12.1, were linked to an increase of 20 and 30% in ripe fruits, respectively, while alleles located at chromosome 5 showed increased phenolics levels at immature and mature green stages. The other five QTLs, carrying the cultivated allele, showed a negative effect on the accumulation of phenolics in the fruit peel. The cultivated QTL *ph*12.2 had a particularly negative impact, leading to a 40% decrease of phenolics. The QTL in chromosome 1 was linked to *SlMYB12* (Solyc01g079620), *LONG CHAIN ACYL SYNTHETASE 1* (*LACS*1; Solyc01g079240) and *DE-ETIOLATED1* (*DET1*; Solyc01g056340) genes, although closer to the first one. The QTL *ph*4.1 co-located with *PASTICCINO2* (*PAS2*; Solyc04g014370), *ph*5.1 with *GLABRA1* (*GL1*; Solyc05g008250), and *ph*8.1 with *FIDDLEHEAD* (*FDH*; Solyc08g067260) and *KCS11-like1* (Solyc08g067410). Many of these genes have roles in cuticle deposition, but could be also involved in the phenolics transport to the cuticle, as the expression analysis seemed to suggest for *FDH* (Barraj Barraj et al., 2021) (Table 1; Supplementary Data S2). Furthermore, phenolics gene expression seems to be regulated by a complex combination of additive and epistatic interactions, with tight links to the cuticle wax and polysaccharides synthesis, as reported by España et al. (2014) and Heredia et al. (2015). An epistatic relationship between a region in chromosome 11 and the QTL found in chromosome 12 (*ph*12.1) was detected and, although the region on chromosome 11 had no significant effect on the phenolics content, it increased the effect of *ph*12.1 (Barraj Barraj et al., 2021).

Tomato mutants, carrying natural or induced mutations, may also represent a source of valuable alleles regarding metabolic regulation of different pathways (Bovy et al., 2010; Chattopadhyay et al., 2021). In that regard, *high pigment* (*hp*) tomatoes comprise a particularly interesting source of variability since their phenotype is the result of exaggerated accumulation of many antioxidants and photoprotective metabolites, including flavonoids, in their fruits (Levin et al., 2006; Azari et al., 2010; Chattopadhyay et al., 2021). The different tomato *hp* mutants contain mutations in either the *DET1* gene (*hp-2*, *hp-2$^{j}$*, and *hp-2$^{dg}$*), a negative regulator of photomorphogenesis, or the *UV DAMAGED DNA BINDING* protein 1 (*DDB1*) gene (*hp-1* and *hp-1$^{w}$*) which interacts with the *DET1* gene (Levin et al., 2006; Azari et al., 2010). Several studies reported the metabolic profiles of the different *hp* mutants and showed that these mutations not only led to the accumulation of high levels of flavonoids, but also to increased levels of other bioactive compounds, such as carotenoids (Bino et al., 2005; Levin et al., 2006; Long et al., 2006; Rutley et al., 2021). The *hp* mutations have been extensively introgressed into elite cultivars, mainly to increase the lycopene content in processing tomatoes (Levin et al., 2006). Likewise, different combinations of *hp* mutations have been stacked in the same genetic background to shed light on their genetic control and their impact on metabolite content. The homozygote double-mutants carrying *Anthocyanin fruit* (*Aft*) with *hp-1$^{w}$* and *Aft* with *hp-2$^{j}$* showed a significant increase in the accumulation of both anthocyanins and rutin compared to the single-mutation lines (van Tuinen et al., 2006). In addition, the stacking of *Aft* with *hp-1* showed promising results by synergistically increasing the production of delphinidin-, petunidin- and malvidin-type anthocyanins and quercetin- and kaempferol-type flavonols in the fruit (Sapir et al., 2008).

*Eggplant*

Being the most abundant phenolic acid in eggplant fruits, studies of the phenolic profile have naturally focused on CGA. Although few QTL approaches have been carried out for deciphering the genetic basis of CGA biosynthesis in eggplant, the development of a growing number of intra and interspecific mapping populations is leading to an escalation of these studies. A $F_2$ population derived from an intraspecific crossing ('305E40' × '67/3') was used to identify two stable QTLs for CGA content on linkage groups E04 (*CGAE04*) and E06 (*CGAE06*) (Toppino et al., 2016). In both cases the positive allele came from the '305E40' parent. No candidate genes pointing at the phenylpropanoid pathway could be identified within the QTLs' regions. However, according to the authors, *CGAE04* and *CGAE06* could be related to quinic acid, since they were syntenic to two tomato ILs containing QTLs for that moiety of CGA (Toppino et al., 2016) (Table 1; Supplementary Data S2). A stable QTL associated with a reduction of CGA content was identified in a recent work (Rosa-Martínez et al., 2022a) using the first developed set of interspecific eggplant ILs with genome fragments of its wild relative *S. incanum* (Gramazio et al., 2017). The QTL (*cga7*) was located at the end of chromosome 7 (129-135 Mb), where a potential candidate gene (SMEL_007g290860.1), orthologous to the tomato Solyc09g007920, which encodes for phenylalanine ammonia-lyase 1 (*Sl*PAL1), was identified by the authors (Table 1; Supplementary Data S2). In a different trial, under two nitrogen (N) regimes, four QTLs for decreased CGA content, on chromosomes 3, 5, 7 and 12, as well as two QTLs negatively controlling total phenolics content, on chromosomes 3 and 12, were identified (Table 1; Supplementary Data S2) (Rosa-Martínez et al., 2022b). Most of these QTLs were only detected under high-N conditions, whereas the QTL for CGA on chromosome 3 was identified under the two N treatments (*cga3.HN* and *cga3.LN*). However, the introgressed fragment large size hampered the identification of underlying candidate genes and, therefore, it will require the use of strategies such as obtaining sub-ILs, to narrow it down (Rosa-Martínez et al., 2022b). Finally, the evaluation of a set of $BC_2$ and $BC_3$, derived from the cross of *S. melongena* 'MEL3' and its distant wild relative *S. elaeagnifolium* 'ELE2' (García-Fortea et al., 2019), under N-restrictive conditions, led to the identification of a QTL increasing CGA content on chromosome 5 (3.9–4.5 Mb) (Villanueva et al., 2021). Interestingly, a distinct pattern of phenolic acid accumulation, associated with a larger array of hydroxycinnamate conjugates other than CGA was detected in this population. The authors were able to identify a QTL associated with the unusual phenolic profile of *S. elaeagnifolium* on chromosome 1, as well as two QTLs associated with a greater total phenolic HPLC-peaks area and a lower CGA HPLC-peak area, located on chromosomes 6 and 1, respectively (Villanueva et al., 2021) (Table 1; Supplementary Data S2).

Studies about the flavonoid accumulation, other than anthocyanins, in eggplant are scarce, due to the fact that their concentration is negligible compared to CGA and anthocyanins at commercial maturity (Lo Scalzo et al., 2021). A recent metabolic study of a $F_6$ recombinant inbred line (RIL) population (Sulli et al., 2021) derived from the eggplant lines '305E40' and '67/3', identified one mQTL on chromosome 10 associated with the level of rutin (P-RUT.10.1). Two other mQTLs were associated with the level of the flavonol glycoside kaempferol 3-*O*-beta-D-sophoroside. One of them was also located on chromosome 10 (P-KSOPH.10.1) near P-RUT.10.1, while the other was located on chromosome 7 (P-KSOPH.7.1). Several candidate genes for these QTLs were also suggested by the authors, encompassing a PYL4 abscisic acid receptor (SMEL_010g352880), which is known to positively regulate flavonoid/anthocyanin biosynthesis in tomato (Diretto et al., 2020), peroxidase-encoding genes and other transcription factors (Sulli et al., 2021) (Table 1; Supplementary Data S2).

*Pepper*

Only a few studies aimed at identification of QTLs associated with phenolic compounds were found for pepper at the time of writing. In the first work, authors used a genetic linkage map constructed from an interspecific $F_2$ population derived from the cross between *C. annuum* AC1979 and *C. chinense* No. 4661 to locate QTLs and genes (Wahyuni et al., 2014). From the comprehensive untargeted metabolic profiling, which led to the identification of more than 500 semi-polar compounds, the authors employed both the QTL and candidate gene approaches to identify one mQTL for the phenolic acid derivative ferulic acid-hexose located in the linkage group P11 (Wahyuni et al., 2014) (Table 1; Supplementary Data S2). In terms of flavonoids, this study

identified 12 mQTLs for different flavonoid glycosides encompassing quercetin, apigenin, luteolin, flavanone derivatives, and naringenin chalcone, scattered over the linkage groups P02, P06, P07, P09 and P10. Moreover, 15 eQTLs for flavonoid pathway genes were identified: (i) eQTLs for *CHS-1*, *CHS-2*, *CHI-2* and *FLS* co-localized on P01 and overlapped with a mQTL for naringenin chalcone, eQTLs for *CHS-1*, *CHS-2*, *CHI-1*, *CHI-2* and *FS-2* co-localized with a naringenin chalcone mQTL on P09, (iii) eQTLs for *F3'H-1* and *F3'H-4*, and *F3H* were detected on P03, P04 and P08, respectively (Wahyuni et al., 2014) (Table 1; Supplementary Data S2). In a study aimed at identifying QTLs associated with $\alpha$-glucosidase inhibitory activity of flavonoids and other compounds present in pepper leaves and fruits (Park et al., 2020), the authors found a QTL on chromosome 9, which co-located with the cluster of mQTLs and eQTLs identified by Wahyuni et al. (2014). In a recent work, Wu et al. (2022) used a $F_2$ population, resulting from the cross between the flavonoid-contrasting *C. annuum* accessions Long Sweet (P12) and AC2212 (P24), to identify a major QTL (*fc5.1*) controlling the accumulation of chalcones, flavonols and flavones located at 228 Mbp on pepper chromosome 5. Further analysis of this region, in a NILs (Near Isogenic Lines) population derived from the original $F_2$ population, identified three *CHS* (Ca5g17030, Ca5g17040 and Ca5g17060) and a homolog to the tomato TF gene *SlMYB12*, *CaMYB12-like* (Ca05g18430). RNA-seq later confirmed the up-regulation of *CaMYB12-like* in NILs carrying the high-flavonoid P12 allele along with 17 other genes involved in the flavonoid pathway. Since this network of genes is scattered across several chromosomes, *CaMYB12*-like must be *trans*-regulating the entire pathway. Virus-induced gene silencing (VIGS) confirmed that *CaMYB12-like*-silenced plants had a significant decrease of flavonoid content and structural flavonoid pathway genes expression levels, compared to mock-silenced plants (Wu et al., 2022).

## 2.3 Regulation of phenylpropanoid biosynthesis pathway

Although the phenylpropanoid biosynthetic pathway is conserved across these species, the accumulation profile of metabolites in tomato, eggplant and pepper differs significantly (Rosa-Martínez et al., 2021). Moreover, studies have proven that phenylpropanoids accumulate in a tissue-specific and developmental-dependent ways (Calumpang et al., 2020; Liu et al., 2020; Lo Scalzo et al., 2021). This implies that, in addition to the structural genes, the pathway is tightly controlled by regulatory complexes. Transcription factors (TFs) of the R2R3-type MYB, Zinc Finger, bHLH, MADS, WRKY and WD40 families have been reported to regulate the phenylpropanoid metabolism in plants, including the Solanaceae (Tohge et al., 2017; Zhu et al., 2018; Sacco et al., 2019; Liu et al., 2020). Among these, the R2R3-type MYB TFs are the most relevant group in tomato, eggplant and pepper.

The role of *MYB12* in the regulation of phenylpropanoid metabolism has been thoroughly characterized in tomato. The study of the metabolomic and molecular profiles of pink-colored fruits of two sets of ILs with wild introgressions from *S. pennellii* (Adato et al., 2009) and *S. chmielewskii* (Ballester et al., 2010) linked *SlMYB12* to the genomic region harboring the *y* mutation on chromosome 1, and to the regulation of naringenin chalcone accumulation in the fruit peel. This was demonstrated by downregulation/upregulation (Adato et al., 2009), gene silencing (Ballester et al., 2010), overexpression (Wang et al., 2018) and by studying mutations in its sequence (Fernandez-Moreno et al., 2016; Jung et al., 2017). These determined that *SlMYB12* controls the expression of the structural genes *SlPAL*, *Sl4CL*, *SlCHS*, *SlCHI*, *SlF3H*, *SlF3'H* and *SlFLS*, being the main activator of the flavonol biosynthetic branch. In addition, Zhang et al. (2015) demonstrated that *AtMYB12*, *SlMYB12*'s homolog of *Arabidopsis*, is not only able to bind to specific phenylpropanoid structural genes, but also to genes associated with the primary metabolism, such as 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (*DAHPS*) and plastidial enolase (*ENO*), and thus exerting its activity at several metabolic points. This complex regulatory action of *SlMYB12* was also pointed out by Fernandez-Moreno et al. (2016) through transcriptomics and by Zhu et al. (2018). In this way, besides increasing the demand for flavonoid synthesis, *MYB12* would increase the flow of carbon entering the shikimate pathway and the formation of aromatic amino acids that serve as substrates for the synthesis of phenolic compounds (Zhang et al., 2015). Other MYB-type TFs have been reported to act as regulators at different levels of both primary and phenylpropanoid metabolism in tomato. Wu et al. (2020) reported a significant impact of *SlMYB72* on several vital metabolic pathways, including the phenylpropanoid pathway. Downregulation of *SlMYB72* led to increased levels of total flavonoids, quercetin, kaempferol, gallic acid and chlorogenic acid, whereas *SlMYB72*-upregulated plants showed no significant differences compared to their azygous controls. *SlMYB72* was reported as a significant player in the regulation of the phenylpropanoid and flavonoid pathways, since it directly influenced the expression of *4CL*, *CHS1* and *CHS2* genes (Wu et al., 2020).

In eggplant, efforts in this area so far resulted in the identification of more than 100 R2R3-MYB TFs (Li et al., 2021), categorized into 20 subgroups (SGs), of which four (SG4 to SG7) were reported to modulate the phenylpropanoid pathway, mainly anthocyanin and flavonol biosynthesis. Within these SGs, ten genes were identified by the authors. *SmMYB1* (SG6), was isolated from an occidental eggplant variety with the purpose of characterizing its expression level in different tissues and its regulation mechanism (Docimo et al., 2016). *SmMYB1* is orthologous to *SlANT1* and the locus *A* in *C. annuum*, widely characterized in tomato and pepper, respectively, as main regulators of anthocyanin synthesis. The authors found that *SmMYB1* was responsible for differentially regulating both the chlorogenic acid and anthocyanin biosynthesis in eggplant. Through *N. benthamiana* transient expression of *SmMYB1* and a C-terminal truncated form of the protein, the authors demonstrated the essential role of the C-terminal region in anthocyanin biosynthesis, while its deletion does not limit the capability of *SmMYB1* to regulate chlorogenic acid accumulation. These results pointed out that TFs involved in the phenylpropanoid pathway might contain several action domains that can bind to different substrates and exert their regulatory action at different points along the pathway (Docimo et al., 2016).

In pepper, using sequence homology with *SlMYB12*, Wahyuni et al. (2014) located a candidate gene for *CaMYB12* in chromosome 1 in the $F_2$ population derived from *C. annuum* × *C. chinense* genetic linkage map. In addition, through metabolic and gene expression analyses in ripe fruits of this population, the authors identified a mQTL for naringenin chalcone and eQTLs associated with *CaMYB12* and other genes encoding specific enzymes in the flavonoid pathway (*CaCHS-1*, *CaCHS-2*, *CaCHI-1* and *CaFLS*), which colocalized at the *CaMYB12* locus on chromosome 1 (Wahyuni et al., 2014). This suggests that a similar regulatory action of *MYB12* transcription factor may control the phenylpropanoid metabolism in pepper and tomato.

In a recent publication, Arce-Rodríguez et al. (2021) performed *in silico* genome-wide analyses of the MYB family in *C. annuum* to identify 235 non-redundant MYB proteins. Coupled with RNA-seq and RT-qPCR assays, the authors reported several *CaMYB*s putatively regulating phenylpropanoid biosynthesis in pepper. Thus, *CaMYB32*, *CaMYB33* and *CaMYB93* expression patterns were shown to be co-expressing with *PAL*. Of these, *CaMYB93* expression correlated with the *PAL* expression profile throughout fruit development, whereas *CaMYB32* and *CaMYB33* transcription factors showed to be highly correlated with flower and early fruit development expression. Moreover, the phylogenetic analysis of *CaMYB*s sequences grouped them in 46 clades (C1–C46) indicating both gene conservation and specialization, by grouping some *CaMYB*s with *A. thaliana* and tomato MYBs and some in *Capsicum*-only groups, respectively. Hence, *CaMYB97*, *CaMYB98*, *CaMYB99* and *CaMYB100*, which grouped together with *AtMYB12* and *SlMYB12*, could be implicated in the regulation of flavonoid biosynthesis in pepper (Arce-Rodríguez et al., 2021).

Islam et al. (2021) followed a similar approach to identify MYB TFs in the *C. chinense*, *C. baccatum* and *C. annuum* genomes, yielding 251, 240 and 245 unique TFs genes, respectively. Based on expression patterns during fruit developmental stages, researchers identified several *C. chinense* and *C. annuum* MYB genes co-expressing with *CHS* and *DFR*, two structural genes of the flavonoid pathway, suggesting that *CcMYB16*, *CcMYB28*, *CcMYB100*, *Cc*A, *CcDIV4*, *CcMYB46*, and *CcMYB74* may be regulating anthocyanin/flavonoid biosynthesis in *Capsicum* fruits. In addition, *CcPHR8* and *CaHHO1*, two TFs present within a reported QTL controlling fruit pungency, were also observed to be co-expressing with genes *C4H* and *DFR*, while *CcMYB48* co-expressed with *PAL*, indicating potential involvement in the phenylpropanoid pathway. Through phylogeny the authors confirmed Arce-Rodríguez et al. (2021) results regarding the clustering of MYBs 98, 99 and 100 with *AtMYB12* and thus having a potential role on flavonol glycoside accumulation. Moreover, *CcMYB115* and *CcA* clustered with *AtMYB75* and *AtMYB90* which have been related to anthocyanin and phenylpropanoid biosynthesis, respectively (Islam et al., 2021).

# 3 Breeding strategies for improving phenolics content in tomato, eggplant and pepper

The current knowledge on the genetics of the phenylpropanoid pathway of tomato, eggplant and pepper, including QTLs, structural and regulatory genes, as well as the elite germplasm, provides enough tools to improve phenolics levels, and thus the health-promoting function, of three of the most consumed vegetables in the world. Several works have proved the effectiveness of combining molecular tools and trait variability through conventional and genetic engineering strategies.

## 3.1 Conventional breeding

Great diversity has been found among the varieties of tomato, eggplant and pepper with respect to phenolic acids and flavonoid concentration in fruit. For instance, in tomato, rutin values ranged from 2 to 231 mg $kg^{-1}$ of fresh weight (FW) in a collection of commercial materials and between 22 to 80 mg $kg^{-1}$ FW in a collection of traditional varieties (Bovy et al., 2010; Rosa-Martínez et al., 2021). Stommel and Whitaker (2003) found differences of nearly 20-fold for total hydroxycinnamic acid content in a collection of 101 cultivated eggplant accessions from the USDA-ARS core collection. Mori et al. (2013) also found varietal variations in CGA content between 0.1 and 2.5 mg $g^{-1}$ FW among 34 eggplant cultivars and lines with diverse growth habits, fruit shapes, sizes, and colors. Huge variation has been reported for phenolic acids and flavonoids in pepper varieties, such as vanillic acid (3.1–13.3 mg $kg^{-1}$ FW), ferulic acid (2.2–12.5 mg $kg^{-1}$ FW), quercetin (3.3–783 mg $kg^{-1}$ FW) or luteolin (0.2–103 mg $kg^{-1}$ FW) (Lemos et al., 2019). In addition, many elite materials come from mapping populations. These are incredibly useful genetic resources as they carry favorable alleles for the accumulation of metabolites, usually introgressed from wild accessions, such as *S. pennellii* (LA0716), *S. chmielewskii* (LA1840), *S. habrochaites* (LA1777), *S. pimpinellifolium* (TO-937), S. *incanum* (MM577) and *S. elaeagnifolium* (ELE2), into a cultivated genetic background, that can be readily incorporated into breeding programs (Prohens et al., 2017).

Therefore, the main aspects of conventional breeding should be cataloguing, selecting and hybridizing these elite materials. In addition, to successfully employ the available genetic resources, the identification of QTLs and the underlying genes is of paramount importance. The application of fine-mapping tools and the development of functional markers could have an unprecedent impact in the development of high-phenolic varieties without the negative impact of linkage drag (Salgotra and Neal Stewart, 2020).

As outlined before, the flavonoid pathway is disabled in the tomato fruit flesh of the cultivated germplasm, rendering it impossible to restore just by cross-pollinating the best materials in the primary genepool (Willits et al., 2005). Fortunately, within the wild genepool, there are materials that could address both the restoration of the flavonoid pathway in the fruit flesh and the increase of levels of flavonols in the fruit peel (Tohge et al., 2020). The aforementioned works of Willits et al. (2005); Hanson et al. (2014) and Ballester et al. (2016) are proof of that.

Marker-assisted selection and pyramiding of genes and QTLs have been proven to be effective strategies for improving several traits simultaneously or a single quantitative trait with polygenic control (Sacco et al., 2013; Rigano et al., 2014; Rigano et al., 2016). Numerous QTLs and genes involved in the phenylpropanoid

pathway are scattered throughout the genome. The high recombination frequencies among these would make it easier to combine several favorable alleles in a single line. In tomato, this has been successfully used to increase phenolics content (Sacco et al., 2013; Rigano et al., 2014; Rigano et al., 2016). It has also served to demonstrate that the introgressed alleles of *S. pennellii* had a synergistic effect over phenolic acids concentration, mainly gallic and ferulic acids, while being safe for human consumption (Rigano et al., 2014). Also, the combination of transcriptomics, metabolomics and genomics of two pyramided genotypes, obtained by crossing *S. pennellii* IL12-4 and IL7-3, provided insights into the function of structural and regulatory genes involved in the branching of the phenylpropanoid pathway (Rigano et al., 2016). It demonstrated the central role of one 4-coumarate-CoA ligase in the redirection of the pathway towards the synthesis of cinnamic acids or flavonoids, as well as the relevancy of the regulatory action of TFs (Rigano et al., 2016).

In eggplant, higher levels of antioxidant activity, total phenolics and chlorogenic acid have been reported in the wild relative *S. incanum* compared to *S. melongena* (Stommel and Whitaker, 2003; Kaur et al., 2014). Prohens et al. (2013) evaluated the $BC_1$ population derived from their crossing for total phenolics and hydroxycinnamic acids content and found a number of plants presenting a good combination of phenolic acids content and fruit weight or flesh browning. Therefore, these plants could constitute a powerful resource to be introduced into breeding pipelines.

Finally, in pepper, the locus *fc5.1*, carrying a high-flavonoid allele, was introgressed into the commercial cultivar backgrounds of Yolo Wonder and Sweet Banana. The resulting $BC_1S_1$ plants revealed an additive effect of this allele on the flavonoid content, showing significant increases of chalcones, flavones and flavonols (Wu et al., 2022).

## 3.2 Genetic engineering

Genetic engineering tools have enabled a great deal of important discoveries regarding the tomato phenylpropanoid pathway. In several cases, it led to outstanding results, unlikely to be achieved through conventional techniques (Schijlen et al., 2006; Bovy et al., 2007). However, the commercialization of genetically modified fruits is not yet legally accepted in some parts of the world, such as Europe. Even the precise gene-editing technology CRISPR-Cas9 has been considered a genetically-modifying technique and is thus under the same limitations (Biswas et al., 2021).

Despite those limitations, CRISPR/Cas9 is becoming the most widespread gene editing technology. Hence, it has been successfully established in Solanaceous crops and has proven its value in applied research directed at understanding the genotype-phenotype relationship. While being primarily reported for tomato, its application in eggplant and pepper is growing (Van Eck, 2018). Genome-editing technologies are enabling researchers to reproduce directly in elite lines the allelic variations responsible for improvement of the traits of interest (Rothan et al., 2019), saving years during the breeding process and avoiding linkage drag of undesired genes/alleles (Deng et al., 2018; Zhu et al., 2018). Some examples of its application in Solanaceae encompass the rewiring of the GABA metabolic pathway in tomato (Li et al., 2018b), the generation of knock-out mutants with reduced PPO activity and browning of eggplant's fruit flesh (Maioli et al., 2020), and the enhancement of resistance against anthracnose in pepper (Mishra et al., 2021). CRISPR/Cas9 has been successfully applied to edit the phenylpropanoid pathway in tomato. Concretely, researchers generated pink-fruited tomatoes *via* knockout of *SlMYB12* in the genetic background of different red-fruited lines. This led to transparent peels and reduced transcription of *CHS* (Deng et al., 2018; Zhu et al., 2018). By doing so, the authors showed that it is an efficient way to engineer fruit color and the metabolic pathway in tomato without jeopardizing other traits.

Breeding for enhanced flavonoid content in tomato can be approached in two ways: increasing the amount stored in the peel or activating the flavonoid pathway in the fruit flesh. Since the flavonols quercetin and kaempferol have higher bioactive properties, the main goal should be to increase their content. Through transgenesis, Muir et al. (2001) expressed in tomato plants the *Petunia hybrida CHI*, which led to an increase of peel quercetin and kaempferol glycosides content up to 78-fold. Despite that, flesh flavonoid content remained unaltered. On the other hand, the simultaneous ectopic expression of four *P. hybrida* genes (*CHI, CHS, F3H* and *FLS*) was able to produce high flavonol content in the peel and modest levels in the flesh (Colliver et al., 2002; Verhoeyen et al., 2002). Further experiments with transformed tomato plants carrying isolated or different combinations of *P. hybrida* genes led to increased accumulation of naringenin glycosides when *PhCHS* was ectopically expressed, and to those glycosides being converted to flavonol glycosides when *PhCHS* and *PhFLS* were stacked in the same background (Colliver et al., 2002; Verhoeyen et al., 2002). In another approach to restore the flavonoid pathway in the tomato flesh, Bovy et al. (2002) developed transgenic plants expressing the maize TFs *Lc* (*Leaf color*; MYC-type) and *C1* (*colorless-1*; MYB-type) genes, simultaneously, under the fruit-specific E8 promotor control. Authors reported increases of up to 60-fold of kaempferol glycosides in the fruit flesh (Bovy et al., 2002).

Modulation of regulatory genes has been a primary target towards improving the fruit content in flavonoids. Davuluri et al. (2005) used RNAi to suppress the accumulation of *DET1* transcripts in transgenic fruits to increase fruit bioactive compounds content while avoiding the negative impacts on plant development. This resulted in significant increases of naringenin chalcone, flavonol glycosides, chlorogenic acid, and carotenoids with no significant impact on other quality and growth parameters. Lim and Li (2017a) proposed ectopic co-expression of two TFs, *Delila* (bHLH-type) and *Rosea1* (MYB-type) (*Del*/*Ros1*) from *Antirrhinum majus*, in combination with the *Allium cepa CHI* gene in order to produce transgenic 'Rubion' tomatoes with high levels of both anthocyanins and flavonols. The strategy aimed at alleviating the major rate-limiting factor of the tomato flavonoid pathway by upregulating the expression of *CHI* followed by redirection of the pathway towards the anthocyanins. Authors reported a significant increase of both flavonols and anthocyanins in the fruit peel and flesh, reaching up

to 200-fold the total flavonol content of wild-type tomatoes (Lim and Li, 2017a). Likewise, Lim and Li (2017b) expressed the *AcCHI* in combination with the *AtPAP1* gene. The combination led to a 130-fold increase in rutin, along with a 30-fold increase of total anthocyanins, compared to the wild-type peel (Lim and Li, 2017b). It is also worth mentioning that the *AcFLS* gene was able to increase rutin content by as much as 3.5-fold in 'Rubion' transgenic tomatoes, in contrast to what was observed with the *PhFLS* gene, which was only effective when in combination with *PhCHI* (Verhoeyen et al., 2002; Lim and Li, 2017b). Regarding phenolic acids, the contents of chlorogenic acid, caffeic acid, and coumaric acid were significantly higher in the *AtPAP1*-only lines compared to the wild type, and only modest increases were observed in the gene-stacked lines (Lim and Li, 2017b).

MYB-type TF genes have been successfully used to increase the flavonoid content in tomato. Luo et al. (2008) developed transgenic 'MicroTom' and 'Moneymaker' plants expressing the *A. thaliana* transcription factor *AtMYB12* and reported strikingly-high amounts of rutin, kaempferol rutinoside, as well as high levels of caffeoylquinic acid and chlorogenic acid, in both fruit peel and flesh. Authors reported that flavonoid content made up to 10% of the fruit dry weight and that it had no significant effect on the carotenoid accumulation of transformed tomatoes (Luo et al., 2008). Zhang et al. (2015) combined *AtMYB12* with *Del/Ros1* to generate the Indigo phenotype, exhibiting intense blue-purple-colored fruits. These fruits had a two-fold increase of chlorogenic acid and a three-fold increase of flavonols compared to the *AtMYB12*-carrying line, plus a two-fold increase of anthocyanins compared to the purple *Del/Ros1* tomatoes. The combination of TFs was able to activate all the genes encoding enzymes of primary metabolism through to the flavonoid biosynthesis in both the peel and flesh of tomato fruits (Zhang et al., 2015). Likewise, the expression of the *AtMYB12* homolog, *AtMYB11*, in the genetic background of *S. lycopersicum* var. 'CSL' led to an increase of both flavonoid and caffeoylquinic acids content (Li et al., 2015). Alterations of the flavonol chemical profile were greater in the fruit peel, and transformed plants accumulated up to 18- and 33-fold the amount of rutin and kaempferol rutinoside, respectively, compared to the controls. In the flesh, flavonols accumulation was modest, showing a 3-fold increase of kaempferol rutinoside compared to the controls (Li et al., 2015).

Wang et al. (2018) developed *SlMYB12*-overexpressing transgenic tomato plants using the cultivars 'MicroTom', 'CSl09-03' and 'Sheng Nv-Guo'. This led to a five-fold increase of naringenin chalcone, 15-fold increase of rutin and five-fold increase of kaempferol rutinoside, on average (Wang et al., 2018), in accordance to the findings of Adato et al. (2009). Recently, Wu et al. (2020) developed *SlMYB72*-downregulated transgenic lines of 'MicroTom' and reported that total flavonoid content, as well as quercetin and kaempferol derivatives, gallic, and chlorogenic acids content, were significantly increased compared to upregulated plants (Wu et al., 2020).

Both eggplant and pepper are recalcitrant species to *in vitro* organogenesis and therefore the application of genetic engineering techniques has lagged behind compared to tomato (Gammoudi et al., 2017; García-Fortea et al., 2021). In eggplant, a recent work used agroinfiltration for the transient overexpression of *SmHQT* and showed a 2-fold increased CGA content (Kaushik et al., 2020). Using VIGS of *SmCHS* in eggplant cultivar 'Zhongnong Changfeng' ripening fruits, Wang and Fu (2018) were able to produce silenced plants showing white peel, resulting from a 2-fold decrease of naringenin and delphinidin. Strikingly, the authors reported an increase of some flavonols and isoflavonoids, compared to mock-inoculated plants, in spite of blocking *SmCHS*. Wang and Fu (2018) hypothesized that chalcone reductase and isoflavone synthase are able to use malonyl-CoA to synthesize isoflavonoids. In relation to the increase of some flavonols, the authors stated that the adoption of an alternative route through caffeoyl-CoA and the mobilization of flavonoids from other tissues may explain this result. The fact that *SmCHS2* kept 31% of its activity after the blocking may also help explain the increase in flavonols (Wang and Fu, 2018). Moreover, authors discussed the involvement of *CHS* in epidermal cell expansion during ripening, after silenced plants yielded uneven, curved and shorter fruits compared to controls. These alterations were also observed at the cellular level, with epidermal cells being compactly arranged, shorter and wider in the white-pigmented sectors compared to the controls. Finally, *CHS*-silenced fruits showed a depressed gravitropic response, implying that either *CHS* or certain flavonoids could be involved in the fruit gravitropic response (Wang and Fu, 2018). This is in agreement to previous findings in tomato (Schijlen et al., 2007; España et al., 2014). Finally, in pepper, application of VIGS targeting MYB and WD40 TFs led to promising results. Aguilar-Barragán and Ochoa-Alejo (2014) showed that the accumulation of anthocyanins in fruits of MYB- and WD40-silenced plants was significantly decreased, compared to the controls. Both transformants showed lower expression of *CHS*, *F3′5′H*, *DFR* and *3GT*. In MYB-silenced plants *CHI* was also negatively affected, whereas in WD40-silenced plants it showed no significant difference. Aversely, *F3H* expression was not affected in MYB-silenced plants while for WD40-transformants it did.

## 4 Potential drawbacks of breeding for increased phenolics

Due to the wide array of functions of phenolic compounds in plant development, and to the interrelationships among the different enzymes and regulatory elements of the phenylpropanoid pathway and with other pathways of plant metabolism, increasing their content in vegetables may affect other traits. That should not be overlooked during the breeding process.

In this respect, some studies have associated increased content of phenolics with lower fruit organoleptic quality. For instance, *SlMYB12* has been reported to act upon genes related to the primary metabolism in tomato (Zhang et al., 2015; Fernandez-Moreno et al., 2016). In this way, the flow of carbon was channeled towards the phenylpropanoid pathway, which indirectly decreased the sugars accumulated in the fruit, which ultimately could affect its flavor. Interestingly, pleiotropic effects related to fertilization and fruit

appearance have also been observed in tomato when silencing a structural gene of the flavonoid pathway. Schijlen et al. (2007) blocked *CHS* through RNAi leading to a dramatic decrease of total flavonoid levels in transgenic plants. This was accompanied by a delay of fruit development and smaller parthenocarpic fruits with an altered reddish color and a dull appearance, compared to the wild-type. This shows that flavonoids, despite being considered secondary metabolites, play a fundamental role in different processes of plant reproduction, fruit development and growth.

On the other hand, in eggplant, enzymatic browning occurs when phenolics, which are mostly confined in the vacuole, are released by the rupture of cellular compartments and oxidized by polyphenol oxidases (PPOs), resulting in brown coloration of the flesh, and thus, reduces the apparent fruit quality. That may suggest that increasing the phenolics content in the fruit would lead to higher browning, and indeed, this has been a major concern in eggplant breeding (Plazas et al., 2013; Kaushik et al., 2017). However, a correlation study among browning, phenolics, and PPO activity, casted doubts on a strong relationship between total phenolics or chlorogenic acid contents and fruit flesh browning (Kaushik et al., 2017). Also, several studies have shown that there is variation within the cultivated genepool for selecting varieties with increased phenolics and reduced PPO activity (Mennella et al., 2010; Mennella et al., 2012; Plazas et al., 2013). Another possibility could be through a CRISPR/Cas9-based mutagenesis approach, as mentioned earlier (Maioli et al., 2020).

Phenolic compounds have been associated to astringency and bitterness of foods (Drewnowski and Gomez-Carneros, 2000), which would also be a drawback of increasing phenolics content in tomato, eggplant and pepper. Nevertheless, at the concentrations normally found in these vegetables, it has no appreciable impact on those characteristics, which are mostly caused by saponins, glycoalkaloids and other compounds (Rosa-Martínez et al., 2022a). Also, due to the role of different flavonoids in the fruit cuticle and epidermal cells (España et al., 2014; Heredia et al., 2015), increasing their content could lead to a tougher fruit skin, which may in some cases be unpalatable to the consumer. In any case, pleiotropic effects of the modification of the phenylpropanoid pathway should be well studied as they may prove to be an advantage, like the production of parthenocarpic fruits (Schijlen et al., 2007), or a disadvantage (lower fruit quality) in commercialization of the product.

## 5 Concluding remarks

We present herein a comprehensive review of the literature on the genetics of the phenylpropanoid pathway, as well as on the breeding strategies available to increase their content in the fruits of tomato, eggplant and pepper. We compiled ~267 QTLs linked to phenolic acids, flavonoids and total phenolics content in the three Solanaceae species, along with many candidate genes putatively controlling those QTLs effects. Tomato stands out as the most studied species (217 out of 267 QTLs), a result of its status as model species, whereas eggplant (16 out of 267 QTLs) and pepper (34 out of 267 QTLs) are clearly lagging behind. Notwithstanding, the availability of genomic resources, such as synteny maps and high-quality annotated genomes, has undoubtably contributed to boosting our understanding of the phenylpropanoid pathway and closing the gap among tomato, eggplant and pepper, in spite of the disparity in available genetic resources. Due to their important roles in plant health and development, as well as in the human body, phenolic compounds are gaining interest in breeding programs. We hope that this review facilitates the development of new tomato, eggplant and pepper varieties with improved phenolic profile and greater functional potential. To that end, breeders must take into account not only the qualitative differences among these crops but also the genetic components involved. Hence, structural genes, transcriptional and post-transcriptional regulators, including promoters, repressors, cellular transporters and other types of proteins, should all be taken into account to increase the phenolics content in tomato, eggplant and pepper. Furthermore, the qualitative variations are intrinsically linked to fruit development and are tissue-specific, hence, an accurate understanding on how transcript levels change with these and other factors, would translate into far better results. Identification and analyses of allelic variation of many of the candidate genes governing phenolics accumulation QTLs remains a task at hand, as well as their putative epistatic interactions with other QTLs and genes. Precise genome editing along with next-generation sequencing techniques are important tools towards that end. Ultimately, the development of functional markers could transform the way we improve phenolics content in the fruits of tomato, eggplant and pepper.

## Author contributions

ER-M and LP-D did the literature research, drafted the manuscript and made the tables and figures. AB supervised the literature research and helped to structure the manuscript. JP, MP, YT, and AB provided comments and helped in the writing of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by grants CIPROM/2021/020, funded by Conselleria d'Innovació, Universitats, Ciència i Societat Digital (Generalitat Valenciana, Spain), and PID2021-128148OB-I00, funded by MCIN/AEI/10.13039/501100011033/ and by "ESF Investing in your future", as well as by European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 677379 (G2P-SOL project). ER-M is grateful to MCIN/AEI/10.13039/501100011033/ and "ESF Investing in your future" for a pre-doctoral grant (BES-2016-077482). LP-D is grateful to the Spanish Ministerio de Universidades for a Margarita Salas postdoctoral grant funded by the European Union NextGenerationEU plan.

## Acknowledgments

The authors would like to thank Pietro Gramazio for his assistance in the production of the figures.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1135237/full#supplementary-material

## References

Acquadro, A., Barchi, L., Portis, E., Nourdine, M., Carli, C., Monge, S., et al. (2020). Whole genome resequencing of four Italian sweet pepper landraces provides insights on sequence variation in genes of agronomic value. *Sci. Rep.* 10, 9189. doi: 10.1038/s41598-020-66053-2

Adato, A., Mandel, T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., et al. (2009). Fruit-surface flavonoid accumulation in tomato is controlled by a *SlMYB12*-regulated transcriptional network. *PloS Genet.* 5 (12), e1000777. doi: 10.1371/journal.pgen.1000777

Aguilar-Barragán, A., and Ochoa-Alejo, N. (2014). Virus-induced silencing of *MYB* and *WD40* transcription factor genes affects the accumulation of anthocyanins in chilli pepper fruit. *Biol. Plant* 58 (3), 567–574. doi: 10.1007/s10535-014-0427-4

Ahmed, O. M., Hassan, M. A., Abdel-Twab, S. M., and Abdel Azeem, M. N. (2017). Navel orange peel hydroethanolic extract, naringin and naringenin have anti-diabetic potentials in type 2 diabetic rats. *Biomed. Pharmacother.* 94, 197–205. doi: 10.1016/j.biopha.2017.07.094

Alarcón-Flores, M. I., Romero-González, R., Martínez Vidal, J. L., and Garrido Frenich, A. (2016). Multiclass determination of phenolic compounds in different varieties of tomato and lettuce by ultra high performance liquid chromatography coupled to tandem mass spectrometry. *Int. J. Food Prop.* 19 (3), 494–507. doi: 10.1080/10942912.2014.978010

Alseekh, S., Tohge, T., Wendenberg, R., Scossa, F., Omranian, N., Li, J., et al. (2015). Identification and mode of inheritance of quantitative trait loci for secondary metabolite abundance in tomato. *Plant Cell* 27 (3), 485–512. doi: 10.1105/tpc.114.132266

Arce-Rodríguez, M. L., Martínez, O., and Ochoa-Alejo, N. (2021). Genome-wide identification and analysis of the *MYB* transcription factor gene family in chili pepper (*Capsicum* spp.). *Int. J. Mol. Sci.* 22, 2229. doi: 10.3390/ijms22052229

Aza-González, C., Herrera-Isidrón, L., Núñez-Palenius, H. G., de la Vega, O. M., and Ochoa-Alejo, N. (2013). Anthocyanin accumulation and expression analysis of biosynthesis-related genes during chili pepper fruit development. *Biol. Plant* 57, 49–55. doi: 10.1007/s10535-012-0265-1

Azari, R., Tadmor, Y., Meir, A., Reuveni, M., Evenor, D., Nahon, S., et al. (2010). Light signaling genes and their manipulation towards modulation of phytonutrient content in tomato fruits. *Biotechnol. Adv.* 28, 108–118. doi: 10.1016/j.biotechadv.2009.10.003

Ballester, A.-R., Molthoff, J., de Vos, R., Hekkert, B., te, L., Orzaez, D., et al. (2010). Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant Physiol.* 152, 71–84. doi: 10.1104/pp.109.147322

Ballester, A. R., Tikunov, Y., Molthoff, J., Grandillo, S., Viquez-Zamora, M., de Vos, R., et al. (2016). Identification of loci affecting accumulation of secondary metabolites in tomato fruit of a *Solanum lycopersicum* × *Solanum chmielewskii* introgression line population. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01428

Barchi, L., Pietrella, M., Venturini, L., Minio, A., Toppino, L., Acquadro, A., et al. (2019). A chromosome-anchored eggplant genome sequence reveals key events in solanaceae evolution. *Sci. Rep.* 9, 11769. doi: 10.1038/s41598-019-47985-w

Barchi, L., Rabanus-Wallace, M. T., Prohens, J., Toppino, L., Padmarasu, S., Portis, E., et al. (2021). Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. *Plant J.* 107 (2), 579–596. doi: 10.1111/tpj.15313

Barraj Barraj, R., Segado, P., Moreno-González, R., Heredia, A., Fernández-Muñoz, R., and Domínguez, E. (2021). Genome-wide QTL analysis of tomato fruit cuticle deposition and composition. *Hortic. Res.* 8, 113. doi: 10.1038/s41438-021-00548-5

Bino, R. J., Ric De Vos, C. H., Lieberman, M., Hall, R. D., Bovy, A., Jonker, H. H., et al. (2005). The light-hyperresponsive *high pigment - 2dg* mutation of tomato: alterations in the fruit metabolome. *New Phytol.* 166 (2), 427–438. doi: 10.1111/j.1469-8137.2005.01362.x

Biswas, S., Zhang, D., and Shi, J. (2021). CRISPR/Cas systems: opportunities and challenges for crop breeding. *Plant Cell Rep.* 40, 979–998. doi: 10.1007/S00299-021-02708-2

Bovy, A., De Vos, R., Kemper, M., Schijlen, E., Almenar Pertejo, M., Muir, S., et al. (2002). High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes *LC* and *C1*. *Plant Cell* 14, 2509–2526. doi: 10.1105/tpc.004218

Bovy, A., Elio, A. E., Ae, S., and Hall, R. D. (2007). Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. *Metabolomics* 3, 399–412. doi: 10.1007/s11306-007-0074-2

Bovy, A. G., Gómez-Roldán, V., and Hall, R. D. (2010). "Strategies to optimize the flavonoid content of tomato fruit," in *Recent advances in polyphenol research*. Eds. C. Santos-Buelga, M. T. Escribano-Bailon and V. Lattanzio (Chichester, UK: Blackwell Publishing Ltd), 138–162. doi: 10.1002/9781119427896

Calumpang, C. L. F., Saigo, T., Watanabe, M., and Tohge, T. (2020). Cross-species comparison of fruit-metabolomics to elucidate metabolic regulation of fruit polyphenolics among solanaceous crops. *Metabolites* 10, 209. doi: 10.3390/metabo10050209

Chattopadhyay, T., Hazra, P., Akhtar, S., Maurya, D., Mukherjee, A., and Roy, S. (2021). Skin colour, carotenogenesis and chlorophyll degradation mutant alleles: Genetic orchestration behind the fruit colour variation in tomato. *Plant Cell Rep.* 40, 767–782. doi: 10.1007/s00299-020-02650-9

Chen, L., and Kang, Y. H. (2013). Anti-inflammatory and antioxidant activities of red pepper (*Capsicum annuum* l.) stalk extracts: Comparison of pericarp and placenta extracts. *J. Funct. Foods* 5, 1724–1731. doi: 10.1016/j.jff.2013.07.018

Chioti, V., Zeliou, K., Bakogianni, A., Papaioannou, C., Biskinis, A., Petropoulos, C., et al. (2022). Nutritional value of eggplant cultivars and association with sequence variation in genes coding for major phenolics. *Plants* 11, 2267. doi: 10.3390/PLANTS11172267/S1

Clifford, M. N., Jaganath, I. B., Ludwig, I. A., and Crozier, A. (2017). Chlorogenic acids and the acyl-quinic acids: Discovery, biosynthesis, bioavailability and bioactivity. *Nat. Prod. Rep.* 34, 1391–1421. doi: 10.1039/c7np00030h

Colanero, S., Perata, P., and Gonzali, S. (2020). What's behind purple tomatoes? insight into the mechanisms of anthocyanin synthesis in tomato fruits. *Plant Physiol.* 182, 1841–1853. doi: 10.1104/pp.19.01530

Colliver, S., Bovy, A., Collins, G., Muir, S., Robinson, S., De Vos, C. H. R., et al. (2002). Improving the nutritional content of tomatoes through reprogramming their flavonoid biosynthetic pathway. *Phytochem. Rev.* 1, 113–123. doi: 10.1023/A:1015848724102

Colonna, V., D'Agostino, N., Garrison, E., Albrechtsen, A., Meisner, J., Facchiano, A., et al. (2019). Genomic diversity and novel genome-wide association with fruit morphology in *Capsicum*, from 746k polymorphic sites. *Sci. Rep.* 9, 10067. doi: 10.1038/s41598-019-46136-5

Cosme, P., Rodríguez, A. B., Espino, J., and Garrido, M. (2020). Plant phenolics: Bioavailability as a key determinant of their potential health-promoting applications. *Antioxidants* 9 (12), 1263. doi: 10.3390/antiox9121263

Dadáková, K., Heinrichová, T., Lochman, J., and Kašparovský, T. (2020). Production of defense phenolics in tomato leaves of different age. *Molecules* 25, 4952. doi: 10.3390/molecules25214952

Davuluri, G. R., Tuinen, A., Fraser, P. D., Manfredonia, A., Newman, R., Burgess, D., et al. (2005). Fruit-specific RNAi-mediated suppression of *DET1* enhances carotenoid and flavonoid content in tomatoes. *Nat. Biotechnol.* 23, 1–13. doi: 10.1038/nbt1108.Fruit-specific

Deng, L., Wang, H., Sun, C., Li, Q., Jiang, H., Du, M., et al. (2018). Efficient generation of pink-fruited tomatoes using CRISPR/Cas9 system. *J. Genet. Genomics* 45, 51–54. doi: 10.1016/J.JGG.2017.10.002

Derlindati, E., Dall'Asta, M., Ardigò, D., Brighenti, F., Zavaroni, I., Crozier, A., et al. (2012). Quercetin-3-O-glucuronide affects the gene expression profile of M1 and M2a human macrophages exhibiting anti-inflammatory effects. *Food Funct.* 3, 1144–1152. doi: 10.1039/c2fo30127j

Dias, M. C., Pinto, D. C. G. A., and Silva, A. M. S. (2021). Plant flavonoids: Chemical characteristics and biological activity. *Molecules* 26, 5377. doi: 10.3390/molecules26175377

Di Matteo, A., Ruggieri, V., Sacco, A., Rigano, M. M., Carriero, F., Bolger, A., et al. (2013). Identification of candidate genes for phenolics accumulation in tomato fruit. *Plant Sci.* 205–206, 87–96. doi: 10.1016/j.plantsci.2013.02.001

Di Matteo, A., Sacco, A., Ruggieri, V., Trotta, N., Nunziata, A., and Barone, A. (2010). Transcriptional network controlling antioxidants in tomato fruit. *J. Biotechnol.* 150S, S1–S576. doi: 10.1016/j.jbiotec.2010.08.287

Diretto, G., Frusciante, S., Fabbri, C., Schauer, N., Busta, L., Wang, Z., et al. (2020). Manipulation of $\beta$-carotene levels in tomato fruits results in increased ABA content and extended shelf life. *Plant Biotechnol. J.* 18, 1185–1199. doi: 10.1111/pbi.13283

Docimo, T., Francese, G., Ruggiero, A., Batelli, G., De Palma, M., Bassolino, L., et al. (2016). Phenylpropanoids accumulation in eggplant fruit: Characterization of biosynthetic genes and regulation by a MYB transcription factor. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.01233

Doganlar, S., Frary, A., Daunay, M. C., Lester, R. N., and Tanksley, S. D. (2002). A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the solanaceae. *Genetics* 161, 1697–1711. doi: 10.1093/genetics/161.4.1697

Dong, N., and Lin, H. (2021). Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *J. Integr. Plant Biol.* 63, 180–209. doi: 10.1111/jipb.13054

Drewnowski, A., and Gomez-Carneros, C. (2000). Bitter taste, phytonutrients, and the consumer: A review. *Am. J. Clin. Nutr.* 72, 1424–1435. doi: 10.1093/ajcn/72.6.1424

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

España, L., Heredia-Guerrero, J. A., Reina-Pinto, J. J., Fernández-Muñoz, R., Heredia, A., and Domínguez, E. (2014). Transient silencing of CHALCONE SYNTHASE during fruit ripening modifies tomato epidermal cells and cuticle properties. *Plant Physiol.* 166, 1371–1386. doi: 10.1104/pp.114.246405

Fernandez-Moreno, J. P., Tzfadia, O., Forment, J., Presa, S., Rogachev, I., Meir, S., et al. (2016). Characterization of a new pink-fruited tomato mutant results in the identification of a null allele of the SlMYB12 transcription factor. *Plant Physiol.* 171, 1821–1836. doi: 10.1104/pp.16.00282

Fratianni, F., D'acierno, A., Cozzolino, A., Spigno, P., Riccardi, R., Raimo, F., et al. (2020). Biochemical characterization of traditional varieties of sweet pepper (*Capsicum annuum* l.) of the campania region, southern Italy. *Antioxidants* 9 (6), 556. doi: 10.3390/antiox9060556

Galano, A., Mazzone, G., Alvarez-Diduk, R., Marino, T., Alvarez-Idaboy, J. R., and Russo, N. (2016). Food antioxidants: Chemical insights at the molecular level. *Annu. Rev. Food Sci. Technol.* 7, 335–352. doi: 10.1146/annurev-food-041715-033206

Gammoudi, N., Pedro, T. S., Ferchichi, A., and Gisbert, C. (2017). Improvement of regeneration in pepper: A recalcitrant species. *Vitr. Cell. Dev. Biol. - Plant* 54, 145–153. doi: 10.1007/S11627-017-9838-1

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2

García-Fortea, E., García-Pérez, A., Gimeno-Páez, E., Martínez-López, M., Vilanova, S., Gramazio, P., et al. (2021)."Ploidy modification for plant breeding using in vitro organogenesis: A case in eggplant," in *Crop breeding*. Ed. P. Tripodi (New York, NY: Humana), 197–206. doi: 10.1007/978-1-0716-1201-9_14

García-Fortea, E., Gramazio, P., Vilanova, S., Fita, A., Mangino, G., Villanueva, G., et al. (2019). First successful backcrossing towards eggplant (*Solanum melongena*) of a new world species, the silverleaf nightshade (*S. elaeagnifolium*), and characterization of interspecific hybrids and backcrosses. *Sci. Hortic.* 246, 563–573. doi: 10.1016/j.scienta.2018.11.018

García-Valverde, V., Navarro-González, I., García-Alonso, J., and Periago, M. J. (2013). Antioxidant bioactive compounds in selected industrial processing and fresh consumption tomato cultivars. *Food Bioprocess Technol.* 6, 391–402. doi: 10.1007/s11947-011-0687-3

Gebhardt, C. (2016). The historical role of species from the solanaceae plant family in genetic research. *Theor. Appl. Genet.* 129, 2281–2294. doi: 10.1007/s00122-016-2804-1

Gramazio, P., Prohens, J., Plazas, M., Andujar, I., Herraiz, F. J., Castillo, E., et al. (2014). Location of chlorogenic acid biosynthesis pathway and polyphenol oxidase genes in a new interspecific anchored linkage map of eggplant. *BMC Plant Biol.* 14, 350. doi: 10.1186/s12870-014-0350-z

Gramazio, P., Prohens, J., Plazas, M., Mangino, G., Herraiz, F. J., and Vilanova, S. (2017). Development and genetic characterization of advanced backcross materials and an introgression line population of *Solanum incanum* in a *S. melongena* background. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01477

Hallmann, E., and Rembialkowska, E. (2012). Characterisation of antioxidant compounds in sweet bell pepper (*Capsicum annuum* l.) under organic and conventional growing systems. *J. Sci. Food Agric.* 92, 2409–2415. doi: 10.1002/jsfa.5624

Hanson, P., Schafleitner, R., Huang, S. M., Tan, C. W., Ledesma, D., and Yang, R. Y. (2014). Characterization and mapping of a QTL derived from *Solanum habrochaites* associated with elevated rutin content (quercetin-3-rutinoside) in tomato. *Euphytica* 200, 441–454. doi: 10.1007/s10681-014-1180-7

Heredia, A., Heredia-Guerrero, J. A., and Domínguez, E. (2015). *CHS* silencing suggests a negative cross-talk between wax and flavonoid pathways in tomato fruit cuticle. *Plant Signal. Behav.* 10 (5), e1019979. doi: 10.1080/15592324.2015.1019979

Hirakawa, H., Shirasawa, K., Miyatake, K., Nunome, T., Negoro, S., Ohyama, A., et al. (2014). Draft genome sequence of eggplant (*Solanum melongena* l.): The representative *Solanum* species indigenous to the old world. *DNA Res.* 21, 649–660. doi: 10.1093/dnares/dsu027

Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., Hill, T. A., Jaffe, D., Williams, S. R., et al. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Hortic. Res.* 5, 4. doi: 10.1038/s41438-017-0011-0

Iijima, Y., Nakamura, Y., Ogata, Y., Tanaka, I., Sakurai, N., Suda, K., et al. (2008). Metabolite annotations based on the integration of mass spectral information. *Plant J.* 54, 949–962. doi: 10.1111/j.1365-313X.2008.03434.x

Islam, K., Rawoof, A., Ahmad, I., Dubey, M., Momo, J., and Ramchiary, N. (2021). *Capsicum chinense MYB* transcription factor genes: Identification, expression analysis, and their conservation and diversification with other solanaceae genomes. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.721265

Jeong, W. Y., Jin, J. S., Cho, Y. A., Lee, J. H., Park, S., Jeong, S. W., et al. (2011). Determination of polyphenols in three *Capsicum annuum* l. (bell pepper) varieties using high-performance liquid chromatography tandem mass spectrometry: Their contribution to overall antioxidant and anticancer activity. *J. Sep. Sci.* 34, 2967–2974. doi: 10.1002/jssc.201100524

Jung, H. J., Veerappan, K., Hwang, I., Goswami, G., Chung, M. Y., and Nou, I. S. (2017). New SNPs and InDel variations in *SlMYB12* associated with regulation of pink color in tomato. *Trop. Plant Biol.* 10, 126–133. doi: 10.1007/s12042-017-9191-x

Kajikawa, M., Maruhashi, T., Hidaka, T., Nakano, Y., Kurisu, S., Matsumoto, T., et al. (2019). Coffee with a high content of chlorogenic acids and low content of hydroxyhydroquinone improves postprandial endothelial dysfunction in patients with borderline and stage 1 hypertension. *Eur. J. Nutr.* 58, 989–996. doi: 10.1007/s00394-018-1611-7

Kaur, C., Nagal, S., Nishad, J., Kumar, R., and Sarika, (2014). Evaluating eggplant (*Solanum melongena* l) genotypes for bioactive properties: A chemometric approach. *Food Res. Int.* 60, 205–211. doi: 10.1016/j.foodres.2013.09.049

Kaushik, P., Gramazio, P., Vilanova, S., Raigón, M. D., Prohens, J., and Plazas, M. (2017). Phenolics content, fruit flesh colour and browning in cultivated eggplant, wild relatives and interspecific hybrids and implications for fruit quality breeding. *Food Res. Int.* 102, 392–401. doi: 10.1016/j.foodres.2017.09.028

Kaushik, P., Kumar, P., and Kumar, S. (2020). Enhancement of chlorogenic content of the eggplant fruit with eggplant hydroxycinnamoyl CoA-quinate transferase gene *via* novel agroinfiltration protocol. *Pharmacogn. Mag.* 16, S450–S454. doi: 10.4103/pm.pm_537_19

Kaushik, P., Prohens, J., Vilanova, S., Gramazio, P., and Plazas, M. (2016). Phenotyping of eggplant wild relatives and interspecific hybrids with conventional and phenomics descriptors provides insight for their potential utilization in breeding. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00677

Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J. M., Lee, H.-A., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* 46, 270–278. doi: 10.1038/ng.2877

Kim, S., Park, J., Yeom, S. I., Kim, Y. M., Seo, E., Kim, K. T., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 18, 210. doi: 10.1186/s13059-017-1341-9

Ku, Y. S., Ng, M. S., Cheng, S. S., Lo, A. W. Y., Xiao, Z., Shin, T. S., et al. (2020). Understanding the composition, biosynthesis, accumulation and transport of flavonoids in crops for the promotion of crops as healthy sources of flavonoids for human consumption. *Nutrients* 12, 1717. doi: 10.3390/nu12061717

Lee, J. (2019). "Development and evolution of molecular markers and genetic maps in capsicum species," in *The capsicum genome*. Eds. N. Ramchiary and C. Kole (Cham: Springer), 85–103. doi: 10.1007/978-3-319-97217-6_5

Lee, H.-Y., Ro, N.-Y., Patil, A., Lee, J.-H., Kwon, J.-K., and Kang, B.-C. (2020). Uncovering candidate genes controlling major fruit-related traits in pepper *via* genotype-by-sequencing based QTL mapping and genome-wide association study. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01100

Lemos, V. C., Reimer, J. J., and Wormit, A. (2019). Color for life: Biosynthesis and distribution of phenolic compounds in pepper (*Capsicum annuum*). *Agriculture* 9, 81. doi: 10.3390/agriculture9040081

Levin, I., Ric De Vos, C. H., Tadmor, Y., Bovy, A., Lieberman, M., Oren-Shamir, M., et al. (2006). High pigment tomato mutants - more than just lycopene (a review). *Isr. J. Plant Sci.* 54, 179–190. doi: 10.1560/IJPS_54_3_179

Li, Y., Chen, M., Wang, S., Ning, J., Ding, X., and Chu, Z. (2015). *AtMYB11* regulates caffeoylquinic acid and flavonol synthesis in tomato and tobacco. *Plant Cell. Tissue Organ Cult.* 122, 309–319. doi: 10.1007/s11240-015-0767-6

Li, J., He, Y. J., Zhou, L., Liu, Y., Jiang, M., Ren, L., et al. (2018a). Transcriptome profiling of genes related to light-induced anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) before purple color becomes evident. *BMC Genomics* 19, 201. doi: 10.1186/s12864-018-4587-z

Li, R., Li, R., Li, X., Fu, D., Zhu, B., Tian, H., et al. (2018b). Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ-aminobutyric acid levels in *Solanum lycopersicum*. *Plant Biotechnol. J.* 16, 415–427. doi: 10.1111/pbi.12781

Li, D., Qian, J., Li, W., Yu, N., Gan, G., Jiang, Y., et al. (2021). A high-quality genome assembly of the eggplant provides insights into the molecular basis of disease resistance and chlorogenic acid synthesis. *Mol. Ecol. Resour.* 21, 1274–1286. doi: 10.1111/1755-0998.13321

Lim, W., and Li, J. (2017a). Co-Expression of onion *chalcone isomerase* in *Del/Ros1*-expressing tomato enhances anthocyanin and flavonol production. *Plant Cell. Tissue Organ Cult.* 128, 113–124. doi: 10.1007/s11240-016-1090-6

Lim, W., and Li, J. (2017b). Synergetic effect of the onion *CHI* gene on the *PAP1* regulatory gene for enhancing the flavonoid profile of tomato skin. *Sci. Rep.* 7, 12377. doi: 10.1038/s41598-017-12355-x

Liu, Y., Lv, J., Liu, Z., Wang, J., Yang, B., Chen, W., et al. (2020). Integrative analysis of metabolome and transcriptome reveals the mechanism of color formation in pepper fruit (*Capsicum annuum* l.). *Food Chem.* 306, 125629. doi: 10.1016/j.foodchem.2019.125629

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in solanaceous vegetables: A review. *Front. Chem.* 6. doi: 10.3389/fchem.2018.00052

Long, M., Millar, D. J., Kimura, Y., Donovan, G., Rees, J., Fraser, P. D., et al. (2006). Metabolite profiling of carotenoid and phenolic pathways in mutant and transgenic lines of tomato: Identification of a high antioxidant fruit line. *Phytochemistry* 67, 1750–1757. doi: 10.1016/j.phytochem.2006.02.022

López-Velázquez, J. G., Delgado-Vargas, F., López-Ángulo, G., García-Armenta, E., López-López, M. E., Ayón-Reyna, L. E., et al. (2020). Phenolic profile associated with chilling tolerance induced by the application of a hot water treatment in bell pepper fruit. *J. Food Sci.* 85, 2080–2089. doi: 10.1111/1750-3841.15310

Lo Scalzo, R., Florio, F. E., Fibiani, M., Speranza, G., Rabuffetti, M., Gattolin, S., et al. (2021). Scrapped but not neglected: Insights into the composition, molecular modulation and antioxidant capacity of phenols in peel of eggplant (*Solanum melongena* l.) fruits at different developmental stages. *Plant Physiol. Biochem.* 167, 678–690. doi: 10.1016/j.plaphy.2021.08.037

Luo, J., Butelli, E., Hill, L., Parr, A., Niggeweg, R., Bailey, P., et al. (2008). AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. *Plant J.* 56, 316–326. doi: 10.1111/j.1365-313X.2008.03597.x

Luo, Y., Shang, P., and Li, D. (2017). Luteolin: A flavonoid that has multiple cardio-protective effects and its molecular mechanisms. *Front. Pharmacol.* 8. doi: 10.3389/fphar.2017.00692

Luthria, D. L., Mukhopadhyay, S., and Krizek, D. T. (2006). Content of total phenolics and phenolic acids in tomato (*Lycopersicon esculentum* mill.) fruits as influenced by cultivar and solar UV radiation. *J. Food Compos. Anal.* 19, 771–777. doi: 10.1016/j.jfca.2006.04.005

Luthria, D., Singh, A. P., Wilson, T., Vorsa, N., Banuelos, G. S., and Vinyard, B. T. (2010). Influence of conventional and organic agricultural practices on the phenolic content in eggplant pulp: Plant-to-plant variation. *Food Chem.* 121, 406–411. doi: 10.1016/j.foodchem.2009.12.055

Maioli, A., Gianoglio, S., Moglia, A., Acquadro, A., Valentino, D., Milani, A. M., et al. (2020). Simultaneous CRISPR/Cas9 editing of three PPO genes reduces fruit flesh browning in *Solanum melongena* l. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.607161

Marchiosi, R., dos Santos, W. D., Constantin, R. P., de Lima, R. B., Soares, A. R., Finger-Teixeira, A., et al. (2020). Biosynthesis and metabolic actions of simple phenolic acids in plants. *Phytochem. Rev.* 19, 865–906. doi: 10.1007/s11101-020-09689-2

Marín, A., Ferreres, F., Tomás-Barberán, F. A., and Gil, M. I. (2004). Characterization and quantitation of antioxidant constituents of sweet pepper (*Capsicum annuum* l.). *J. Agric. Food Chem.* 52, 3861–3869. doi: 10.1021/jf0497915

Martí, R., Leiva-Brondo, M., Lahoz, I., Campillo, C., Cebolla-Cornejo, J., and Roselló, S. (2018). Polyphenol and l-ascorbic acid content in tomato as influenced by high lycopene genotypes and organic farming at different environments. *Food Chem.* 239, 148–156. doi: 10.1016/j.foodchem.2017.06.102

Menda, N., Semel, Y., Peled, D., Eshed, Y., and Zamir, D. (2004). *In silico* screening of a saturated mutation library of tomato. *Plant J.* 38, 861–872. doi: 10.1111/j.1365-313X.2004.02088.x

Mennella, G., Lo Scalzo, R., Fibiani, M., DAlessandro, A., Francese, G., Toppino, L., et al. (2012). Chemical and bioactive quality traits during fruit ripening in eggplant (*S. melongena* l.) and allied species. *J. Agric. Food Chem.* 60, 11821–11831. doi: 10.1021/jf3037424

Mennella, G., Rotino, G. L., Fibiani, M., D'Alessandro, A., Franceses, G., Toppino, L., et al. (2010). Characterization of health-related compounds in eggplant (*Solanum melongena* l.) lines derived from introgression of allied species. *J. Agric. Food Chem.* 58, 7597–7603. doi: 10.1021/jf101004z

Meyer, R. S., Little, D. P., Whitaker, B. D., and Litt, A. (2019). "The genetics of eggplant nutrition," in *The eggplant genome. compendium of plant genomes*. Ed. M. A. Chapman (Cham: Springer), 23–32. doi: 10.1007/978-3-319-99208-2_3

Mishra, R., Mohanty, J. N., Mahanty, B., and Joshi, R. K. (2021). A single transcript CRISPR/Cas9 mediated mutagenesis of *CaERF28* confers anthracnose resistance in chilli pepper (*Capsicum annuum* l.). *Planta* 254, 5. doi: 10.1007/s00425-021-03660-x

Moco, S., Bino, R. J., Vorst, O., Verhoeven, H. A., De Groot, J., Van Beek, T. A., et al. (2006). A liquid chromatography-mass spectrometry-based metabolome database for tomato. *Plant Physiol.* 141, 1205–1218. doi: 10.1104/pp.106.078428

Moglia, A., Lanteri, S., Comino, C., Hill, L., Knevitt, D., Cagliero, C., et al. (2014). Dual catalytic activity of hydroxycinnamoyl-coenzyme a quinate transferase from tomato allows it to moonlight in the synthesis of both mono- and dicaffeoylquinic acids. *Plant Physiol.* 166, 1777–1787. doi: 10.1104/pp.114.251371

Mokhtar, M., Soukup, J., Donato, P., Cacciola, F., Dugo, P., Riazi, A., et al. (2015). Determination of the polyphenolic content of a *Capsicum annuum* l. extract by liquid chromatography coupled to photodiode array and mass spectrometry detection and evaluation of its biological activity. *J. Sep. Sci.* 38, 171–178. doi: 10.1002/jssc.201400993

Monforte, A. J., and Tanksley, S. D. (2000). Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: A tool for gene mapping and gene discovery. *Genome* 43, 803–813. doi: 10.1139/g00-043

Morales-Soto, A., García-Salas, P., Rodríguez-Pérez, C., Jiménez-Sánchez, C., Cádiz-Gurrea, M. L., Segura-Carretero, A., et al. (2014). Antioxidant capacity of 44 cultivars of fruits and vegetables grown in Andalusia (Spain). *Food Res. Int.* 58, 35–46. doi: 10.1016/j.foodres.2014.01.050

Morales-Soto, A., Gómez-Caravaca, A. M., García-Salas, P., Segura-Carretero, A., and Fernández-Gutiérrez, A. (2013). High-performance liquid chromatography coupled to diode array and electrospray time-of-flight mass spectrometry detectors for a comprehensive characterization of phenolic and other polar compounds in three pepper (*Capsicum annuum* l.) samples. *Food Res. Int.* 51, 977–984. doi: 10.1016/j.foodres.2013.02.022

Moreno-Ramírez, Y. D. R., Martínez-Ávila, G. C. G., González-Hernández, V. A., Castro-López, C., and Torres-Castillo, J. A. (2018). Free radical-scavenging capacities, phenolics and capsaicinoids in wild piquin chili (*Capsicum annuum* var. *glabriusculum*). *Molecules* 23, 2655. doi: 10.3390/molecules23102655

Mori, T., Umeda, T., Honda, T., Zushi, K., Wajima, T., and Matsuzoe, N. (2013). Varietal differences in the chlorogenic acid, anthocyanin, soluble sugar, organic acid, and amino acid concentrations of eggplant fruit. *J. Hortic. Sci. Biotechnol.* 88, 657–663. doi: 10.1080/14620316.2013.11513021

Mubarak, A., Bondonno, C. P., Liu, A. H., Considine, M. J., Rich, L., Mas, E., et al. (2012). Acute effects of chlorogenic acid on nitric oxide status, endothelial function, and blood pressure in healthy volunteers: A randomized trial. *J. Agric. Food Chem.* 60, 9130–9136. doi: 10.1021/jf303440j

Mudrić, S., Gašić, U. M., Dramićanin, A. M., Ćirić, I., Milojković-Opsenica, D. M., Popović-Đorđević, J. B., et al. (2017). The polyphenolics and carbohydrates as indicators of botanical and geographical origin of Serbian autochthonous clones of red spice paprika. *Food Chem.* 217, 705–715. doi: 10.1016/j.foodchem.2016.09.038

Muhlemann, J. K., Younts, T. L. B., and Muday, G. K. (2018). Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. *Proc. Natl. Acad. Sci. U. S. A.* 115, E11188–E11197. doi: 10.1073/pnas.1811492115

Muir, S. R., Collins, G. J., Robinson, S., Hughes, S., Bovy, A., Ric De Vos, C. H., et al. (2001). Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nat. Biotechnol.* 19, 470–474. doi: 10.1038/88150

Niggeweg, R., Michael, A. J., and Martin, C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat. Biotechnol.* 22, 746–754. doi: 10.1038/nbt966

Ofner, I., Lashbrooke, J., Pleban, T., Aharoni, A., and Zamir, D. (2016). *Solanum pennellii* backcross inbred lines (BILs) link small genomic bins with tomato traits. *Plant J.* 87, 151–160. doi: 10.1111/tpj.13194

Ökmen, B., Şiğva, H.Ö., Gürbüz, N., Ülger, M., and Frary, A. (2011). Quantitative trait loci (QTL) analysis for antioxidant and agronomically important traits in tomato (*Lycopersicon esculentum*). *Turkish J. Agric. For.* 35, 501–514. doi: 10.3906/tar-1008-1207

Olmstead, R. G., Bohs, L., Migid, H. A., Santiago-Valentin, E., Garcia, V. F., and Collier, S. M. (2008). A molecular phylogeny of the solanaceae. *Taxon* 57, 1159–1181. doi: 10.2307/3284540

Ou, L., Li, D., Lv, J., Chen, W., Zhang, Z., Li, X., et al. (2018). Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence–absence variation analyses. *New Phytol.* 220, 360–363. doi: 10.1111/NPH.15413

Özçelik, B., Kartal, M., and Orhan, I. (2011). Cytotoxicity, antiviral and antimicrobial activities of alkaloids, flavonoids, and phenolic acids. *Pharm. Biol.* 49, 396–402. doi: 10.3109/13880209.2010.519390

Park, D., Barka, G. D., Yang, E. Y., Cho, M. C., Yoon, J. B., and Lee, J. (2020). Identification of QTLs controlling α-glucosidase inhibitory activity in pepper (*Capsicum annuum* l.) leaf and fruit using genotyping-by-sequencing analysis. *Genes (Basel).* 11, 1116. doi: 10.3390/genes11101116

Pelinson, L. P., Assmann, C. E., Palma, T. V., da Cruz, I. B. M., Pillat, M. M., Mânica, A., et al. (2019). Antiproliferative and apoptotic effects of caffeic acid on SK-Mel-28 human melanoma cancer cells. *Mol. Biol. Rep.* 46, 2085–2092. doi: 10.1007/s11033-019-04658-1

Plazas, M., López-Gresa, M. P., Vilanova, S., Torres, C., Hurtado, M., Gramazio, P., et al. (2013). Diversity and relationships in key traits for functional and apparent quality in a collection of eggplant: Fruit phenolics content, antioxidant activity, polyphenol oxidase activity, and browning. *J. Agric. Food Chem.* 61, 8871–8879. doi: 10.1021/JF402429K/ASSET/IMAGES/MEDIUM/JF-2013-02429K_0004.GIF

Plazas, M., Prohens, J., Cuñat, A. N., Vilanova, S., Gramazio, P., Herraiz, F. J., et al. (2014). Reducing capacity, chlorogenic acid content and biological activity in a collection of scarlet (*Solanum aethiopicum*) and gboma (*S. macrocarpon*) eggplants. *Int. J. Mol. Sci.* 15, 17221–17241. doi: 10.3390/ijms151017221

Pott, D. M., Osorio, S., and Vallarino, J. G. (2019). From central to specialized metabolism: An overview of some secondary compounds derived from the primary metabolism for their role in conferring nutritional and organoleptic characteristics to fruit. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00835

Povero, G., Gonzali, S., Bassolino, L., Mazzucato, A., and Perata, P. (2011). Transcriptional analysis in high-anthocyanin tomatoes reveals synergistic effect of *Aft* and *atv* genes. *J. Plant Physiol.* 168, 270–279. doi: 10.1016/J.JPLPH.2010.07.022

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Díez, M. J., et al. (2017). Introgressiomics: A new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica* 213, 158. doi: 10.1007/s10681-017-1938-9

Prohens, J., Plazas, M., Raigón, M. D., Seguí-Simarro, J. M., Stommel, J. R., and Vilanova, S. (2012). Characterization of interspecific hybrids and first backcross generations from crosses between two cultivated eggplants (*Solanum melongena* and *S. aethiopicum* kumba group) and implications for eggplant breeding. *Euphytica* 186, 517–538. doi: 10.1007/s10681-012-0652-x

Prohens, J., Whitaker, B. D., Plazas, M., Vilanova, S., Hurtado, M., Blasco, M., et al. (2013). Genetic diversity in morphological characters and phenolic acids content resulting from an interspecific cross between eggplant, *Solanum melongena*, and its wild ancestor (*S. incanum*). *Ann. Appl. Biol.* 162, 242–257. doi: 10.1111/aab.12017

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 5135–5140. doi: 10.1073/pnas.1400975111

Ramaroson, M., Koutouan, C., Helesbeux, J., Hamama, L., Geoffriau, E., and Briard, M. (2022). Role of phenylpropanoids nad flavonoids in plant resistance to pests and diseases. *Molecules* 27, 8371. doi: 10.3390/molecules27238371

Ribes-Moya, A. M., Adalid, A. M., Raigón, M. D., Hellín, P., Fita, A., and Rodríguez-Burruezo, A. (2020). Variation in flavonoids in a collection of peppers (*Capsicum* sp.) under organic and conventional cultivation: effect of the genotype, ripening stage, and growing system. *J. Sci. Food Agric.* 100, 2208–2223. doi: 10.1002/jsfa.10245

Rigano, M. M., Raiola, A., Docimo, T., Ruggieri, V., Calafiore, R., Vitaglione, P., et al. (2016). Metabolic and molecular changes of the phenylpropanoid pathway in tomato (*Solanum lycopersicum*) lines carrying different *Solanum pennellii* wild chromosomal regions. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01484

Rigano, M. M., Raiola, A., Tenore, G. C., Monti, D. M., Giudice, R. D., Frusciante, L., et al. (2014). Quantitative trait loci pyramiding can improve the nutritional potential of tomato (*Solanum lycopersicum*) fruits. *J. Agric. Food Chem.* 62, 11519–11527. doi: 10.1021/jf502573n

Righini, S., Rodriguez, E. J., Berosich, C., Grotewold, E., Casati, P., and Falcone Ferreyra, M. L. (2019). Apigenin produced by maize flavone synthase I and II protects plants against UV-b-induced damage. *Plant Cell Environ.* 42, 495–508. doi: 10.1111/pce.13428

Rinaldi, R., Van Deynze, A., Portis, E., Rotino, G. L., Toppino, L., Hill, T., et al. (2016). New insights on eggplant/tomato/pepper synteny and identification of eggplant and pepper orthologous QTL. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01031

Rosa-Martínez, E., Adalid-Martínez, A. M., Dolores García-Martínez, M., Mangino, G., Raigón, M. D., Plazas, M., et al. (2022a). Fruit composition of eggplant lines with introgressions from the wild relative *S. incanum*: Interest for breeding and safety for consumption. *Agronomy* 12, 266. doi: 10.3390/agronomy12020266

Rosa-Martínez, E., Dolores García-Martínez, M., María Adalid-Martínez, A., Pereira-Dias, L., Casanova, C., Soler, E., et al. (2021). Fruit composition profile of pepper, tomato and eggplant varieties grown under uniform conditions. *Food Res. Int.* 147, 110531. doi: 10.1016/j.foodres.2021.110531

Rosa-Martínez, E., Villanueva, G., Sahin, A., Gramazio, P., García-Martínez, M. D., Raigón, M. D., et al. (2022b). Characterization and QTL identification in eggplant introgression lines under two n fertilization levels. *Hortic. Plant J.* 107545. doi: 10.1016/j.hpj.2022.08.003

Rothan, C., Diouf, I., and Causse, M. (2019). Trait discovery and editing in tomato. *Plant J.* 97, 73–90. doi: 10.1111/tpj.14152

Rousseaux, M. C., Jones, C. M., Adam, D., Chetelat, R., Bennett, A., and Powell, A. (2005). QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. *Theor. Appl. Genet.* 111, 1396–1408. doi: 10.1007/s00122-005-0071-7

Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M. M., Parisi, M., et al. (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biol.* 14, 337. doi: 10.1186/s12870-014-0337-9

Rutley, N., Miller, G., Wang, F., Harper, J. F., Miller, G., and Lieberman-Lazarovich, M. (2021). Enhanced reproductive thermotolerance of the tomato *high pigment 2* mutant is associated with increased accumulation of flavonols in pollen. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.672368

Sacco, A., Di Matteo, A., Lombardi, N., Trotta, N., Punzo, B., Mari, A., et al. (2013). Quantitative trait loci pyramiding for fruit quality traits in tomato. *Mol. Breed.* 31, 217–222. doi: 10.1007/s11032-012-9763-2

Sacco, A., Raiola, A., Calafiore, R., Barone, A., and Rigano, M. M. (2019). New insights in the control of antioxidants accumulation in tomato by transcriptomic analyses of genotypes exhibiting contrasting levels of fruit metabolites. *BMC Genomics* 20, 43. doi: 10.1186/s12864-019-5428-4

Salgotra, R. K., and Neal Stewart, C. (2020). Functional markers for precision plant breeding. *Int. J. Mol. Sci.* 21, 4792. doi: 10.3390/ijms21134792

Santana-Gálvez, J., Villela Castrejón, J., Serna-Saldívar, S. O., and Jacobo-Velázquez, D. A. (2020). Anticancer potential of dihydrocaffeic acid: a chlorogenic acid metabolite. *CYTA - J. Food* 18, 245–248. doi: 10.1080/19476337.2020.1743762

Sapir, M., Oren-Shamir, M., Ovadia, R., Reuveni, M., Evenor, D., Tadmor, Y., et al. (2008). Molecular aspects of anthocyanin fruit tomato in relation to *high pigment-1*. *J. Hered.* 99, 292–303. doi: 10.1093/jhered/esm128

Scarano, A., Chieppa, M., and Santino, A. (2020). Plant polyphenols-biofortified foods as a novel tool for the prevention of human gut diseases. *Antioxidants* 9, 1225. doi: 10.3390/antiox9121225

Schijlen, E., Ric De Vos, C. H., Jonker, H., Van Den Broeck, H., Molthoff, J., Van Tunen, A., et al. (2006). Pathway engineering for healthy phytochemicals leading to the production of novel flavonoids in tomato fruit. *Plant Biotechnol. J.* 4, 433–444. doi: 10.1111/j.1467-7652.2006.00192.x

Schijlen, E. G. W. M., Ric De Vos, C. H., Martens, S., Jonker, H. H., Rosin, F. M., Molthoff, J. W., et al. (2007). RNA Interference silencing of chalcone synthase, the first step in the flavonoid biosynthesis pathway, leads to parthenocarpic tomato fruits. *Plant Physiol.* 144, 1520–1530. doi: 10.1104/pp.107.100305

Scossa, F., Roda, F., Tohge, T., Georgiev, M. I., and Fernie, A. R. (2019). The hot and the colorful: Understanding the metabolism, genetics and evolution of consumer preferred metabolic traits in pepper and related species. *CRC. Crit. Rev. Plant Sci.* 38, 339–381. doi: 10.1080/07352689.2019.1682791

Siddique, M. I., Lee, H.-Y., Ro, N.-Y., Han, K., Venkatesh, J., Solomon, A. M., et al. (2019). Identifying candidate genes for *Phytophthora capsici* resistance in pepper (*Capsicum annuum*) *via* genotyping-by-sequencing-based QTL mapping and genome-wide association study. *Sci. Rep.* 9, 1–15. doi: 10.1038/s41598-019-46342-1

Silva-Navas, J., Moreno-Risueno, M. A., Manzano, C., Téllez-Robledo, B., Navarro-Neila, S., Carrasco, V., et al. (2016). Flavonols mediate root phototropism and growth through regulation of proliferation-to-differentiation transition. *Plant Cell* 28, 1372–1387. doi: 10.1105/tpc.15.00857

Singh, A. P., Luthria, D., Wilson, T., Vorsa, N., Singh, V., Banuelos, G. S., et al. (2009). Polyphenols content and antioxidant capacity of eggplant pulp. *Food Chem.* 114, 955–961. doi: 10.1016/j.foodchem.2008.10.048

Singh, A. P., Wang, Y., Olson, R. M., Luthria, D., Banuelos, G. S., Pasakdee, S., et al. (2017). LC-MS-MS analysis and the antioxidant activity of flavonoids from eggplant skins grown in organic and conventional environments. *Food Nutr. Sci.* 8, 873–888. doi: 10.4236/fns.2017.89063

Slimestad, R., Fossen, T., and Verheul, M. J. (2008). The flavonoids of tomatoes. *J. Agric. Food Chem.* 56, 2436–2441. doi: 10.1021/jf073434n

Slimestad, R., and Verheul, M. (2009). Review of flavonoids and other phenolics from fruits of different tomato (*Lycopersicon esculentum* mill.) cultivars. *J. Sci. Food Agric.* 89, 1255–1270. doi: 10.1002/jsfa.3605

Stommel, J. R., and Whitaker, B. D. (2003). Phenolic acid content and composition of eggplant fruit in a germplasm core subset. *J. Am. Soc Hortic. Sci.* 128, 704–710. doi: 10.21273/jashs.128.5.0704

Stommel, J. R., Whitaker, B. D., Haynes, K. G., and Prohens, J. (2015). Genotype × environment interactions in eggplant for fruit phenolic acid content. *Euphytica* 205, 823–836. doi: 10.1007/s10681-015-1415-2

Sulli, M., Barchi, L., Toppino, L., Diretto, G., Sala, T., Lanteri, S., et al. (2021). An eggplant recombinant inbred population allows the discovery of metabolic QTLs controlling fruit nutritional quality. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.638195

Sunseri, F., Polignano, G. B., Alba, V., Lotti, C., Bisignano, V., Mennella, G., et al. (2010). Genetic diversity and characterization of African eggplant germplasm collection. *Afr. J. Plant Sci.* 4, 231–241. doi: 10.5897/ajps.9000128

Sytar, O., Zivcak, M., Bruckova, K., Brestic, M., Hemmerich, I., Rauh, C., et al. (2018). Shift in accumulation of flavonoids and phenolic acids in lettuce attributable to changes in ultraviolet radiation and temperature. *Sci. Hortic.* 239, 193–204. doi: 10.1016/j.scienta.2018.05.020

Szymański, J., Bocobza, S., Panda, S., Sonawane, P., Cárdenas, P. D., Lashbrooke, J., et al. (2020). Analysis of wild tomato introgression lines elucidates the genetic basis of

transcriptome and metabolome variation underlying fruit traits and pathogen response. *Nat. Genet.* 52, 1111–1121. doi: 10.1038/s41588-020-0690-6

Tanksley, S. D., Bernatzky, R., Lapitan, N. L., and Prince, J. P. (1988). Conservation of gene repertoire but not gene order in pepper and tomato. *Proc. Natl. Acad. Sci. U. S. A.* 85, 6419–6423. doi: 10.1073/PNAS.85.17.6419

The Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119

Tieman, D., Zhu, G., Resende, M. F. R., Lin, T., Nguyen, C., Bies, D., et al. (2017). A chemical genetic roadmap to improved tomato flavor. *Science* 355 (6323), 391–394. doi: 10.1126/science.aal1556

Tohge, T., de Souza, L. P., and Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.* 68, 4013–4028. doi: 10.1093/jxb/erx177

Tohge, T., and Fernie, A. R. (2016). Specialized metabolites of the flavonol class mediate root phototropism and growth. *Mol. Plant* 9, 1554–1555. doi: 10.1016/j.molp.2016.10.019

Tohge, T., Scossa, F., Wendenburg, R., Frasse, P., Balbo, I., Watanabe, M., et al. (2020). Exploiting natural variation in tomato to define pathway structure and metabolic regulation of fruit polyphenolics in the lycopersicum complex. *Mol. Plant* 13, 1027–1046. doi: 10.1016/J.MOLP.2020.04.004

Toppino, L., Barchi, L., Lo Scalzo, R., Palazzolo, E., Francese, G., Fibiani, M., et al. (2016). Mapping quantitative trait loci affecting biochemical and morphological fruit properties in eggplant (*Solanum melongena* l.). *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00256

Valanciene, E., Jonuskiene, I., Syrpas, M., Augustiniene, E., Matulis, P., Simonavicius, A., et al. (2020). Advances and prospects of phenolic acids production, biorefinery and analysis. *Biomolecules* 10, 874. doi: 10.3390/BIOM10060874

Van Eck, J. (2018). Genome editing and plant transformation of solanaceous food crops. *Curr. Opin. Biotechnol.* 49, 35–41. doi: 10.1016/j.copbio.2017.07.012

van Tuinen, A., Vos, R., Hall, R. D., Linus, H. W., van der Plas, L. H. W., Bowler, C., et al. (2006). "Use of metabolomics for identification of tomato genotypes with enhanced nutritional value derived from natural light-hyperresponsive mutants," in *Plant genetic engineering. vol. 7. metabolic engineering and molecular farming*. Ed. K Jaiwal Pawan (Huston, Texas: Studium Press, LLC), 240–256.

Verhoeyen, M. E., Bovy, A., Collins, G., Muir, S., Robinson, S., De Vos, C. H. R., et al. (2002). Increasing antioxidant levels in tomatoes through modification of the flavonoid biosynthetic pathway. *J. Exp. Bot.* 53, 2099–2106. doi: 10.1093/jxb/erf026

Villanueva, G., Rosa-Martínez, E., Sahin, A., García-Fortea, E., Plazas, M., Prohens, J., et al. (2021). Evaluation of advanced backcrosses of eggplant with *Solanum elaeagnifolium* introgressions under low n conditions. *Agronomy* 11, 1770. doi: 10.3390/agronomy11091770

Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/mp/ssp106

Wahyuni, Y., Ballester, A. R., Sudarmonowati, E., Bino, R. J., and Bovy, A. G. (2011). Metabolite biodiversity in pepper (*Capsicum*) fruits of thirty-two diverse accessions: Variation in health-related compounds and implications for breeding. *Phytochemistry* 72, 1358–1370. doi: 10.1016/j.phytochem.2011.03.016

Wahyuni, Y., Stahl-Hermes, V., Ballester, A. R., de Vos, R. C. H., Voorrips, R. E., Maharijaya, A., et al. (2014). Genetic mapping of semi-polar metabolites in pepper fruits (*Capsicum* sp.): Towards unravelling the molecular regulation of flavonoid quantitative trait loci. *Mol. Breed.* 33, 503–518. doi: 10.1007/s11032-013-9967-0

Wang, S., Chu, Z., Jia, R., Dan, F., Shen, X., Li, Y., et al. (2018). *SlMYB12* regulates flavonol synthesis in three different cherry tomato varieties. *Sci. Rep.* 8, 1582. doi: 10.1038/s41598-018-19214-3

Wang, C., and Fu, D. (2018). Virus-induced gene silencing of the eggplant chalcone synthase gene during fruit ripening modifies epidermal cells and gravitropism. *J. Agric. Food Chem.* 66, 2623–2629. doi: 10.1021/acs.jafc.7b05617

Wang, L., Lam, P. Y., Lui, A. C. W., Zhu, F. Y., Chen, M. X., Liu, H., et al. (2020). Flavonoids are indispensable for complete male fertility in rice. *J. Exp. Bot.* 71, 4715–4728. doi: 10.1093/jxb/eraa204

Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H., and Bao, C. (2020). A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Hortic. Res.* 7, 153. doi: 10.1038/s41438-020-00391-0

Weng, C. J., and Yen, G. C. (2012). Chemopreventive effects of dietary phytochemicals against cancer invasion and metastasis: Phenolic acids, monophenol, polyphenol, and their derivatives. *Cancer Treat. Rev.* 38, 76–87. doi: 10.1016/j.ctrv.2011.03.001

Whitaker, B. D., and Stommel, J. R. (2003). Distribution of hydroxycinnamic acid conjugates in fruit of commercial eggplant (*Solanum melongena* l.) cultivars. *J. Agric. Food Chem.* 51, 3448–3454. doi: 10.1021/jf026250b

Widhalm, J. R., and Dudareva, N. (2015). A familiar ring to it: Biosynthesis of plant benzoic acids. *Mol. Plant* 8, 83–97. doi: 10.1016/j.molp.2014.12.001

Wildermuth, M. C. (2006). Variations on a theme: synthesis and modification of plant benzoic acids. *Curr. Opin. Plant Biol.* 9, 288–296. doi: 10.1016/j.pbi.2006.03.006

Williams, R. J., Spencer, J. P. E., and Rice-Evans, C. (2004). Flavonoids: antioxidants or signalling molecules? *Free Radic. Biol. Med.* 36, 838–849. doi: 10.1016/j.frirabiomed.2004.01.001

Willits, M. G., Kramer, C. M., Prata, R. T. N., De Luca, V., Potter, B. G., Stephens, J. C., et al. (2005). Utilization of the genetic resources of wild species to create a nontransgenic high flavonoid tomato. *J. Agric. Food Chem.* 53, 1231–1236. doi: 10.1021/jf049355i

Wu, F., Eannetta, N. T., Xu, Y., and Tanksley, S. D. (2009). A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. *Theor. Appl. Genet.* 118, 927–935. doi: 10.1007/s00122-008-0950-9

Wu, Y., Popovsky-Sarid, S., Tikunov, Y., Borovsky, Y., Baruch, K., Visser, R. G. F., et al. (2022). *CaMYB12-like* underlies a major QTL for flavonoid content in pepper (*Capsicum annuum*) fruit. *New Phytol.* 237, 2255–2267. doi: 10.1111/nph.18693

Wu, M., Xu, X., Hu, X., Liu, Y., Cao, H., Chan, H., et al. (2020). *SlMYB72* regulates the metabolism of chlorophylls, carotenoids, and flavonoids in tomato fruit. *Plant Physiol.* 183, 854–868. doi: 10.1104/pp.20.00156

Zhang, Y., Butelli, E., Alseekh, S., Tohge, T., Rallapalli, G., Luo, J., et al. (2015). Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* 6, 8635. doi: 10.1038/ncomms9635

Zhang, Y., Butelli, E., De Stefano, R., Schoonbeek, H. J., Magusin, A., Pagliarani, C., et al. (2013). Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. *Curr. Biol.* 23, 1094–1100. doi: 10.1016/J.CUB.2013.04.072

Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. *Cell* 172, 249–261. doi: 10.1016/j.cell.2017.12.019

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 April 2023
DOI 10.3389/fpls.2023.1159223

Check for updates

OPEN ACCESS

EDITED BY
Eleni Tani,
Agricultural University of Athens, Greece

REVIEWED BY
Dhananjay K. Pandey,
Amity University, Jharkhand, India
Erpei Lin,
Zhejiang Agriculture and Forestry University, China

*CORRESPONDENCE
Yehua He
heyehua@hotmail.com
Chengjie Chen
ccj0410@gmail.com

†These authors have contributed equally to this work

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

RECEIVED 05 February 2023
ACCEPTED 17 March 2023
PUBLISHED 14 April 2023

CITATION
Yi W, Luan A, Liu C, Wu J, Zhang W, Zhong Z, Wang Z, Yang M, Chen C and He Y (2023) Genome-wide identification, phylogeny, and expression analysis of GRF transcription factors in pineapple (*Ananas comosus*). *Front. Plant Sci.* 14:1159223. doi: 10.3389/fpls.2023.1159223

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

# Genome-wide identification, phylogeny, and expression analysis of GRF transcription factors in pineapple (*Ananas comosus*)

Wen Yi[1†], Aiping Luan[2†], Chaoyang Liu[1], Jing Wu[1], Wei Zhang[1], Ziqin Zhong[1], Zhengpeng Wang[1], Mingzhe Yang[1], Chengjie Chen[1*] and Yehua He[1*]

[1]Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in South China, Ministry of Agriculture and Rural Areas, College of Horticulture, South China Agricultural University, Guangzhou, China, [2]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

**Background:** Pineapple is the only commercially grown fruit crop in the Bromeliaceae family and has significant agricultural, industrial, economic, and ornamental value. GRF (growth-regulating factor) proteins are important transcription factors that have evolved in seed plants (embryophytes). They contain two conserved domains, QLQ (Gln, Leu, Gln) and WRC (Trp, Arg, Cys), and regulate multiple aspects of plant growth and stress response, including floral organ development, leaf growth, and hormone responses. The GRF family has been characterized in a number of plant species, but little is known about this family in pineapple and other bromeliads.

**Main discoveries:** We identified eight GRF transcription factor genes in pineapple, and phylogenetic analysis placed them into five subfamilies (I, III, IV, V, VI). Segmental duplication appeared to be the major contributor to expansion of the *AcGRF* family, and the family has undergone strong purifying selection during evolution. Relative to that of other gene families, the gene structure of the GRF family showed less conservation. Analysis of promoter *cis*-elements suggested that *AcGRF* genes are widely involved in plant growth and development. Transcriptome data and qRT-PCR results showed that, with the exception of *AcGRF5*, the *AcGRF*s were preferentially expressed in the early stage of floral organ development and *AcGRF2* was strongly expressed in ovules. Gibberellin treatment significantly induced *AcGRF7/8* expression, suggesting that these two genes may be involved in the molecular regulatory pathway by which gibberellin promotes pineapple fruit expansion.

**Conclusion:** AcGRF proteins appear to play a role in the regulation of floral organ development and the response to gibberellin. The information reported here provides a foundation for further study of the functions of *AcGRF* genes and the traits they regulate.

KEYWORDS

pineapple, GRF gene family, phylogeny analysis, flower organ, expression analysis

# 1 Introduction

Pineapple is one of the three most widely cultivated tropical fruit crops in the world and the only commercially grown member of the Bromeliaceae family (Lobo and Paull, 2017). Because of its excellent flavor and texture, it has been called the queen of fruits (Baruwa, 2013). Its unique shape, fiber content, and nutritional value give it an important place in medicine and industry (Sanewski et al., 2018). However, adverse environmental factors such as temperature extremes, as well as changes in hormone levels, can seriously affect pineapple growth and development, reducing yield and quality. Exploring the mechanisms that regulate flower and fruit development and stress responses is important for maintaining the commercial value of pineapple.

Transcription factors regulate the ability of cells to express different genes and thereby control development (Riechmann et al., 2000; Levine and Tjian, 2003). Growth-regulating factor (GRF) proteins are widespread transcription factors in plants, and their highly conserved N-terminal QLQ and WRC domains are their most prominent feature (Van der Knaap et al., 2000). These two domains have different functions; the QLQ domain is involved in protein interaction and is rich in aromatic/hydrophobic amino acids (Kim and Kende, 2004), whereas the WRC domain mediates DNA-binding ability (Osnato et al., 2010; Kim et al., 2012; Kuijt et al., 2014).

Since the first GRF protein was discovered in rice (*Oryza sativa*) (Van der Knaap et al., 2000; Omidbakhshfard et al., 2015), GRFs have been shown to influence almost all plant growth and developmental processes, including leaf growth (Kim and Kende, 2004; Kim and Lee, 2006; Arvidsson et al., 2011; Debernardi et al., 2014; Wu et al., 2014; Zhang et al., 2021), floral organ development (Arvidsson et al., 2011; Pu et al., 2012; Liang et al., 2013; Liu H. et al., 2014; Pajoro et al., 2014; Lee et al., 2017), root development (Bao et al., 2014), seed oil content (Bao et al., 2014), plant lifespan, and stress responses (Kim et al., 2012; Casadevall et al., 2013; Liu J. et al., 2014). Understanding the roles of GRFs is therefore important for plant growth research and genetic improvement.

Recent studies have shown that *AtGRF*s from the model plant *Arabidopsis thaliana* can significantly improve the transgenic transformation efficiency of a variety of crops, in addition to performing some of the functions above. For example, overexpression of *AtGRF5* and its orthologs increases transformation efficiency, callus cell proliferation, and transgenic bud formation in *Beta vulgaris*, corn (*Zea mays*), soybean (*Glycine max*), and *Helianthus annuus* (Debernardi et al., 2020; Kong et al., 2020). Expression of a GRF4–GIF1 fusion protein significantly increased the regeneration efficiency, regeneration rate, and somatic embryogenesis of wheat (*Triticum aestivum*) and rice (*Oryza sativa*) (Debernardi et al., 2020). Although there are multiple literature reports on the functions of GRF family members, detailed genome-wide phylogenetic and functional studies of GRF genes are not yet available for pineapple or other bromeliads.

To better understand the evolutionary dynamics of GRF genes in pineapple and explore their potential regulatory roles in flower and fruit development and hormone and stress responses, we identified eight AcGRF genes in pineapple and performed a series of analyses, documenting their chromosome locations, motif compositions, evolutionary relationships, genomic collinearity, and selection pressure. We also investigated the potential roles of these *AcGRF*s using interaction network prediction and gene expression analysis. These results provide insights into the functions of GRF family members in pineapple growth and development.

# 2 Materials and methods

## 2.1 Data sources and sequence retrieval

All protein sequences of pineapple were obtained from the Pineapple Genome Project (Ming et al., 2015). Sequences of nine *A. thaliana* GRF genes and 12 rice GRF genes were obtained from previous studies (Kim et al., 2003; Choi et al., 2004). The corresponding protein sequences were downloaded from the *Arabidopsis* Information Resource Library (TAIR) (http://www.arabidopsis.org/) and the UniProt Database (https://www.uniprot.org) (Boutet et al., 2016; Cheng et al., 2017). Protein sequences from *Phalaenopsis equestris* and *Nymphaea colorata* were obtained from recent studies (Zhang et al., 2017; Zhang and Chen, 2019), and the *Amborella trichopoda* proteins were downloaded from the PLAZA database (https://bioinformatics.psb.ugent.be/plaza/) (Van Bel et al., 2017). Protein sequences for *Vitis vinifera*, *Sorghum bicolor*, *Musa acuminata*, and other species were obtained from Phytozome (http://www.phytozome.net/).

## 2.2 Identification and classification of *GRFs*

We used two strategies to identify *GRF* genes in pineapple. First, we performed a local BlastP search of pineapple protein sequences using GRF protein sequences from *A. thaliana* as queries. We then obtained hidden Markov models of the QLQ (PF08880) and WRC (PF08879) domains from the Pfam database (http://pfam.xfam.org/) and used them to query the pineapple protein files using TBtools with an E value of 1e−10 (Chen C. et al., 2020). We identified 21 protein sequences as candidate GRFs in pineapple. We confirmed the presence of GRF core sequences using the Batch CD-Search and SMART programs, further examining all candidate genes that appeared to contain QLQ and WRC domains in the BlastP and HMMER search results. Each candidate gene was then manually checked, and the structurally annotated gene was corrected to ensure that there were conserved heptapeptide sequences at the N terminus of the predicted QLQ and WRC domains. Thirteen protein sequences with no or incomplete QLQ or WRC domains were removed. We identified GRF genes in 23 additional species using the same method, and the complete set of GRF protein sequences was used to study their evolutionary relationships. The ExPASy ProtParam database (https://web.expasy.org/protparam/) was used to analyze the physicochemical properties of the AcGRF genes/proteins, including coding region length, number of amino acids, molecular weight (MW), and theoretical isoelectric point (pI).

The subcellular localizations of the AcGRFs were predicted using Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/#), the transmembrane domains using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/), and the signal peptides using SignalP (http://www.cbs.dtu.dk/services/SignalP/).

We next characterized the evolutionary relationships among *GRF* genes from various species and assigned the putative pineapple *GRF* genes to specific subfamilies. Multiple-sequence alignments of amino acid sequences were constructed using MAFFT with default parameters, and a species-tree of 26 taxons was constructed using their protein sequences with the maximum likelihood method and 1,000 bootstrap replicates in OrthoFinder with the following parameters: orthofinder -f dataset -M msa -S diamond -t fasttree -t 16 -a 16.

## 2.3 Chromosome locations, gene structures, and conserved motifs of the AcGRFs

We obtained chromosome locations and genetic structures of each *AcGRF* gene from the pineapple genome annotation file (Ming et al., 2015). The data were then integrated and plotted using TBtools (Chen C. et al., 2020). We identified conserved motifs shared among the AcGRF proteins (Bailey et al., 2009) using MEME tools with the following parameters: maximum number of cardinal orders, 10; minimum width, 20; maximum width, 50.

## 2.4 Duplication, collinearity, and evolutionary analysis of the *AcGRF* gene family

We performed collinearity analysis using the method described in TBtools (Chen et al., 2022). We first prepared the pineapple genome and annotation file, then used it as input for collinearity analysis using the "One Step MCScanX Wrapper" function with the following parameters: CPUs for BlastP, 8; e-value, 1e−3; number of blast hits, 10. Dispersed, proximal, tandem, and segmental/WGD duplicates in the *AcGRF* family were identified using TBtools. We connected segmentally duplicated gene pairs by red and green arcs in the Circos plot. We also performed collinearity analysis of GRF genes in pineapple and other plant species, including *A. thaliana*, *V. vinifera*, *A. trichopoda*, *N. colorata*, *O. sativa*, *S. bicolor*, *M. acuminata*, *Phalaenopsis equestris*, and *Spirodela polyrhiza*. TBtools was used to visualize a portion of the results.

## 2.5 Identification of *cis*-elements in the *AcGRF* promoters

The 2,000-bp sequence upstream of each *AcGRF* gene was extracted using TBtools and defined as the promoter region. PlantCARE software (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was then used to predict the *cis*-elements in the promoter region (Lescot, 2002), and the results were visualized using TBtools.

## 2.6 Expression profiles of *AcGRF* genes in different pineapple tissues

Earlier transcriptome studies have generated data on *AcGRF* gene expression in floral organs at different developmental stages (Wang et al., 2020) and in various fruit tissues during development (Mao et al., 2018). The transcriptome data from both projects were downloaded from NCBI using the project IDs PRJEB38680 and PRJNA483249. The two datasets were analyzed separately using FPKM values. TBtools was used to transform the FPKM values to $\log_2$(FPKM+1) values and generate an expression heatmap for the relevant genes.

## 2.7 Hormone and stress treatments, RNA extraction, and RT-PCR

Well-grown "Shenwan" calli were sampled after exposure to 0.1 mM jasmonic acid (JA), 0.1 mM abscisic acid (ABA), 0.1 mM auxin (IAA), 0.1 mM gibberellin (GA), or 150 mM NaCl in suspension culture medium for 0, 4, 8, 16, 24, 36, and 48 h. All materials were immediately frozen in liquid nitrogen and stored at −80°C for subsequent RNA extraction.

The CWBIO RNApure Plant Kit (DNase I) was used to extract total RNA. RNA quality was checked by agarose gel electrophoresis, and RNA concentration was estimated using a Nanodrop ND-1000 spectrophotometer. First-strand cDNA was synthesized from DNA-free RNA using the HiScript II First-Strand cDNA Synthesis Kit (YEASEN) according to the manufacturer's protocol. qRT-PCR was performed on a Roche LightCycler 480 instrument with SYBR Green Master Mix (YEASEN). The total reaction volume was 10 μL and contained 5 μL SYBR mix, 0.4 μL upstream and downstream primer mix (2.5 μM), 0.5 μL cDNA template, and 4.1 μL $ddH_2O$. Each reaction was performed in triplicate. The reaction conditions were 95°C for 30 s, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. Relative gene expression was calculated by the $2^{-\Delta\Delta Ct}$ method with β-actin as the internal reference gene. Primer sequences used in this study are provided in Table S10.

## 2.8 Protein–protein interaction network prediction

All AcGRF protein sequences were submitted to the STRING website (http://string-db.org), and their rice orthologs were selected as a reference. After completing the BLAST step, the network was built using the proteins with the highest scores (bitscores). Proteins that were not predicted to interact with any other proteins were removed.

## 2.9 miRNA target prediction for *AcGRFs*

Pineapple miRNAs were obtained from previous studies (Yusuf et al., 2015), and *AcGRF* genes targeted by miRNAs were predicted using psRNATarget (https://www.zhaolab.org/psRNATarget/) with default parameters while selecting target accessibility as described in Dai et al. (2018). We visualized the interactions among the predicted miRNAs and the corresponding target *AcGRF* genes using TBtools v1.106.

# 3 Results and analysis

## 3.1 Physicochemical properties of AcGRF transcription factors

We identified eight *GRF* gene family members with intact QLQ and WRC domains in pineapple and named them *AcGRF1* to *AcGRF8* based on their chromosome locations (Table 1). The eight *GRFs* encoded proteins with 249 (AcGRF4) to 612 (AcGRF6) amino acids, molecular weights (MWs) from 25.1 kDa (AcGRF4) to 65.7 kDa (AcGRF7), and isoelectric points (pIs) from 6.57 (AcGRF5) to 9.63 (AcGRF4). All AcGRF proteins had a negative hydrophilicity score (GRAVY), indicating that they are hydrophilic proteins. Subcellular localization predictions indicated that seven AcGRF proteins were localized to the nucleus and AcGRF4 was localized to the cell membrane.

## 3.2 Characterization of AcGRF sequences

All the pineapple GRFs contained a highly conserved WRC protein domain with the RTDGKKWRC motif. Seven contained the canonical QLQ (Gln-Leu-Gln) motif; the exception was AcGRF5, in which Leu was replaced by Met (Gln-Met-Gln) (Figure 1A). A similar result was reported in *A. thaliana*, in which the Leu in AtGRF9 is replaced by Phe (Kim et al., 2003).

The scattered distribution of *AcGRF* gene members on different chromosomes is shown in Figure 1B. There are two *GRF* genes on

TABLE 1 Characteristics of AcGRFs.

| Gene name | Gene ID | Chr location | CDS (bp) | Exon no. | Protein length (aa) | MW (kDa) | pI | GRAVY score | Subcellular localization |
|---|---|---|---|---|---|---|---|---|---|
| AcGRF1 | Aco009479 | LG01:1916306-1917558 | 903 | 2 | 300 | 33.4 | 8.18 | −0.519 | Nucleus |
| AcGRF2 | Aco023268 | LG02:11260633-11262425 | 1,167 | 4 | 388 | 43.3 | 7.74 | −0.887 | Nucleus |
| AcGRF3 | Aco020046 | LG03:9349100-9351207 | 927 | 3 | 308 | 34.4 | 9.00 | −0.530 | Nucleus |
| AcGRF4 | Aco015543 | LG03:12472433-12473582 | 750 | 3 | 249 | 25.1 | 9.63 | −0.077 | Cell membrane |
| AcGRF5 | Aco015755 | LG09:10903631-10910491 | 831 | 4 | 276 | 30.8 | 6.57 | −0.852 | Nucleus |
| AcGRF6 | Aco000277 | LG12:3038602-3042504 | 1,839 | 4 | 612 | 64.7 | 8.03 | −0.502 | Nucleus |
| AcGRF7 | Aco013343 | LG15:11084401-11088716 | 1,821 | 4 | 606 | 65.7 | 9.16 | −0.711 | Nucleus |
| AcGRF8 | Aco013172 | LG24:222856-228237 | 1,200 | 5 | 399 | 43.9 | 8.80 | −0.539 | Nucleus |

10.3389/fpls.2023.1159223

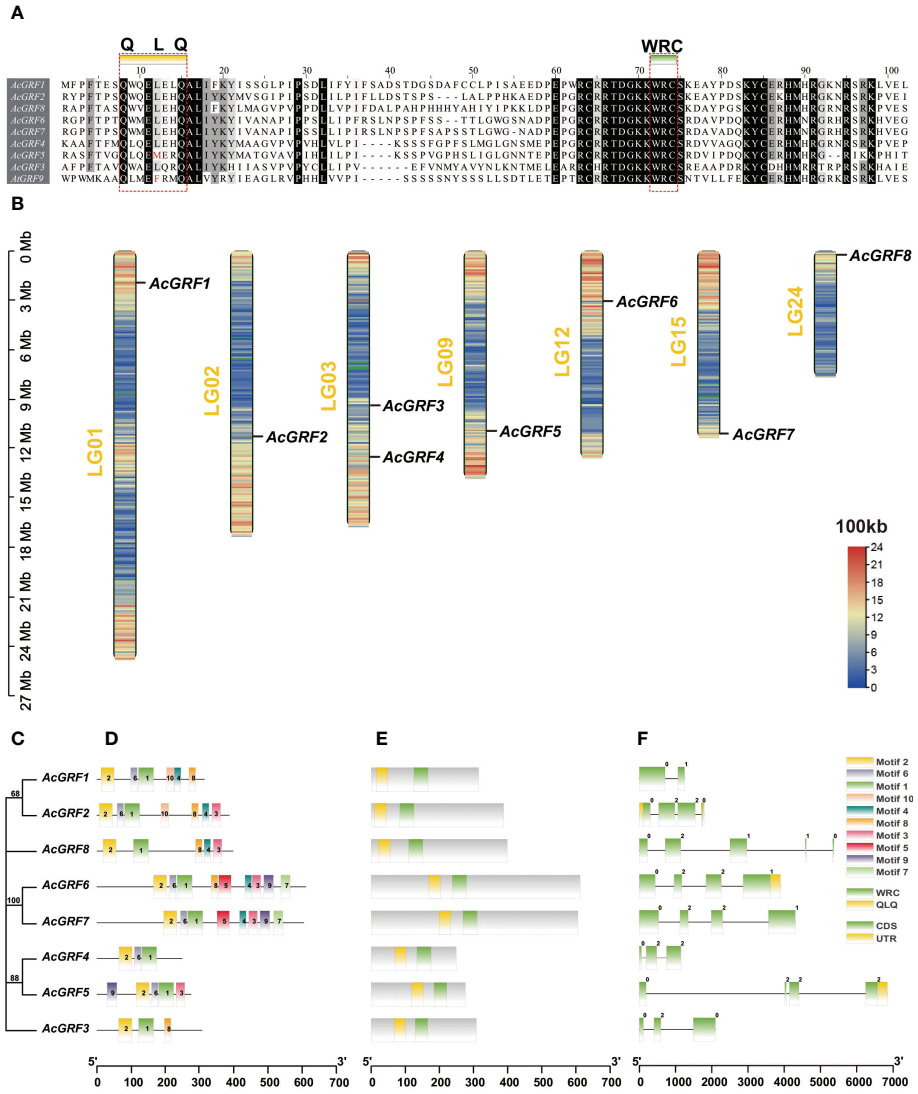



FIGURE 1
Conserved sequences, chromosomal distribution, gene structures, and motif patterns of *AcGRF* genes and their encoded proteins. **(A)** Conserved sequences in AcGRF genes. The QLQ and WRC domains are marked with a red rectangular box, and the amino acids replaced in the QLQ domain are indicated in a red font. **(B)** Chromosomal distribution of *AcGRF* genes. The sliding window size was set to 100 kb, and the color from red to blue indicates high to low gene density. Blank areas on chromosomes are genetic regions for which information on gene distribution is lacking. **(C)** Phylogenetic clustering of AcGRF proteins. **(D)** Motif patterns of AcGRF proteins. **(E)** Conserved domains in AcGRF proteins. **(F)** *AcGRF* gene structures. Yellow boxes indicate 5′ and 3′ UTR regions; green boxes indicate exons; black lines represent introns; and the number (0–2) indicates the intron phase.

chromosome 3, but they are physically distant. Most *AcGRF* genes tend to cluster in regions with higher gene density.

A conserved motif search was performed on the eight AcGRF proteins, and the number of conserved motifs in individual proteins ranged from 3 to 9. Motifs 1 and 2 were components of the WRC and QLQ domains, respectively, and were present in all GRF proteins, whereas other motifs were only present in certain members. AcGRF proteins in the same subgroup had similar motif compositions (Figures 1C–E, S1). Exon numbers ranged from 2 to 5, and most *GRF* genes did not have 5′ or 3′ UTR annotation information (Figure 1F). Four was the most common number of exons (4 *AcGRF* genes), followed by three (2 genes).

## 3.3 Phylogenetic relationships of the AcGRFs

To reveal the evolutionary relationships among GRF genes from various species, we constructed a phylogenetic tree of *GRFs* from 26 species (Figure 2A). GRF genes were identified from 23 additional plant species and two algal species using the same method applied to pineapples (Figures 2B, C). Most species had around 10 GRF genes; exceptions included *Marchantia polymorpha* and *Physcomitrella patens* (1), *Musa acuminata* (19), and *Gossypium raimondii* (41) (Figure 2B). We constructed phylogenetic trees of the GRF protein sequences from these species using the maximum

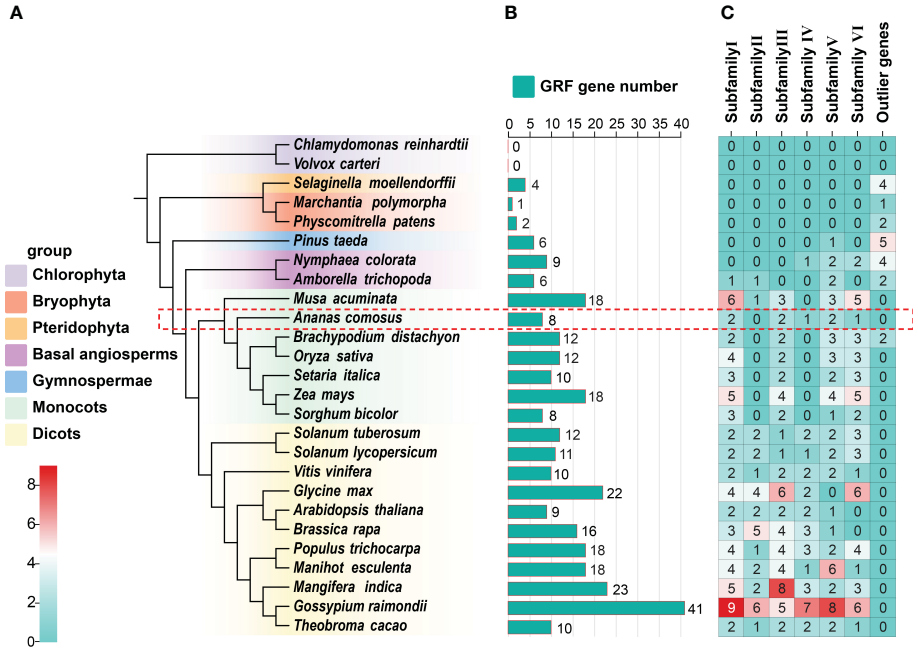



FIGURE 2
Evolutionary relationships among 26 species and their GRF gene compositions. **(A)** Phylogenetic relationships among 26 species. **(B)** Numbers of GRF genes in different species. **(C)** Distribution of GRF genes in various subfamilies. Phylogenetic analyses of 26 species were performed using OrthoFinder (Emms and Kelly, 2019). The *GRF* genes in pineapple are marked with a red rectangular box.

likelihood method, dividing them into specific subfamilies (Figures 2C, S2; Table S1). Pineapples contained members of all subfamilies except subfamily II: *AcGRF6/7* in subfamily I, *AcGRF1/2* in subfamily III, *AcGRF4/5* in subfamily V, and *AcGRF3* and *AcGRF8* in subfamilies IV and VI, respectively (Figure S2). Across all species, subfamily I had the most members, followed by subfamilies III, VI, and V (Figure 2C), which accounted for 21.38%, 18.42%, 17.10%, and 16.78% of all GRF genes in the 26 species, respectively. Interestingly, GRF genes from bryophytes and ferns were not grouped into any of the six subfamilies but were present in a separate outer group. *GRFs* from basal angiosperms (i.e., *N. colorata* and *A. trichopoda*) were found in three of the six subfamilies, and some were in the outer group. Banana (no subfamily IV) and pineapple (no subfamily II) contained members of five *GRF* subfamilies, and other monocots contained members of only four subfamilies (no subfamilies II or IV). All dicots contained members of all six subfamilies, with the exception of soybean, which lacked members of subfamily V.

## 3.4 Origins of AcGRF gene members

Gene duplication is thought to be the main driver of species evolution and a direct cause of gene family expansion (Lynch and Conery, 2000; Moore and Purugganan, 2003; Maere et al., 2005). Pineapple experienced two whole-genome duplication (WGD) events in its early evolution, corresponding to two peaks in Figure S3. We searched for tandem and segmental duplicates among the *AcGRFs* (Tables S2, S3) and visualized them using the Circos plot (Figure 3A). Expansion of the AcGRF gene family appeared to have occurred mainly through segmental duplication, and no tandem duplication events were found. There was one segmentally duplicated gene pair (Figure 3A); the AcGRF1–AcGRF2 gene pair was associated with the WGD events of pineapple.

To understand the ancestral relationships among AcGRF genes and GRF genes from other species, we examined collinear relationships among genes from pineapple and two core eudicots (*A. thaliana*, *V. vinifera*), two basal angiosperms (*A. trichopoda*, *N. colorata*), and five core monocots (*O. sativa*, *S. bicolor*, *P. equestris*, *M. acuminata*, and *S. polyrhiza*). Four GRF genes from *A. thaliana* had collinear relationships with AcGRF genes from pineapple; the numbers of collinear genes in other species were 7 in *V. vinifera*, 14 in *M. acuminata*, 9 in *O. sativa*, 6 in *S. bicolor*, 9 in *S. polyrhiza*, 2 in *P. equestris*, and 3 in *N. colorata* (Table S4; Figures 3B, C). *AcGRF2* had collinear relationships with gene(s) in all species except *N. colorata*; the collinear block in which *AcGRF1* is located is present in species other than rice (*O. sativa*) and *S. bicolor*, and the collinear block in which *AcGRF4* is located is present in species except *A. thaliana* and *P. equestris*. Interestingly, *AcGRF5*-related collinear blocks were found only in *O. sativa*, and *AcGRF3*-related blocks were found only in water lilies (*N. colorata*). We next calculated the Ka/Ks (non-synonymous/synonymous substitution ratio) values of *AcGRF* genes and their orthologs in nine other species (Tables S4, S5; Figure 3D). All AcGRF orthologous gene pairs had Ka/Ks values

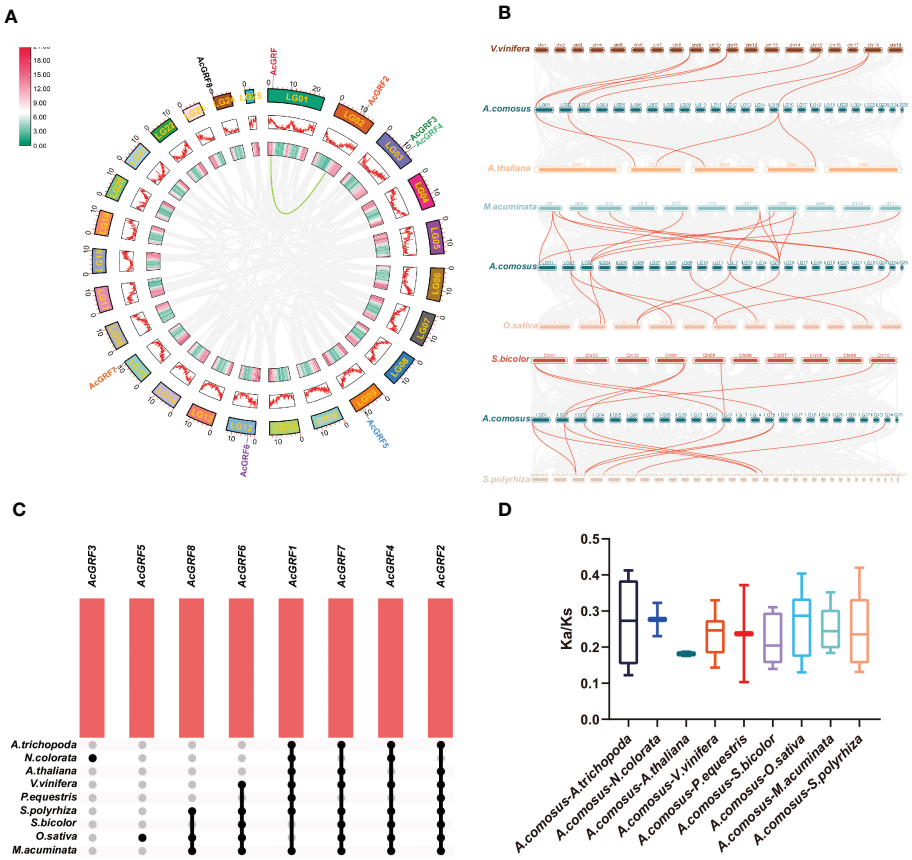



FIGURE 3
Collinearity and Ka/Ks analyses of pineapple *AcGRFs* and their homologs in other species. **(A)** Circos plot of pineapple collinear homologous genes. Genome-wide collinear blocks are set against a gray background, and duplicate AcGRF gene pairs are highlighted by red and green curves. Each pineapple chromosome is accompanied by 100 kb of gene density information, depicted by heat maps and waveform maps. **(B)** Collinearity of GRF genes in pineapple and six representative species. Same-line blocks are set to a gray background, and collinear GRF genes are highlighted with red curves. **(C)** Venn diagram of non-redundant collinear GRF genes in pineapple and other species. **(D)** Box plot of Ka/Ks ratios of *GRF* orthologs.

less than 1, indicating that the AcGRF gene family has experienced strong purifying selection during evolution.

## 3.5 Promoter analysis and expression of AcGRF genes

Gene promoters interact with DNA or regulatory proteins to control gene expression (Biłas et al., 2016). We extracted promoter sequences from the pineapple GRF genes (2,000 bp upstream of the start codon) and identified their *cis*-acting elements. A total of 156 *cis*-elements were detected in the promoter regions of the AcGRF genes and were divided into three classes (Figures 4A, B and Table S6). Hormone response elements included those responsive to IAA (4; 2.56%), GA (4; 2.56%), ABA (14; 8.97%), JA (10; 6.41%), and salicylic acid (5; 3.21%). With the exception of subfamily III members (*AcGRF1/2*), the promoter regions of all other subfamily members contained JA and ABA response elements. Growth- and development-related elements included those related to meristem expression (3; 1.92%; present in *AcGRF1* and *AcGRF8*), endosperm expression (4; 2.56%; present in *AcGRF1/2/7*), light response (72; 46.15%; present in all AcGRF genes), and circadian rhythm control (1; 0.64%; present in *AcGRF4*). Stress response elements included those related to drought induction (5; 3.21%), low temperature response (2; 1.28%), and anaerobic induction (25; 16.03%). Anaerobic induction elements were present in all genes except *AcGRF5*, and other response elements were unevenly distributed among the *AcGRFs*.

Studies have reported that GRF genes have a positive regulatory effect on callus proliferation (Kong et al., 2020), and GRF genes regulate plant growth and development by regulating GA, IAA, and ABA (Kim et al., 2012; Chen Y. et al., 2020; Huang et al., 2022). To investigate whether the expression of AcGRF genes is affected by abiotic stress and hormone treatments, we measured the expression of seven *AcGRF* family members by qRT-PCR after exposure to different hormones or 150 mM NaCl (Figure 4C). *AcGRF1* expression was induced by all tested treatments. The expression of *AcGRF1* and *AcGRF5* gradually increased with ABA exposure time, whereas that of *AcGRF6* showed the opposite trend. GA significantly induced *AcGRF1*, *AcGRF7*, and *AcGRF8* expression and inhibited that of *AcGRF2* and *AcGRF6*. JA significantly induced the expression of *AcGRF6*. IAA significantly induced the expression of *AcGRF8* but inhibited that of *AcGRF6*. Most *AcGRF* genes increased in expression with increasing duration of NaCl

10.3389/fpls.2023.1159223

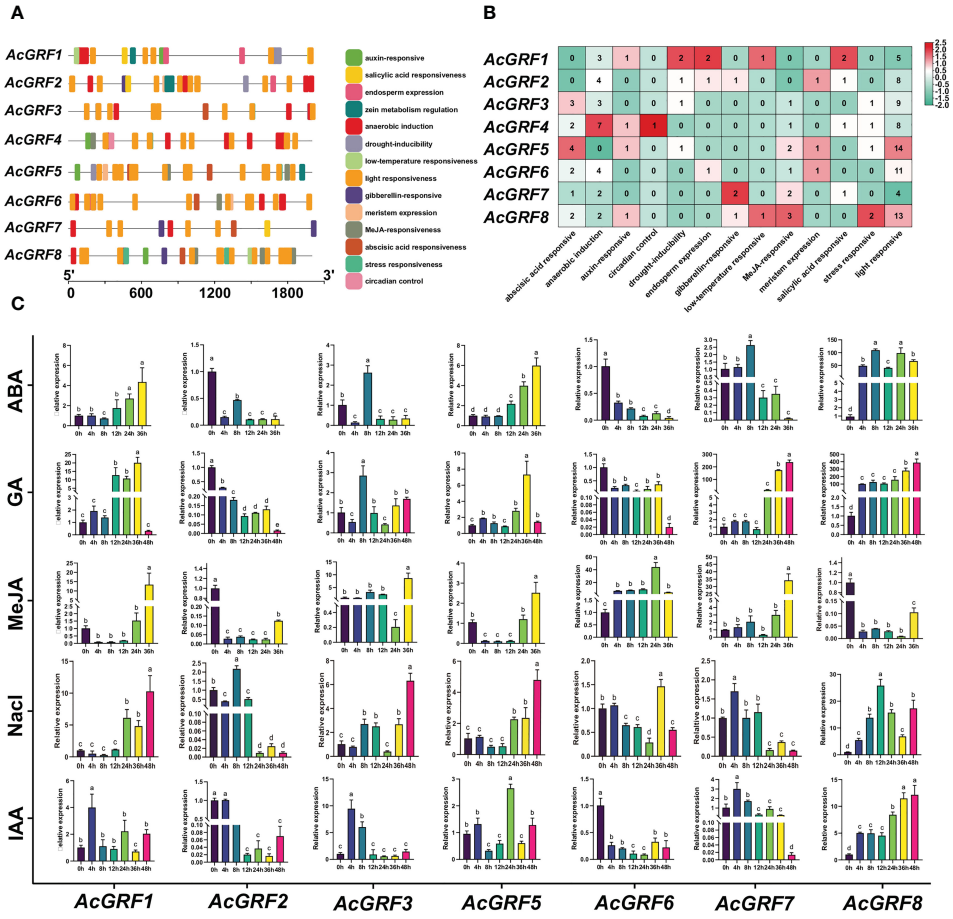



FIGURE 4
*cis*-Elements in AcGRF gene promoters and AcGRF expression profiles under hormone and salt stress treatments. **(A)** *cis*-Acting elements in the 2,000-bp region upstream of the AcGRF genes. **(B)** Distribution of cis-Acting elements of the AcGRF genes. **(C)** AcGRF gene expression profiles under different hormone and salt stress treatments. Data were normalized to the expression of β-actin and expressed as mean ± standard deviation. Different lowercase letters indicate significant differences between treatment times ($P < 0.01$, F test).

exposure. *AcGRF8* had the highest expression levels during ABA and GA treatment, with relative expression levels of 110 and 387, respectively, and *AcGRF7* had a relative expression level of 237 after GA treatment. These findings suggest potential roles for AcGRF genes under different hormone treatments and stress conditions.

We next analyzed the expression profiles of GRF genes at different developmental stages of pineapple tissues using our previously reported transcriptome data (Mao et al., 2018) (Figure 5 and Table S7). The analysis included roots, stems, leaves, petals, stamens, pistils, discs, peduncles, ovules, ovaries, fruit hearts, bracts, sepals, and placenta (Figure 5A). In general, GRF genes from the same clade tended to exhibit similar expression patterns. Subfamily I members *AcGRF6* and *AcGRF7* showed a similar expression in nearly all tissues tested. Interestingly, almost all AcGRF genes were relatively highly expressed in the early stages of floral organ development (Figure 5A), with the exception of *AcGRF5*, which was relatively highly expressed in the late developmental stages of most tissues.

To determine whether individual AcGRF genes are associated with specific developmental stages of pineapple floral organs, we analyzed transcriptome data from five floral organs (sepals, petals, stamens, pistils, and ovules) at different developmental stages (Wang et al., 2020). Almost all AcGRF genes showed the highest expression in ovules at all stages of development, followed by stamens; the only exception was *AcGRF3*, which was specifically expressed in pistils (Figure 5B and Table S8).

To better visualize the different expression patterns of AcGRF genes in various tissues, we created cartoon heatmaps of relevant gene expression from all pineapple *GRFs* (Figure 5C). Subfamily III member *AcGRF1* was highly expressed in petals, stamens, and pistils. *AcGRF2* was highly expressed in petals and ovules, and subfamily VI member *AcGRF8* was expressed at higher levels in pistils and ovules. Subfamily I members *AcGRF6* and *AcGRF7* were highly expressed in petals, pistils, and ovules. *AcGRF4* from subfamily V was highly expressed in pistils, and *AcGRF5* was highly expressed in bracts, sepals, and ovules. Subfamily IV member *AcGRF3* showed relatively high expression in petals. The expression patterns of AcGRF genes thus differ among pineapple tissues, suggesting their potential regulatory roles in pineapple development.

10.3389/fpls.2023.1159223

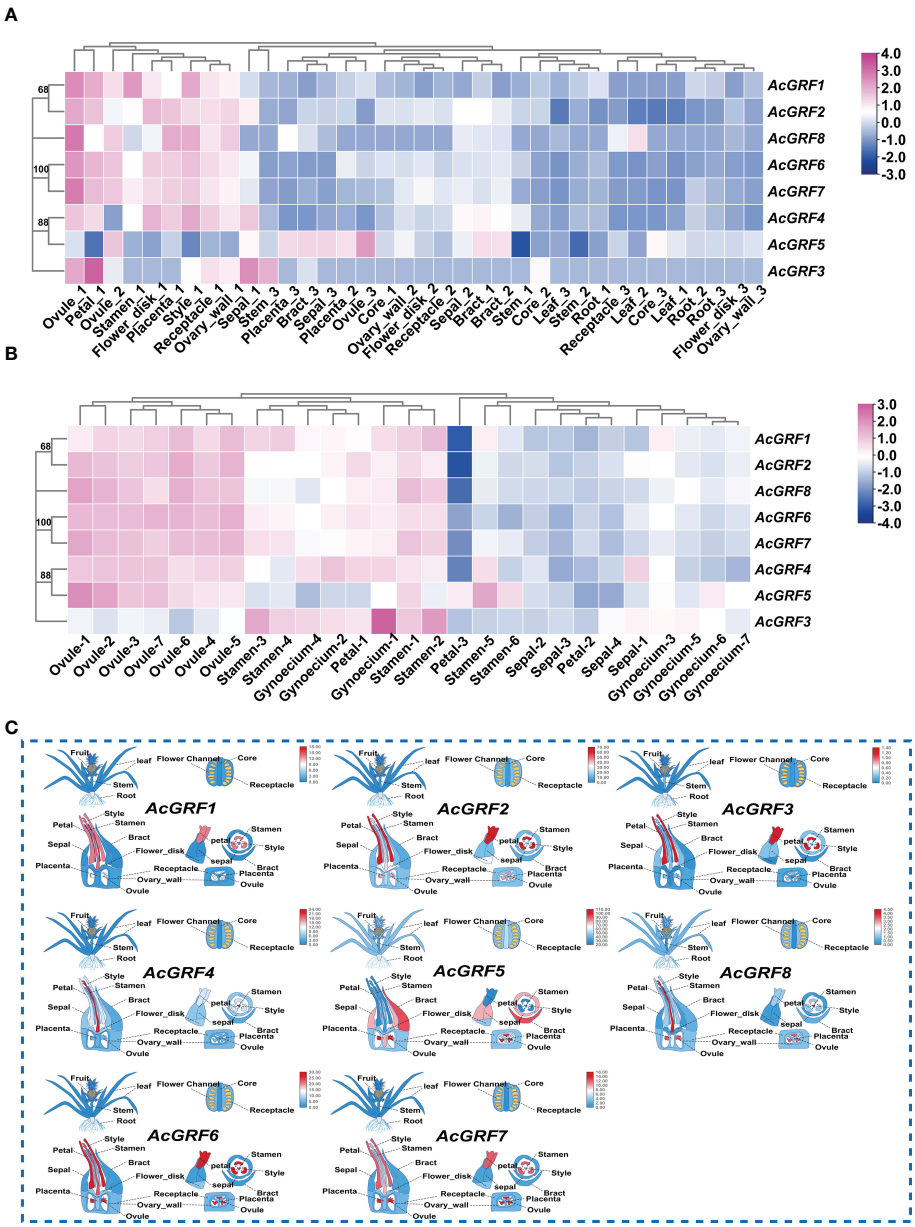



FIGURE 5
Expression of AcGRF genes in different tissues of pineapple. **(A)** Clustered heat map showing the expression patterns of AcGRF genes in various tissues at different developmental stages. **(B)** Expression heat map of AcGRF genes in five floral organs (stamens, pistils, ovules, sepals, and petals) of pineapple at different developmental stages: four stages of sepals (S1–S4), three of petals (S1–S3), five of stamens (S1–S5), and seven of pistils (S1–S7) and ovules (S1–S7). Relative gene expression is quantified as $\log_2$(FPKM+1). Blue, white, and red indicate low, medium, and high expression levels. **(C)** Cartoon heatmap showing the expression patterns of AcGRF genes from different clades in different tissues.

## 3.6 Predicted protein interaction network of AcGRFs

A predicted protein interaction network indicated that AcGRFs had multiple interaction partners (Figure 6A), including growth regulator interaction factor 1 (GIF1), the TCP (teosinte branched1/cincinnata/proliferating cell factor) transcription factor PCF5, and the YABBY transcription factor YAB2. With the exception of AcGRF1/3/4, all other members interacted with one or more proteins. AcGRF2/5/6/7/8 were predicted to interact with 6, 2, 5,

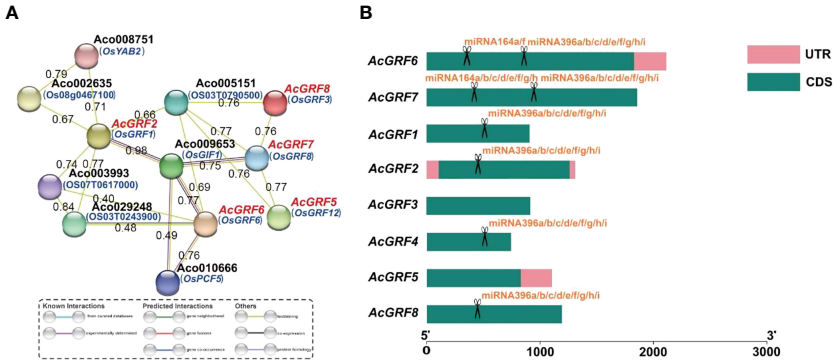



FIGURE 6
Predicted protein interaction network of AcGRF proteins and miRNA target sites in *AcGRF* genes. **(A)** Protein interaction network predicted using *AcGRF* orthologs from rice. **(B)** Predicted miRNA targets in the *AcGRF* genes. The scissors represent the miRNA and its targeted *AcGRF* gene, with penalty score ≤5 and lower expectations indicating higher prediction accuracy.

4, and 2 proteins, respectively. There were also predicted interactions between the AcGRF proteins themselves.

We next performed miRNA target site prediction for the *AcGRF* genes. As shown in Figure 6B, six *AcGRF* genes were predicted to be targeted by two miRNAs (miRNA164/miRNA396). *AcGRF6/7* contained targets for both miRNAs, whereas *AcGRF1/2/4/8* contained targets for miRNA396 only (Table S9). Interestingly, all miRNA target sites were located in the CDS regions.

# 4 Discussion

GRF proteins regulate important processes in plant biology, including leaf growth, floral organ development, root development, seed oil content, plant lifespan, and stress response (Omidbakhshfard et al., 2015). Pineapple is the third most important tropical fruit crop in the world after bananas and citrus, and various biotic and abiotic stresses have a significant impact on its yield and quality. Although GRF genes have been studied in other plant species, they have not yet been characterized in the bromeliad family. We therefore identified the GRF genes in pineapple, analyzed their evolutionary relationships, compared their sequence features, and analyzed their expression patterns in response to hormone and salt stress treatments.

## 4.1 Identification, classification, and characteristics of the *AcGRF* genes

We identified eight AcGRF proteins from the pineapple genome and assigned them to five subfamilies on the basis of phylogenetic analysis (Figures 2A, C). Pineapple has fewer GRF genes than the monocots banana (18) and maize (18) but a similar number to *Setaria italica* and *S. bicolor*. This suggests that *GRF* family expansion in monocots may be associated mainly with lineage-specific WGD events. Pineapple contained members of five of the six major *GRF* subfamilies, similar to rice (four major subfamilies), banana (5), *B. distachyon* (4), *A. thaliana* (5), soybean (5), and *B. rapa* (5) and more than non-angiosperm species such as *M. polymorpha* (0), *S. moellendorffii* (0), and *Pinus taeda* (1). Notably, genes that fell outside the main subfamilies were found only in algae, ferns, gymnosperms, basal angiosperms, and *Brachypodium distachyon*; they may have been preserved from more ancient duplication events or acquired only in specific clades. Among monocots, there were five *GRF* subfamilies in banana (missing subfamily IV) and pineapple (missing subfamily II) and four subfamilies in other species (missing both subfamily II and subfamily IV). By contrast, all dicots except soybean (missing subfamily V) contained members of all six subfamilies. These results suggest that the preservation of specific subfamilies may vary among species with different evolutionary histories. Several genomic evolutionary models have been proposed in model species on the basis of comparative genomic analysis (Wolfe and Shields, 1997; Moore and Purugganan, 2003; Hurley et al., 2005). If duplicated genes are retained, they tend to diverge in their regulatory and coding regions, and differences in coding regions, especially those that alter gene function, may cause amino acid changes or substitutions or changes in exon/intron structure (Xu et al., 2012). In general, proteins from the same family have relatively conserved intron/exon structures, such as β-expansins in *A. thaliana* and rice (Lee and Kende, 2001). However, the number and length of introns are not conserved in GRF genes from pineapple, *A. thaliana*, and rice (Figure 1C), even within the same subfamily, and GRF genes are randomly distributed in the genome (Table 1 and Figure 1B). This suggests that no recent duplication events have occurred in the GRF gene family and that expansions in this gene family reflect early duplication events.

Gain and loss of exons/introns and differences in exon/intron length can result from chromosome rearrangement and fusion (Zan et al., 2020). In pineapple, *AcGRF2* contains three introns, whereas its paralog *AcGRF1* contains only one intron. Likewise, intron lengths were much greater in *AcGRF5* than in its paraphyletic

homologous gene *AcGRF4*. We therefore speculate that differences in intron/exon length or number may have led to different biological functions for *AcGRF4* and *AcGRF5*. Previous studies have reported that the N-terminal QLQ domain is involved in protein–protein interactions (Kim et al., 2003; Kim and Kende, 2004). Here, we found that the QLQ motif of AcGRF5 is actually Gln-Met-Gln, with a Met substituted for Leu (Figure 1A), similar to that of *A. thaliana* AtGRF9 in which Phe is substituted for Leu (Kim et al., 2003). The mutated QLQ domain of AcGRF5 may lead to alterations in its protein interaction activity.

## 4.2 Evolution of the AcGRF genes

Gene duplication, including tandem, segmental, and whole-genome duplication, is one of the main drivers of genome evolution (Moore and Purugganan, 2003), and most seed plants have experienced one or more WGD events in their evolutionary history (Rensing et al., 2007). Pineapple has undergone at least two WGD events (Ming et al., 2015; Ming et al., 2016) (Figure S3), and its gene family expansion reflects the effects of WGDs, as well as tandem and segmental duplications. In this study, we analyzed the presence of GRF genes in 26 species, including algae, mosses, angiosperms, and gymnosperms (Figures 2B, C). GRF genes were present in all species except algae, suggesting that GRF genes appeared in land plants and may have supported their terrestrial adaptation. Two of the eight pineapple GRF genes were associated with segmental duplication events (Figure 3A), consistent with findings in wheat that 26 of 30 GRF genes were associated with such events (Zan et al., 2020). This suggests that segmental duplication may have played an important role in the early expansion of the GRF gene family. The number of *AcGRF* genes was significantly lower in pineapple than in banana (18) compared with other plants used for collinearity analysis, and it was similar to the number of GRF genes in other plant species (Osnato et al., 2010; Kim et al., 2012; Debernardi et al., 2014; Kuijt et al., 2014; Omidbakhshfard et al., 2015). The large number of banana *GRFs* may thus reflect banana-specific duplication events. Notably, there was no significant correlation between the number of GRF genes and genome size. For example, *A. thaliana* has nine *GRFs* and pineapple has eight, despite the pineapple genome being 3.2 times the size (375 Mb) (Ming et al., 2015) of the *A. thaliana* genome (115 Mb) (Hou et al., 2022). After identifying non-redundant GRF genes that showed collinear relationships between pineapple and nine other species, we found that an *AcGRF5* ortholog was present only in pineapple and rice and an *AcGRF3* ortholog was present only in pineapple and water lily (Figure 3C). This suggests that the retention of gene family members can vary among species with different evolutionary histories. All *AcGRF* orthologous genes had Ka/Ks values less than 1 (Figure 3D), indicating that *GRFs* experienced strong purifying selection during evolution.

## 4.3 Inference of biological functions of *AcGRF* genes

Because of the importance of pineapple as a tropical fruit crop and ornamental plant, the mechanisms that regulate its flower and fruit development are of significant interest. The biological functions of pineapple AcGRF genes remain to be clarified, but identification of their presumed orthologs in different species can provide insight into their functions (Nehrt et al., 2011). GRF genes have previously been studied in other species to characterize their regulatory network and understand their functions in a wide range of biological processes (Omidbakhshfard et al., 2015). Here, we used GRFs with known functions in other species to infer potential functions of their pineapple orthologs (Figure S4 and Figure 6). Clade 2 includes *AcGRF1/2*, *AtGRF6*, and *OsGRF1*. Previous studies have shown that *AtGRF6* and *OsGRF1* have a role in controlling leaf size (Kim et al., 2003; Luo et al., 2005) and *OsGRF1* also influences rice stem elongation, juvenile growth, and panicle extension (Van der Knaap et al., 2000; Luo et al., 2005). *AcGRF1/2* in the same clade were highly expressed in early developmental stages of petals and ovules, suggesting that they may have similar functions in pineapple. *ZaGRF6*, *OsGRF3*, and *AcGRF8* were included in clade 3. Heterologous overexpression of *ZaGRF6* increased branching and chlorophyll synthesis and delayed aging in transgenic tobacco, and ectopic overexpression of *OsGRF3* in rice reduced tiller numbers and induced the formation of ectopic roots and shoots on the nodes (Kuijt et al., 2014; Zhang et al., 2021). Promoter analysis revealed the presence of two ABA response elements and one gibberellin response element in the *AcGRF8* gene promoter, and qRT-PCR results showed that *AcGRF8* expression responded significantly to ABA and GA treatment. These results suggest that *AcGRF8* may regulate pineapple aging and root and shoot formation through the abscisic acid and gibberellin pathways. Clade 5 included *AcGRF4/5*, *AtGRF9*, *OsGRF10*, and *ZmGRF10*. In previous studies, overexpression of *ZmGRF10* in maize led to reductions in leaf size and plant height, and knockout and overexpression of *AtGRF9* made petals and other organs of *A. thaliana* larger and smaller, respectively (Wu et al., 2014; Omidbakhshfard et al., 2018). Here, *AcGRF4* was preferentially expressed in pistils and *AcGRF5* in bracts, sepals, and ovules. *AcGRF4* and *AcGRF5* have IAA response elements, and their expression responded to increasing durations of IAA treatment. We therefore suggest that *AcGR4/5* may participate in the regulation of floral organ and leaf size, plant height, and other traits through the auxin pathway. Clade 6 included *AcGRF3*, *AtGRF7*, and *AtGRF8*. Leaves of an *AtGRF7* single-allele mutant were reported to be smaller than those of the wild type (Kim et al., 2012), but this phenomenon was not observed by Lee et al. (2022). *AtGRF8* has been reported to positively regulate cell proliferation (Tsukaya, 2021). We found that *AcGRF3* was preferentially expressed in petals and contained GA-responsive elements in its

promoter. We therefore speculate that *AcGRF3* may regulate pineapple cell proliferation, leaf size, and petal growth through the GA pathway. Clade 8 included *AcGRF6/7*, *PagGRF15*, and *OsGRF7/8*. *PagGRF15* regulates leaf size through its effects on cell expansion during poplar leaf development, and *OsGRF7/8* influence plant structure and leaf growth by regulating gibberellin and indole-3-acetic acid metabolism. Here, *AcGRF6* and *AcGRF7* were preferentially expressed in petals, pistils, and ovules. qRT-PCR experiments showed that GA significantly induced the expression of *AcGRF7* but inhibited that of *AcGRF6*, and IAA significantly inhibited *AcGRF6*. *AcGRF6/7* may therefore regulate pineapple flower and leaf development and plant structure through GA and IAA pathways.

## 5 Conclusion

We identified eight GRF gene family members on seven chromosomes of the pineapple genome and classified them into five of six known subfamilies on the basis of phylogenetic analysis. In contrast to other gene families, the GRFs showed little conservation of gene structure and motif composition. Collinearity analysis showed that early segmental duplications promoted expansion of the pineapple GRF gene family, and purifying selection was the main force acting on the GRF genes. The paralogs *AcGRF1* and *AcGRF2* had different expression profiles in different floral organs, perhaps related to their differences in structure. Transcriptome data suggested that all AcGRF genes were involved in regulation of floral organ development, and *AcGRF1/2/3/6/7* may have functionally redundant roles in petal development. The *AcGRF* promoters contained *cis*-acting elements involved in hormone response, developmental regulation, and stress response, and changes in expression of *AcGRFs* under ABA, GA, IAA, JA, and NaCl treatments may reflect their different regulatory effects. Protein interactions between AcGRF proteins or with YABBY, TCP, and AcGRF-miRNA may contribute to the formation of more complex regulatory networks. These results support further research on the functions and regulatory mechanisms of AcGRF transcription factors during pineapple growth and development.

## 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

YH, CC conceived the project. WY and JW performed the experiments. WY, AL, JW, CL and WZ performed data analysis. WY and AL wrote the original manuscript draft. All authors participated in discussion and revised the manuscript. All authors reviewed and approved the manuscript.

## Funding

This research was funded by the National Natural Science Foundation of China(32272677, 31901355), Special Project of Rural Revitalization Strategy in Guangdong Province(2022-NPY-00-031), the National Key R&D Program of China (2018YFD1000500).

## Acknowledgments

We thank Guiyang Watch Biotechnology for their advice on gene family data analyses. TopEdit (www.topeditsci.com) for its linguistic assistance during the preparation of this manuscript. The authors thank lab members for their assistance.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1159223/full#supplementary-material

# References


Arvidsson, S., Pérez-Rodríguez, P., and Mueller-Roeber, B. (2011). A growth phenotyping pipeline for arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytol.* 191, 895–907. doi: 10.1111/j.1469-8137.2011.03756.x

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bao, M., Bian, H., Zha, Y., Li, F., Sun, Y., Bai, B., et al. (2014). miR396a-mediated basic helix–Loop–Helix transcription factor bHLH74 repression acts as a regulator for root growth in arabidopsis seedlings. *Plant Cell Physiol.* 55, 1343–1353. doi: 10.1093/pcp/pcu058

Baruwa, O. I. (2013). Profitability and constraints of pineapple production in osun state, Nigeria. *J. Hortic. Res.* 21, 59–64. doi: 10.2478/johr-2013-0022

Biłas, R., Szafran, K., Hnatuszko-Konka, K., and Kononowicz, A. K. (2016). Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tissue Organ Culture* 127, 269–287. doi: 10.1007/s11240-016-1057-7

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., et al. (2016). UniProtKB/Swiss-prot, the manually annotated section of the UniProt KnowledgeBase: How to use the entry view. *Plant Bioinform.: Methods*, 1374. 23–54. doi: 10.1007/978-1-4939-3167-5_2

Casadevall, R., Rodriguez, R. E., Debernardi, J. M., Palatnik, J. F., and Casati, P. (2013). Repression of growth regulating factors by the MicroRNA396 inhibits cell proliferation by UV-b radiation in arabidopsis leaves. *Plant Cell* 25, 3570–3583. doi: 10.1105/tpc.113.117473

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, Y., Dan, Z., Gao, F., Chen, P., Fan, F., and Li, S. (2020). Rice GROWTH-REGULATING FACTOR7 modulates plant architecture through regulating GA and indole-3-Acetic acid metabolism. *Plant Physiol.* 184, 393–406. doi: 10.1104/pp.20.00302

Chen, C., Wu, Y., and Xia, R. (2022). A painless way to customize circos plot: From data preparation to visualization using TBtools. *iMeta* 1 (3), e35. doi: 10.1002/imt2.35

Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., and Town, C. D. (2017). Araport11: A complete reannotation of the arabidopsis thaliana reference genome. *Plant J.* 89, 789–804. doi: 10.1111/tpj.13415

Choi, D., Kim, J. H., and Kende, H. (2004). Whole genome analysis of the OsGRF gene family encoding plant-specific putative transcription activators in rice (Oryza sativa l.). *Plant Cell Physiol.* 45, 897–904. doi: 10.1093/pcp/pch098

Dai, X., Zhuang, Z., and Zhao, P. X. (2018). psRNATarget: A plant small RNA target analysis server, (2017 release). *Nucleic Acids Res.* 46, W49–W54. doi: 10.1093/nar/gky316

Debernardi, J. M., Mecchia, M. A., Vercruyssen, L., Smaczniak, C., Kaufmann, K., Inze, D., et al. (2014). Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. *Plant J.* 79, 413–426. doi: 10.1111/tpj.12567

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

Emms, D. M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 1–14. doi: 10.1186/s13059-019-1832-y

Hou, X., Wang, D., Cheng, Z., Wang, Y., and Jiao, Y. (2022). A near-complete assembly of an arabidopsis thaliana genome. *Mol. Plant* 15, 1247–1250. doi: 10.1016/j.molp.2022.05.014

Huang, Y., Chen, J., Li, J., Li, Y., and Zeng, X. (2022). Genome-wide identification and analysis of the growth-regulating factor family in zanthoxylum armatum DC and functional analysis of ZaGRF6 in leaf size and longevity regulation. *Int. J. Mol. Sci.* 23, 9043. doi: 10.3390/ijms23169043

Hurley, I., Hale, M. E., and Prince, V. E. (2005). Duplication events and the evolution of segmental identity. *Evol. Dev.* 7, 556–567. doi: 10.1111/j.1525-142x.2005.05059.x

Kim, J. H., Choi, D., and Kende, H. (2003). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in arabidopsis. *Plant J.* 36, 94–104. doi: 10.1046/j.1365-313x.2003.01862.x

Kim, J. H., and Kende, H. (2004). A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in arabidopsis. *Proc. Natl. Acad. Sci. U. States America* 101, 13374–13379. doi: 10.1073/pnas.0405450101

Kim, J. H., and Lee, B. H. (2006). GROWTH-REGULATING FACTOR4 of arabidopsis thaliana is required for development of leaves, cotyledons, and shoot apical meristem. *J. Plant Biol.* 49, 463–468. doi: 10.1007/bf03031127

Kim, J.-S., Mizoi, J., Kidokoro, S., Maruyama, K., Nakajima, J., Nakashima, K., et al. (2012). Arabidopsis GROWTH-REGULATING FACTOR7 functions as a transcriptional repressor of abscisic acid– and osmotic stress–responsive genes, including DREB2A. *Plant Cell* 24, 3393–3405. doi: 10.1105/tpc.112.100933

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

Kuijt, S. J. H., Greco, R., Agalou, A., Shao, J., 't Hoen, C. C. J., Övernäs, E., et al. (2014). Interaction between the GROWTH-REGULATING FACTOR and KNOTTED1-LIKE HOMEOBOX families of transcription factors. *Plant Physiol.* 164, 1952–1966. doi: 10.1104/pp.113.222836

Lee, Y., and Kende, H. (2001). Expression of β-expansins is correlated with internodal elongation in deepwater rice. *Plant Physiol.* 127, 645–654. doi: 10.1104/pp.010345

Lee, G.-H., Lee, B. H., Jung, J.-H., Lee, S.-J., Mai, T.-T., and Kim, J. H. (2022). Systematic assessment of the positive role of arabidopsis thaliana GROWTH-REGULATING FACTORs in regulation of cell proliferation during leaf growth. *J. Plant Biol.* 65, 413–422. doi: 10.1007/s12374-022-09366-1

Lee, S.-J., Lee, B. H., Jung, J.-H., Park, S. K., Song, J. T., and Kim, J. H. (2017). Growth-regulating factor and grf-interacting factor specify meristematic cells of gynoecia and anthers. *Plant Physiol.* 176, 717–729. doi: 10.1104/pp.17.00960

Lescot, M. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* 424, 147–151. doi: 10.1038/nature01763

Liang, G., He, H., Li, Y., Wang, F., and Yu, D. (2013). Molecular mechanism of microRNA396 mediating pistil development in arabidopsis. *Plant Physiol.* 164, 249–258. doi: 10.1104/pp.113.225144

Liu, H., Guo, S., Xu, Y., Li, C., Zhang, Z., Zhang, D., et al. (2014). OsmiR396d-regulated OsGRFs function in floral organogenesis in rice through binding to their targets OsJMJ706 and OsCR4. *Plant Physiol.* 165, 160–174. doi: 10.1104/pp.114.235564

Liu, J., Rice, J. H., Chen, N., Baum, T. J., and Hewezi, T. (2014). Synchronization of developmental processes and defense signaling by growth regulating transcription factors. *PloS One* 9, e98477. doi: 10.1371/journal.pone.0098477

Lobo, M., and Paull, R. (2017). *Handbook of pineapple technology: production, postharvest science, processing and nutrition* (John Wiley & Sons).

Luo, A.-D., Liu, L., Tang, Z.-S., Bai, X.-Q., Cao, S.-Y., and Chu, C.-C. (2005). Down-regulation of OsGRF1 gene in rice rhd1 mutant results in reduced heading date. *J. Integr. Plant Biol.* 47, 745–752. doi: 10.1111/j.1744-7909.2005.00071.x

Lynch, M., and Conery, J. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., et al. (2005). Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U. States America* 102, 5454–5459. doi: 10.1073/pnas.0501102102

Mao, Q., Chen, C., Xie, T., Luan, A., Liu, C., and He, Y. (2018). Comprehensive tissue-specific transcriptome profiling of pineapple (Ananas comosus) and building an eFP-browser for further study. *PeerJ* 6, e6028. doi: 10.7717/peerj.6028

Ming, R., VanBuren, R., Wai, C. M., Tang, H., Schatz, M. C., Bowers, J. E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47, 1435–1442. doi: 10.1038/ng.3435

Ming, R., Wai, C. M., and Guyot, R. (2016). Pineapple genome: A reference for monocots and CAM photosynthesis. *Trends Genet.* 32, 690–696. doi: 10.1016/j.tig.2016.08.008

Moore, R. C., and Purugganan, M. D. (2003). The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. U. States America* 100, 15682–15687. doi: 10.1073/pnas.2535513100

Nehrt, N. L., Clark, W. T., Radivojac, P., and Hahn, M. W. (2011). Testing the ortholog conjecture with comparative functional genomic data from mammals. *PloS Comput. Biol.* 7, e1002073. doi: 10.1371/journal.pcbi.1002073

Omidbakhshfard, M. A., Fujikura, U., Olas, J. J., Xue, G.-P., Balazadeh, S., and Mueller-Roeber, B. (2018). GROWTH-REGULATING FACTOR 9 negatively regulates arabidopsis leaf growth by controlling ORG3 and restricting cell proliferation in leaf primordia. *PloS Genet.* 14, e1007484. doi: 10.1371/journal.pgen.1007484

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

Osnato, M., Stile, M. R., Wang, Y., Meynard, D., Curiale, S., Guiderdoni, E., et al. (2010). Cross talk between the KNOX and ethylene pathways is mediated by intron-binding transcription factors in barley. *Plant Physiol.* 154, 1616–1632. doi: 10.1104/pp.110.161984

Pajoro, A., Madrigal, P., Muiño, J. M., Matus, J., Jin, J., Mecchia, M. A., et al. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol.* 15, R41. doi: 10.1186/gb-2014-15-3-r41

Pu, C.-X., Ma, Y., Wang, J., Zhang, Y.-C., Jiao, X.-W., Hu, Y.-H., et al. (2012). Crinkly4 receptor-like kinase is required to maintain the interlocking of the palea and

lemma, and fertility in rice, by promoting epidermal cell differentiation. *Plant J.* 70, 940–953. doi: 10.1111/j.1365-313x.2012.04925.x

Rensing, S. A., Ick, J., Fawcett, J. A., Lang, D., Zimmer, A., Van de Peer, Y., et al. (2007). An ancient genome duplication contributed to the abundance of metabolic genes in the moss physcomitrella patens. *BMC Evol. Biol.* 7, 130. doi: 10.1186/1471-2148-7-130

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, 2105–2110. doi: 10.1126/science.290.5499.2105

Sanewski, G., Bartholomew, D., and Paull, R. (2018). *The pineapple: botany, production and uses* (CAB International).

Tsukaya, H. (2021). The leaf meristem enigma: The relationship between the plate meristem and the marginal meristem. *Plant Cell* 33, 3194–3206. doi: 10.1093/plcell/koab190

Van Bel, M., Diels, T., Vancaester, E., Kreft, L., Botzki, A., Van de Peer, Y., et al. (2017). PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* 46, D1190–D1196. doi: 10.1093/nar/gkx1002

Van der Knaap, E., Kim, J. H., and Kende, H. (2000). A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. *Plant Physiol.* 122, 695–704. doi: 10.1104/pp.122.3.695

Wang, L., Li, Y., Jin, X., Liu, L., Dai, X., Liu, Y., et al. (2020). Floral transcriptomes reveal gene networks in pineapple floral growth and fruit development. *Commun.* 3, 500. doi: 10.1038/s42003-020-01235-2

Wolfe, K., and Shields, D. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387, 708–713. doi: 10.1038/42711

Wu, L., Zhang, D., Xue, M., Qian, J., He, Y., and Wang, S. (2014). Overexpression of the maize GRF10, an endogenous truncated growth-regulating factor protein, leads to reduction in leaf size and plant height. *J. Integr. Plant Biol.* 56, 1053–1063. doi: 10.1111/jipb.12220

Xu, G., Guo, C., Shan, H., and Kong, H. (2012). Divergence of duplicate genes in exon-intron structure. *Proc. Natl. Acad. Sci. U. States America* 109, 1187–1192. doi: 10.1073/pnas.1109047109

Yusuf, N. H. M., Ong, W. D., Redwan, R. M., Latip, M. A., and Kumar, S. V. (2015). Discovery of precursor and mature microRNAs and their putative gene targets using high-throughput sequencing in pineapple (Ananas comosus var. comosus). *Gene* 571, 71–80. doi: 10.1016/j.gene.2015.06.050

Zan, T., Zhang, L., Xie, T., and Li, L. (2020). Genome-wide identification and analysis of the growth-regulating factor (GRF) gene family and GRF-interacting factor family in triticum aestivum l. *Biochem. Genet.* 58, 705–724. doi: 10.1007/s10528-020-09969-8

Zhang, L., and Chen, F. (2019). The water lily genome and the early evolution of flowering plants. *Nature* 577, 79–84. doi: 10.1038/s41586-019-1852-5

Zhang, G.-Q., Liu, K.-W., Li, Z., Lohaus, R., Hsiao, Y.-Y., Niu, S.-C., et al. (2017). The apostasia genome and the evolution of orchids. *Nature* 549, 379–383. doi: 10.1038/nature23897

Zhang, B., Tong, Y., Luo, K., Zhai, Z., Liu, X., Shi, Z., et al. (2021). Identification of GROWTH-REGULATING FACTOR transcription factors in lettuce (Lactuca sativa) genome and functional analysis of LsaGRF5 in leaf size regulation. *BMC Plant Biol.* 21, 1–13. doi: 10.1186/s12870-021-03261-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 July 2023
DOI 10.3389/fpls.2023.1170004

Check for updates

OPEN ACCESS

EDITED BY
Christos Bazakos,
Max Planck Institute for Plant Breeding Research, Germany

REVIEWED BY
Peng-Fei Ma,
Chinese Academy of Sciences (CAS), China
Huasheng Peng,
China Academy of Chinese Medical Sciences, China
Vasileios Papasotiropoulos,
University of Patras, Greece

*CORRESPONDENCE
Xiao Ma
maroar@126.com
Yao Ling
ly9729752@163.com

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

RECEIVED 20 February 2023
ACCEPTED 05 July 2023
PUBLISHED 24 July 2023

CITATION
Yuan S, Nie C, Jia S, Liu T, Zhao J, Peng J, Kong W, Liu W, Gou W, Lei X, Xiong Y, Xiong Y, Yu Q, Ling Y and Ma X (2023) Complete chloroplast genomes of three wild perennial *Hordeum* species from Central Asia: genome structure, mutation hotspot, phylogenetic relationships, and comparative analysis.
*Front. Plant Sci.* 14:1170004.
doi: 10.3389/fpls.2023.1170004

COPYRIGHT
© 2023 Yuan, Nie, Jia, Liu, Zhao, Peng, Kong, Liu, Gou, Lei, Xiong, Xiong, Yu, Ling and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Complete chloroplast genomes of three wild perennial *Hordeum* species from Central Asia: genome structure, mutation hotspot, phylogenetic relationships, and comparative analysis

Shuai Yuan[1,2†], Cong Nie[1†], Shangang Jia[3], Tianqi Liu[1], Junming Zhao[1], Jinghan Peng[1], Weixia Kong[1], Wei Liu[1], Wenlong Gou[2], Xiong Lei[2], Yi Xiong[1], Yanli Xiong[1], Qingqing Yu[1], Yao Ling[1*] and Xiao Ma[1*]

[1]College of Grassland Science and Technology, Sichuan Agricultural University, Chengdu, China, [2]Sichuan Academy of Grassland Sciences, Chengdu, China, [3]College of Grassland Science and Technology, China Agricultural University, Beijing, China

*Hordeum* L. is widely distributed in mountain or plateau of subtropical and warm temperate regions around the world. Three wild perennial *Hordeum* species, including *H. bogdanii*, *H. brevisubulatum*, and *H. violaceum*, have been used as forage and for grassland ecological restoration in high-altitude areas in recent years. To date, the degree of interspecies sequence variation in the three *Hordeum* species within existing gene pools is still not well-defined. Herein, we sequenced and assembled chloroplast (cp) genomes of the three species. The results revealed that the cp genome of *H. bogdanii* showed certain sequence variations compared with the cp genomes of the other two species (*H. brevisubulatum* and *H. violaceum*), and the latter two were characterized by a higher relative affinity. Parity rule 2 plot (PR2) analysis illuminated that most genes of all ten *Hordeum* species were concentrated in nucleotide T and G. Numerous single nucleotide polymorphism (SNP) and insertion/deletion (In/Del) events were detected in the three *Hordeum* species. A series of hotspots regions (*tRNA-GGU ~ tRNA-GCA*, *tRNA-UGU ~ ndhJ*, *psbE ~ rps18*, *ndhF ~ tRNA-UAG*, etc.) were identified by mVISTA procedures, and the five highly polymorphic genes (*tRNA-UGC*, *tRNA-UAA*, *tRNA-UUU*, *tRNA-UAC*, and *ndhA*) were proved by the nucleotide diversity (Pi). Although the distribution and existence of cp simple sequence repeats (cpSSRs) were predicted in the three *Hordeum* cp genomes, no rearrangement was found between them. A similar phenomenon has been found in the cp genome of the other seven *Hordeum* species, which has been published so far. In addition, evolutionary relationships were

reappraised based on the currently reported cp genome of *Hordeum* L. This study offers a framework for gaining a better understanding of the evolutionary history of *Hordeum* species through the re-examination of their cp genomes, and by identifying highly polymorphic genes and hotspot regions that could provide important insights into the genetic diversity and differentiation of these species.

KEYWORDS

***Hordeum*, chloroplast genome, parity rule 2, repeated sequences, hotpot, phylogenic tree**

## 1 Introduction

As secretory organs and active metabolic centers, chloroplasts (cp) are considered the source of energy that drives the evolution of early life (Liu et al., 2018). Although most of the genetic information is provided by the nuclear genome, the cp genome is used to perform variation analysis due to its small size and matrilineal inheritance without gene recombination interference (Gumeni et al., 2017; Shen et al., 2018). Therefore, sequence variation in cp genomes plays a key role in studying plant evolution, and genetic diversity (Xiong et al., 2020a). With the advent of high-throughput sequencing technologies, especially Illumina sequencing, sequence and structure information obtained from the whole cp genome has been elucidated in some vital species (Ogihara et al., 2000; Sajjad et al., 2017). Cp genomes contain several functional genes, such as photosynthesis-related genes, expression-related genes, and biosynthesis-related genes (Bailey et al., 2020). Differential gene detection and phylogeny analysis among genera or families using cp genome sequences is another effective method for studying evolutionary patterns due to the conservative property of cp DNA, mainly in content and arrangement mode. Generally, the structure of the cp genome is quadrantal, containing two inverted repeat (IR) sequences divided by a large single-copy (LSC) region and a small single-copy (SSC) region (Wu et al., 2021). However, four specific *Hordeum* species, *H. pubiflorum*, *H. murinum*, *H. marinum*, and *H. bulbosum*, were a noticeable exception to this typical structure with IR loss or missing introns (Bernhardt et al., 2017). It is noteworthy that this phenomenon was rarely reported in plants in the Poaceae family but it was often found in plants in the Leguminaceae family (Xue et al., 2019).

Derived from the *Triticeae* tribe of the Gramineae family, *Hordeum* L. is composed of approximately 45 species or subspecies, which are distributed in the southern and northern hemispheres, with four species diversity centers, including Southwest Asia, Central Asia, North America, and Southern America (Brassac and Blattner, 2015; and Reinert et al., 2019). The genus *Hordeum* consists of one cultivated species, namely *H. vulgare*, and abundant wild species, such as *H. vulgare* subsp. *spontaneum*, *H. bogdanii*, *H. brevisubulatum*, *H.violaceum* (*H. roshevilzii*), etc. Wild species — which gradually undergo environmental selection — often possess favorable genes such as disease resistance and insect resistance genes and thus are considered important germplasms for genetic improvement (Alyr et al., 2020). Investigation of the genetic diversity and kinship between wild and cultivated species may provide a perspective for the development and utilization of advantageous genes and extension of the genetic basis of cultivars. Previous studies have explored the phylogenetic relationships between wild and cultivated and annual and perennial *Hordeum* species, which mainly depended on the mitochondrial genome sequences (Hisano et al., 2016) or partial nuclear single-copy genome sequence analysis (Jonathan and Blattner, 2015). However, there are relatively few reports on the phylogenetic relationships using complete cp genomes of the genus *Hordeum*. Particularly, large-scale phylogenetic analysis of wild perennial species originating from North Central Asia (*H. bogdanii*, *H. brevisubulatum*, and *H. violaceum*) and those distributed elsewhere is still insufficient. Therefore, performing complete cp genome sequencing of these three wild perennial *Hordeum* species to identify some plastid key genes in interspecific genetic differentiation between the wild and cultivated and/or perennial and annual *Hordeum* species is of great significance, to further improve the phylogenetic relationships and genome structure of the genus *Hordeum*.

Here, complete cp genomes of three wild perennial *Hordeum* species, *H. bogdanii*, *H. brevisubulatum*, and *H. violaceum*, were sequenced and annotated, to determine the cp genome size, nucleotide diversity (Pi), repeat sequences, insertions/deletions (In/Dels), single nucleotide polymorphisms (SNPs). Sequence synteny, relative synonymous codon usage, Parity rule 2 (PR2) analysis, rearrangements, and IR expansions or contractions were evaluated among 10 *Hordeum* species (*H. bogdanii, H. brevisubulatum*, *H. violaceum, H. jubatum*, *H. bulbosum*, *H. marinum*, *H. murinum*, *H. pubiflorum*, *H. vulgare* subsp. *spontaneum*, and *H. vulgare*). In addition, phylogenetic relationships of the sequenced *Hordeum* species from other whole sequenced Poaceae species were revealed. Meanwhile, the degree of variation between wild and cultivated and annual and perennial *Hordeum* species was further evaluated. This study contributes to the expansion of the cp genome database.

# 2 Methods

## 2.1 Plant material, DNA extraction and sequencing

Three *Hordeum* species, *H. bogdanii, H. brevisubulatum*, and *H. violaceum*, were from NPGS (National Plant Germplasm System of the United States; Supplementary Table 1). In total, 100 mg leaves were harvested at the three-leaf stage, and then total genomic DNA was extracted using the plant DNA Extraction Kit (Tiangen, Beijing, China) as per manufacture's instruction. DNA concentration was quantified using 0.1% agarose gel, libraries were established, and DNA with good quality was selected and sequenced using the Illumina NovaSeq platform with a read length of PE150.

## 2.2 Chloroplast genome assembly and annotation

The complete circular genome sequence cannot be directly obtained by one-time splicing because of the characteristics of next-generation sequencing (NGS), genomic repeats, a specific structure of the genome, and related factors. Therefore, a different complicated strategy was performed: The kernel modules were assembled using the SPAdes v3.10.1 (Saint Petersburg State University, Saint, Russia) (Safonova et al., 2014) software for the cp genome of three species, which is not dependent on the reference genome. The contig was obtained using the kmer iterative extend seed. The SSPACE v2.0 procedure was used (BaseClear BV, Einsteinweg, Leiden, The Netherlands) (Boetzer et al., 2011) to acquire scaffolds by connecting contig sequences. The gap of scaffolds sequence was constructed using Gapfiller V2.1.1 procedure (BaseClear BV, Einsteinweg, Leiden, The Netherlands) to assemble a complete pseudo sequence (Boetzer and Pirovano, 2012). The alignment-correction method was used to align the sequencing sequence into the pseudo genome, which was later rearranged according to the cp structure of the three species, thereby obtaining a complete cp circular genome sequence.

Cp gene structure annotation plays an important role in cp genome sequencing. Blast v2.2.25 (U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA) (Kent and Brumbaugh, 2002) was used to align CDS sequences of cp genome in NCBI. The gene annotation results of cp genomes for three *Hordeum* species were acquired using a manual correction. Moreover, to obtain gene annotation, rRNA and tRNA sequences of cp genomes were aligned in NCBI (https://www.ncbi.nlm.nih.gov/) database using HMMER v3.1b2 (HHMI/Harvard University, Boston, USA; The European Bioinformatics Institute, Cambridge, UK) (Finn et al., 2011) and Aragorn v1.2.38 programs (Murdoch University, Western Australia, Australia; Lund University, Lund, Sweden) (Dean and Bjorn, 2004). In addition, *H. vulgare* subsp. *spontaneum* (KC912688.1) was used as a reference sequence for quality control of the cp genome after assembly.

## 2.3 Prediction of repetitive sequences

The Simple Sequence Repeats (SSRs) markers are a class of tandem repeats with motifs consisting of several nucleotides group (usually 1~6) as repeating units. The SSR marker is called cpSSR marker on cp genomes. CpSSR were identified and analyzed using the software MISA v1.0 (Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Corrensstr. 3, 06466 Seeland, Germany) (Beier et al., 2017). CpSSR parameters were described as A-B, with A representing the number of repetitions and B representing the total number of the base unit in a sequence. For example, 1-8 indicates more than 8 repetitions of a single-base, 2-5 indicates more than 5 repetitions of a double-base, 3-3 more than three repetitions of triple-base, 4-3, 5-3, 6-3 and so on. Furthermore, the interspersed repeats sequences, which are a different kind of repetitive sequences from tandem repeats and have both forward and palindromic repeats (including reverse and complementary) with a minimum size of 15 bp, sequence coherence of more than 90% and are distributed throughout the genome, were identified using the Vmatch v2.3.0 (http://www.vmatch.de/) program.

## 2.4 Relative synonymous codon usage and parity rule 2 analysis

The degeneracy of codons show that each amino acid has one to six codons. The heterogeneity of synonymous codon usage is called Relative Synonymous Codon Usage (RSCU). To highlight the relative biasness between amino acids and codons, the RSCU was analyzed using the MEGA v10.1.8 program (Kumar et al., 2008).

The complete cp genomes of the three *Hordeum* species sequenced in this study and seven other *Hordeum* species (*H. bulbosum*, *H. jubatum*, *H. marinum*, *H. murinum*, *H. pubiflorum*, *H. vulgare* subsp. *spontaneum*, *H. vulgare*) were downloaded from the NCBI database and used for PR2 analysis to evaluate nucleotide usage bias in the coding genes of them (Wei et al., 2014). Base A, T, C and G content at the third site of synonymous codons were calculated using the MEGA v10.1.8 software.

## 2.5 Analysis of sequences variation and Ka/Ks

SNP (Single Nucleotide Polymorphism) refers to the DNA sequence polymorphism caused by the variation (insertions or deletions (In/Dels)) of a single nucleotide at the genomic level and accounts for more than 90% of known polymorphisms. The cp genomes of three *Hordeum* materials were aligned using MAFFT program, version v7.310 (https://mafft. cbrc. jp/alignment/software/) (Standley, 2013) to identify SNP and In/Dels. In addition, the nucleotide diversity (Pi) and Ka/Ks in this study were calculated using the conjunct genes and protein-coding genes of the three *Hordeum* materials detected. Base mutation, including non-

Synonymous mutations (Ka) and synonymous mutations (Ks) causes changes in amino acids, which ratios > 1 is called a positive selection effect and < 1 is named a purified selection effect. Pi is considered an important tool that able to reveal the variation of size of nucleic acid sequences, and a range of potential molecular markers can be provided based on the regions of high variability for population genetics (Meng et al., 2018). The Ka/Ks and Pi values were calculated using KaKs_Calculator v2.0 (https://sourceforge.net/projects/kakscalculator2/) (Zhang et al., 2006) and VCFTOOLS (Danecek et al., 2011), respectively. Nevertheless, before achieving the above tasks, the CDS sequences of the conjunct genes in each species were globally aligned using MAFFT software.

## 2.6 Multiple Cp genomes alignment

Alignment and collinearity of 10 *Hordeum* species complete cp genomes, *H. bogdanii, H. brevisubulatum*, *H. violaceum, H. jubatum*, *H. bulbosum*, *H. marinum*, *H. murinum*, *H. pubiflorum*, *H. vulgare* subsp. *spontaneum*, and *H. vulgare*, was analyzed using Mauve (Darling et al., 2004) and Mvista tools (http://genome.lbl.gov/vista/mvista/submit.html). The IRSCOPE online software (https://irscope.shinyapps.io/irapp/) was used to evaluate the expansion or contraction of IR and SC regions boundary for six species (*H. bogdanii*, *H. brevisubulatum*, *H. violaceum*, *H. jubatum*, *H. vulgare* subsp. *spontaneum*, and *H. vulgare*).

## 2.7 Phylogenetic analysis

A total of 28 Poaceae species published in NCBI (Supplementary Table 2), and three *hordeum* species (*H. bogdanii* (CNS0491101), *H. brevisubulatum* (CNS0491102), *H. violaceum* (CNS0491103)) that in this study were sequenced to establish the phylogenetic tree. *Saccharum spontaneum* (LN896360.1) and *Sorghum bicolor* (NC008602.1) were the outgroups. MAFFT and RAxML v8.2.10 software (https://cme. h-its. org/exelixis/software. html) that follow GTR model and Hill Climbing algorithm were carried out to achieve the multi-sequence alignment and construction of the phylogenetic tree for different species, respectively.

# 3 Results

## 3.1 Characteristics of Cp genomes of six *Hordeum* species

Due to the loss of the IR region in the cp genomes of *H. pubiflorum*, *H. murinum*, *H. marinum*, and *H. bulbosum*, cp genome characteristics of only six *Hordeum* species, *H. bogdanii*, *H. brevisubulatum*, *H. violaceum*, *H. jubatum*, *H. vulgare* subsp. *spontaneum*, and *H. vulgare* were selected for comparison of cp genome characteristics (Figure 1). This comparison also included IR expansion and contraction. *H. vulgare* had the smallest cp genome size (136,462 bp) compared with that of the other five species (*H. bogdanii* (137,141 bp), *H. brevisubulatum* (137,002 bp), *H. violaceum* (137,032 bp), and *H. spontaneum* (136,536 bp), *H. jubatum* (136,826 bp), while it also had the highest GC content and total number of genes. Illumina paired-end sequencing yielded 26,262,890, 25,330,242, and 25,890,515 ReadSum (pair-end reads) from *H. bogdanii, H. brevisubulatum*, and *H. violaceum*, respectively. Q20 and Q30 (the percentage of bases with a mass value ≥20 and ≥30, respectively) were both more than 85%. The three perennial species (*H. bogdanii, H. brevisubulatum*, and *H. violaceum*) belonged to a typical quadrantal model, consisted of two copies of IR regions (IRs 21,573-21,587 bp), and were separated by LSC (81,128-81,169 bp) and SSC (12,728-12,798 bp) regions, which are the common feature of the majority of plants in the Poaceae family (Figure 1, Table 1). The overall GC content in the cp genomes of *H. bogdanii, H. brevisubulatum*, and *H. violaceum* was 38.23, 28.28, and 38.27%, respectively, and the percentage distributed in the IR regions was the highest than that in LSC and SSC regions. A total of 129, 131, and 131 genes were located in the complete cp genomes of *H. bogdanii, H. brevisubulatum*, and *H. violaceum*, respectively. Thirty-eight ribosomal RNA (rRNA) genes, 8 transfer RNA (tRNA) genes, and 85 messenger RNA (mRNA) genes were distributed in both *H. brevisubulatum* and *H. violaceum*. Interestingly, the annual cultivated species (*H. vulgare*) had the largest number of genes compared with the other five species, but these genes these genes were all attributed to tRNA.

Out of the 113 genes were shared by the five cp *Hordeum* genomes (*H. bogdanii, H. brevisubulatum*, *H. violaceum*, *H. vulgare* subsp. *spontaneum*, and *H. vulgare*) (Table 2). 46 were annotated to photosynthesis-related genes such as the large subunit of rubisco, a subunit of photosystem I, a subunit of photosystem II, a subunit of ATP synthase, cytochrome b/f complex, c-type cytochrome synthesis, and subunit of NADH dehydrogenase. Thirty-four genes were involved in self-replication, of which 30 genes and 4 genes were related to tRNA and rRNA, respectively. In addition, 12 genes encoding ribosomal proteins, as well as 14 genes were assembled into transcription. Interestingly, *trnI-GAU*, *trnG-UCC*, *rps12*, and *rps16* genes were unique to two annual species (*H. vulgare* subsp. *spontaneum* and *H. vulgare*), while *trnT-CGU* and *trnS-CGA* genes were specific to three perennial species (*H. bogdanii, H. brevisubulatum*, and *H. violaceum*). More mutations may accumulate in introns because they are less constrained by natural selection than exons (Xiong et al., 2020b). Ten genes that contained a single intron in three cp genomes were collected (Supplementary Table 3).

## 3.2 Repeat sequence analysis

Two different types of repeat sequences, which includes scattered repetitive sequences (palindrome repeats and direct repeats) and simple sequence repeats (SSR), were carefully analyzed Using MISA v1.0 and Vmatch v2.3.0, respectively. A total of 231 (forward type, 125 and palindromic type, 106), 220 (forward type 115 and palindromic type, 105), and 218 (forward

10.3389/fpls.2023.1170004

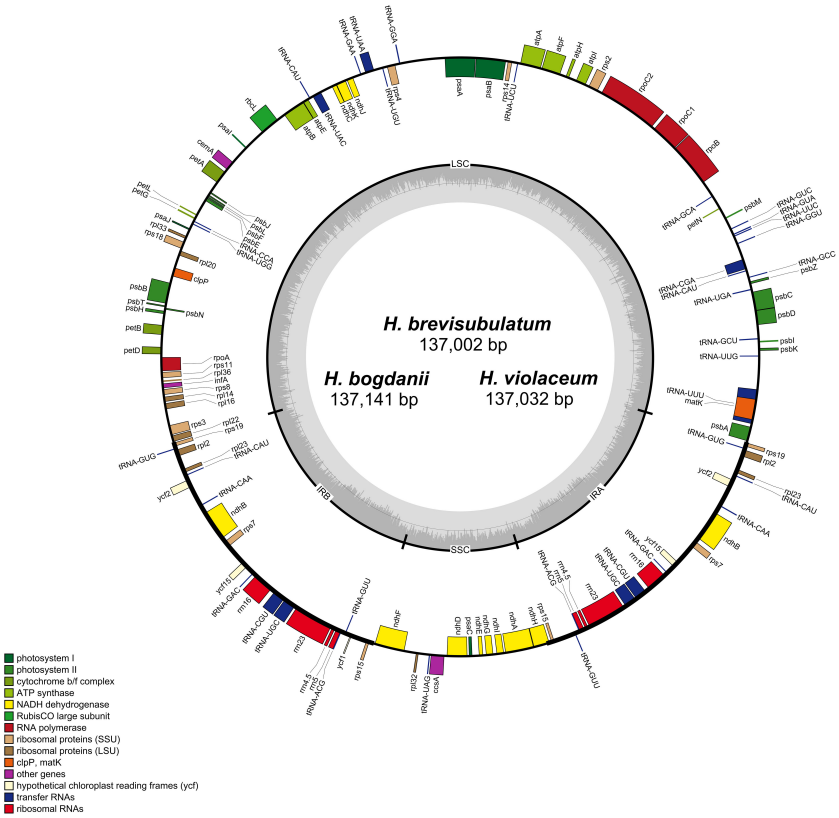



FIGURE 1
Gene maps of *H. bogdanii, H. brevisubulatum, and H. violaceum* cp genomes. Genes inside and outside the circle undergo clockwise and counterclockwise transcription in the gene map. Dark gray and light gray color represent guanine and cytosine (GC) content and adenine and thymine (AT) content, respectively.

type,115 and palindromic type 103) scattered repetitive sequences were predicted in *H. bogdanii, H. brevisubulatum*, and *H. violaceum*, respectively (Figure 2A). Their common characteristic is the number of repeats reached the peak at a repeat length of 15 bp (Figure 2C). SSR, a tandem repeat sequence of dozens of nucleotides generally composed of a series of repeat units (1-6 bp in length), was distributed throughout the genome. A total of 182 SSR in the cp genome of *H. bogdanii* was detected, which was greater than that of *H. brevisubulatum* (178) and *H. violaceum* (176) (Figure 2B). The number of mononucleotides (primarily poly-A or poly-T) accounted for the largest proportion of total SSR, which was above 59% (Figure 2B). Interestingly, trinucleotide (AGC) and tetranucleotide (AACA and AGAA) SSR were found only in *H. bogdanii*, and other types of SSR nucleotides in the cp genome of the three wild perennial *Hordeum* species were predicted with a fixed distribution (Figure 2D), which warrants further investigation in

TABLE 1 Comparison of the six *Hordeum* chloroplast genomes.

| Species | Improvement status | Size (bp) | | | | GC content (%) | | | | tRNA | rRNA | mRNA | Genes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Cp genome | LSC | SSC | IR | Cp genome | LSC | SSC | IR | | | | |
| *H. bogdanii* | Wild, perennial | 137141 | 81169 | 12798 | 21587 | 38.23 | 36.20 | 32.10 | 43.87 | 38 | 8 | 83 | 129 |
| *H. brevisubulatum* | Wild, perennial | 137002 | 81128 | 12728 | 21573 | 38.28 | 36.25 | 32.27 | 43.87 | 38 | 8 | 85 | 131 |
| *H. violaceum* | Wild, perennial | 137032 | 81155 | 12731 | 21573 | 38.27 | 36.24 | 32.26 | 43.87 | 38 | 8 | 85 | 131 |
| *H. jubatum* | Wild, perennial | 136826 | 80901 | 12665 | 21630 | 38.24 | 36.19 | 32.32 | 43.81 | 39 | 8 | 82 | 129 |
| *H. vulgare* subsp. *spontaneum* | Wild, annual | 136536 | 80612 | 12778 | 21573 | 38.30 | 36.30 | 32.25 | 43.84 | 39 | 8 | 83 | 130 |
| *H. vulgare* | Cultivate, annual | 136462 | 81671 | 12701 | 21045 | 38.32 | 36.31 | 32.33 | 43.83 | 48 | 8 | 83 | 139 |

10.3389/fpls.2023.1170004

TABLE 2 List of genes annotated in the plastomes of the three wild perennial *Hordeum* species (*H. bogdanii, H. brevisubulatum*, and *H. violaceum*) from Central Asia and two annual species (*H. vulgare* subsp. *spontaneum* and *H. vulgare*).

| Category | Function | Name of gene | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Self-replication (34) | Ribosomal RNA genes | **rrn4.5** | **rrn5** | **rrn16** | **rrn23** | | | | |
| | Transfer RNA genes | **trnR-ACG** | **trnL-CAA** | **trnV-GAC** | **trnH-GUG** | **trnN-GUU** | **trnA-UGC*** | **trnT-CGU*/bbv** | trnS-CGA*/bbv |
| | | **trnM-CAU** | **trnI-GAU*/$^{vul}$** | trnG-UCC*/vul | trnK-UUU* | trnL-UAA* | trnV-UAC* | trnC-GCA | trnG-GCC |
| | | trnS-GCU | trnS-GGA | trnT-GGU | trnY-GUA | trnD-GUC | trnL-UAG | trnR-UCU | trnS-UGA |
| | | trnP-UGG | trnT-UGU | trnE-UUC | trnQ-UUG | trnF-GAA | trnW-CCA | | |
| Ribosomal proteins (12) | Small subunit of ribosome (SSU) | rps2 | rps3 | rps4 | **rps7** | rps8 | rps11 | **rps12$^{vul}$** | rps14 |
| | | **rps15** | rps16*/vul | rps18 | **rps19** | | | | |
| Transcription (14) | Large subunit of ribosome (LSU) | **rpl2*** | rpl14 | rpl16 | rpl20 | rpl22 | **rpl23** | rpl32 | rpl33 |
| | | rpl36 | | | | | | | |
| | RNA polymerase subunits | rpoA | rpoB | rpoC1 | rpoC2 | | | | |
| | Translation initiation factor | infA | | | | | | | |
| Photosynthesis related genes (46) | RubisCO large subunit | rbcL | | | | | | | |
| | Subunits of photosystem I | psaA | psaB | psaC | psaI | psaJ | | | |
| | Subunits of photosystem II | psbA | psbB | psbC | psbD | psbE | psbF | psbH | psbI |
| | | psbJ | psbK | psbL | psbM | psbN | psbT | psbZ | |
| | Subunits of ATP synthase | atpA | atpB | atpE | atpF* | atpH | atpI | | |
| | Cytochrome b/f complex | petA | petB | petD | petG | petL | petN | | |
| | C-type cytochrome synthesis gene | ccsA | | | | | | | |
| | Subunits of NADH dehydrogenase | ndhA* | **ndhB*** | ndhC | ndhD | ndhE | ndhF | ndhG | ndhH |
| | | ndhI | ndhJ | ndhK | | | | | |
| Other genes (6) | Maturase | matK | | | | | | | |
| | Protease | clpP | | | | | | | |
| | Chloroplast envelope membrane protein | cemA | | | | | | | |
| | Hypothetical open reading frames | ycf1 | **ycf2** | ycf3$^{\#/bvv}$ | ycf4$^{bvv}$ | | | | |
| Unknown function (1) | Plant protein of unknown function | **ycf15$^{bbb}$** | | | | | | | |

*, gene containing a single intron; #, gene containing two introns; Genes in bold correspond to genes that are located in the IRs and hence are duplicated; bbv, genes that are particular for *H. bogdanii, H. brevisubulatum* and *H. violaceum*; vul, genes that are particular for *H. vulgare* subsp. spontaneum, and *H. vulgare*; bvv, genes that are particular for *H. brevisubulatum*, *H. violaceum*, *H. vulgare* subsp. *spontaneum*, and *H. vulgare*.

the future. The mononucleotide T was repeated 13 times and was unique to *H. brevisubulatum* and *H. violaceum* (Figure 2D). Furthermore, the majority of SSR were distributed in the LSC region, of which the proportion of *H. bogdanii* was 75.6%, slightly lower than that of *H. brevisubulatum* (76%) and *H. violaceum* (76%) (Figure 2E).

## 3.3 Relative synonymous codon usage and PR2-plot analysis

RSCU, which is caused by the unequal usage of a synonymous codon, was further analyzed (Figure 3). Each amino acid corresponds to at least one codon and at most six codons owing to the redundancy

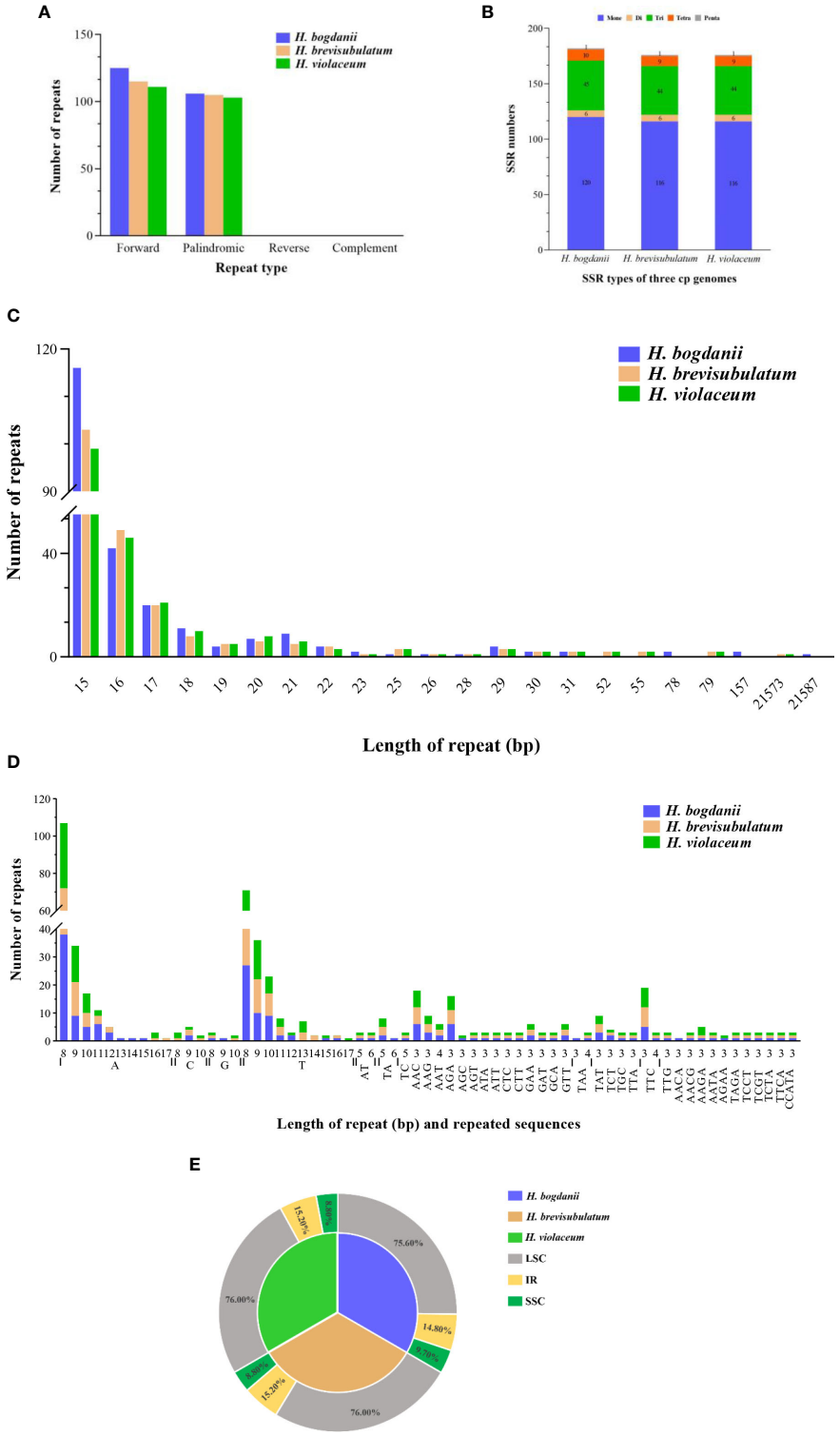



FIGURE 2
Simple sequence repeats (SSRs) and scattered repetitive sequences in the three *Hordeum* cp genomes. **(A)** frequency of repeat types; **(B)** compare of the number of SSR type in the three *Hordeum* cp genomes; **(C)** frequency of repeats length; **(D)** motifs in the cp genome of *Hordeum*; **(E)** Distribution region of repeating sequences of three *Hordeum* cp genome. IR, inverted repeat; LSC, large single-copy; SSC, small single-copy.

of codons. RSCU values for the initial codon (AUG) were 1.987, 1.983, and 1.987 in *H. bogdanii*, *H. brevisubulatum*, and *H. violaceum*, respectively. RSCU values for termination codons, UAA, UAG, and UGA, were 1.771, 0.651, and 0.578 in *H. bogdanii*, 1.730, 0.671, and 0.600 in *H. brevisubulatum*, and 1.730, 0.671, and 0.600 in *H. violaceum*, respectively. Codons with RSCU values >1, which are usually considered to be preferred codons, accounted for 51.61% (32/62) of codons, and the third nucleotide of most codons was biased

10.3389/fpls.2023.1170004

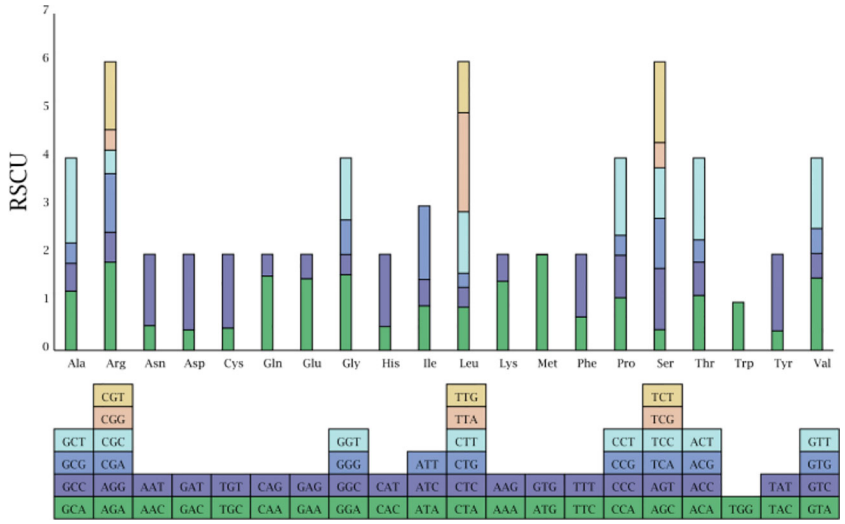



FIGURE 3
Relative frequency of synonymous codon for the twenty amino acids in the three *Hordeum* species chloroplast genomes.

towards either A or U. Notably, only one codon, UGG (corresponding to tryptophan), showed no bias in the three *Hordeum* species, and its RSCU was 1.00.

Forty-four coding sequences (CDS, ≥300 bp long) containing start (ATG) and stop (TAG, TGA, TAA) codons were collected from the 10 cp genomes, to carry out PR2-plot analysis to further understand codon bias (Figure 4). The results showed that the 44 genes of the 10 species were not evenly distributed within the four regions, but mainly in $G_3/(G_3+C_3) > 0.5$ and $A_3/(A_3+T_3) < 0.5$ regions. This suggests that there may be a bias towards G and T bases at the third position of synonymous codons, which needs further investigation.

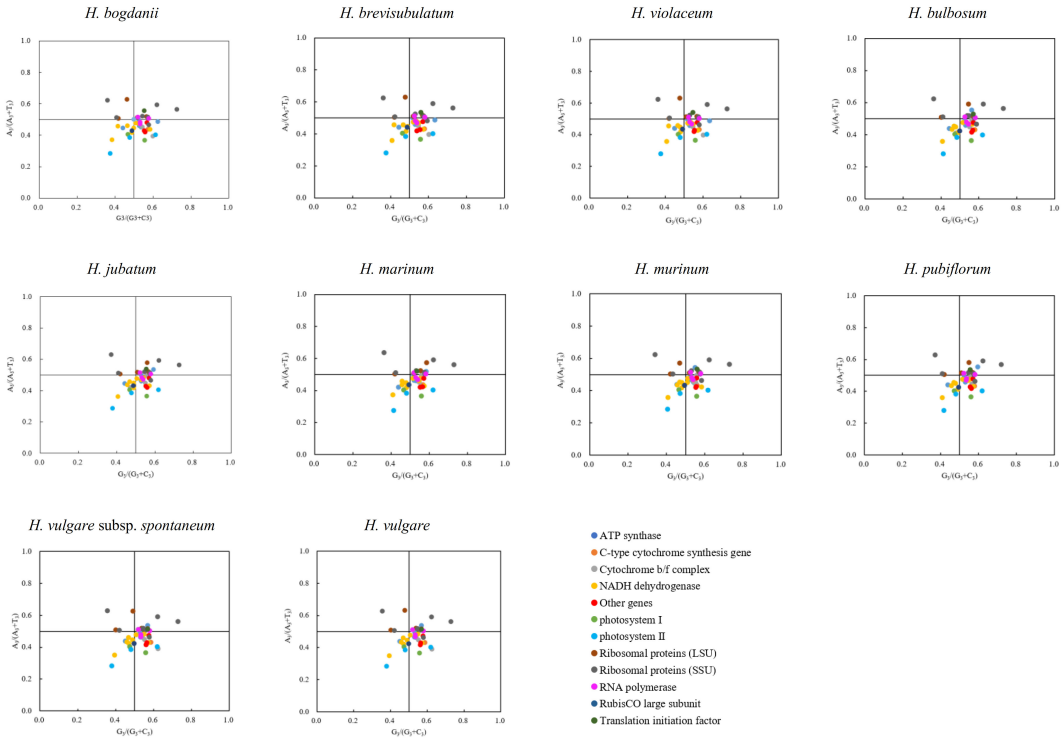



FIGURE 4
PR2-plot analysis of cp genomes ten *Hordeum* species. Base A, T, C and G content at the third site of synonymous codons were replaced through $A_3$, $T_3$, $C_3$ and $G_3$, respectively.

## 3.4 In/Dels and SNPs

InDels and SNPs (mainly containing Tn (transition) and Tv (transversion)) were detected among the three *Hordeum* cp genomes using MAFFT software (Standley, 2013). A total of 109, 112, and 33 In/Dels were identified in *H. bogdanii* vs *H. brevisubulatum*, *H. bogdanii* vs *H. violaceum*, and *H. brevisubulatum* vs. *H. violaceum*, respectively, in which 4 InDels were discovered in the coding sequence (Supplementary Table 4). There were similar quantities of Tn and Tv in both *H. bogdanii* vs *H. brevisubulatum* (Tn = 61, Tv = 304) and *H. bogdanii* vs *H. violaceum* (Tn = 66, Tv = 298), most of which were encoded in the noncoding sequence. However, 19 Tn (2 coding, 17 noncoding) and 60 Tv (23 coding, 37 noncoding) were detected during *H. brevisubulatum* vs *H. violaceum*. Interestingly, we found that both InDels and SNPs were mainly concentrated in LSC and the intergenic region for each pairwise comparison, while InDels did not occurred in the IR region of *H. brevisubulatum* vs *H. violaceum* (Figure 5).

The non-synonymous/Synonymous mutation ratio (Ka/Ks) ratio of 83 common protein-coding genes in cp genomes of the three *Hordeum* species was calculated using Ka/Ks Calculator software (Zhang et al., 2006) (Supplementary Table 3). Ka/Ks values of *H. bogdanii* vs *H. brevisubulatum*, *H. bogdanii* vs *H. violaceum*, and *H. brevisubulatum* vs *H. violaceum* were 16, 19, and 2, respectively. In addition, the Ka/Ks values of some genes (*ropB*, *atpI*, *psaB*, etc.) could not be computed because Ka or/and Ka was 0, which suggests that these genes were relatively conservative without any Ka or Ks nucleotide substitution. Pi values were calculated using VCFTOOLS software. A total of 101 common genes in the three wild perennial *Hordeum* species were examined, whose Pi values ranged between 0 to 0.1674 (Figure 6). It is noteworthy that relatively higher Pi values (Pi ≥ 0.1) were detected in five genes, including *tRNA-UGC*, *tRNA-UAA*, *tRNA-UUU*, *tRNA-UAC*, and *ndhA*. Meanwhile, these genes were also among those with Ka/Ks > 1. Moreover, other genes with a Pi ≥ 0.1 were found in single-copy (SC) rather than IR regions, except for *tRNA-UGC*.

## 3.5 Whole Cp genomes comparison with ten *Hordeum* species

To understand the sequence divergence between wild and cultivated, as well as annual and perennial species in genus *Hordeum*, and elaborate further on the evolutionary events that occurred, including gene mutation, rearrangement and loss, we analyzed and compared the cp genomes of two annual species (one cultivated species, *H. vulgare* and one wild species, *H. vulgare* subsp. *spontaneum*), and eight perennial wild species (*H. bogdanii, H. brevisubulatum, H. violaceum, H. bulbosum, H. jubatum, H. marinum, H. murinum*, and *H. pubiflorum*) were compared and analyzed. It was found that the coding region is more conservative than the non-coding region, as well as the divergence frequency was higher in the LSC and SSC region than in IR region (Figure 7). The two annual species (especially *H. vulgare*) had many conserved regions compared with the other eight wild perennial species, this was the case in the CNS (Conserved Noncoding Sequences) of LSC and SSC regions. The highly variable regions are called hotspots regions, and these regions were mainly concentrated in small RNA molecules such as *tRNA-GGU ~ tRNA-GCA*, *tRNA-UGU ~ ndhJ*, *psbE ~ rps18*, *ndhF ~ tRNA-UAG*. Furthermore, MAUVE software revealed rearrangement events with scanty genes in the cp genomes of 10 species (Supplementary Figure 1).

## 3.6 IR expansion and contraction

Expansion and contraction of IR regions, recognized as an evolutionary event, are generally concentrated in the junction of IR/SSC or IR/LSC. Moreover, this phenomenon is the primary cause of the variation of cp genomes size. Therefore, the IR borders of six species in the *Hordeum* genus were compared to explore their differences. The species studied included two annuals (including one cultivated species, *H. vulgare* and one wild species, *H. vulgare* subsp. *spontaneum*), and four perennial wild species (*H. bogdanii, H. brevisubulatum, H.*

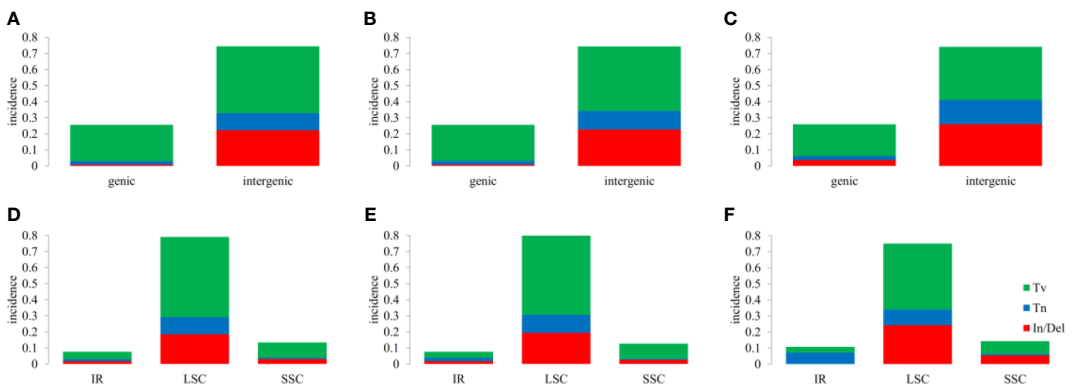



FIGURE 5
Overview of single nucleotide polymorphisms (SNPs) and Insertions/Deletions (In/Dels). **(A, B)**, **(C, D)**, and **(E, F)** the differences between *H. bogdanii* vs *H. brevisubulatum*, *H. bogdanii* vs *H. violaceum* and *Hordeum brevisubulatum* vs *Hordeum. violaceum*. Tv, transversion; Tn, transition; In/Del, insertion/deletion; IR, inverted repeat; LSC, large single-copy; SSC, small single-copy.

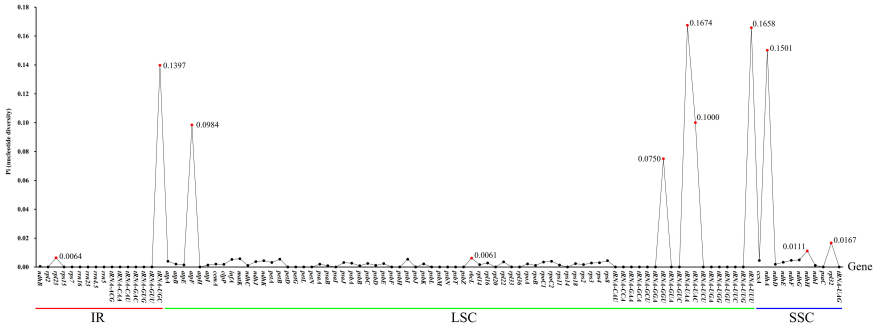



FIGURE 6
The nucleotide diversity (Pi) calculated by 101 genes shared in three wild perennial *Hordeum* species. Genes with Ka/Ks value > 1 are highlighted in red; The genes above the red line, green line and blue line were located in IR, LSC and SSR regions, respectively.

*violaceum*, and *H. jubatum*) (Figure 8). The results showed significant differences in the junction sites between the annual and perennial species. The genes *ndhF-ndhH* and *rpl2-trnH-psbA-rpl22-rps19* were found close in SSC/IR and LSC/IR boundaries, respectively. The *ndhH* genes of the other five species ranged from 207 (*H. bogdanii*, *H. brevisubulatum*, *H. violaceum*) to 216 (*H. vulgare*) bp in IRa region throughout the SSC/IRa junction, with the exception of *H. vulgare* subsp. *spontaneum*. Two genes, *trnH* and *rpl2*, were found near the junction of LSC/IR region in *H. vulgare*, whereas the genes around this junction region of the other five species were *rpl22* and *rps19* genes. Additionally, we observed that only *e ndhH* gene for *H. vulgar* was separated from SSC/IRb boundary with 1 bp.

## 3.7 Phylogenetic relationships

The phylogenetic position of Triticeae was identified based on the cp genomes sequences of three studied *Hordeum* species and


FIGURE 7
Alignment of the ten *Hordeum* species cp genome sequences. Exon, untranslated region (UTR), conserved noncoding sequences (CNS), and mRNA were marked by different colors. The x-axis and level a clinic columnar strip express the paratactic and sequences stability in the cp genome and the peaks represent hotspot regions.

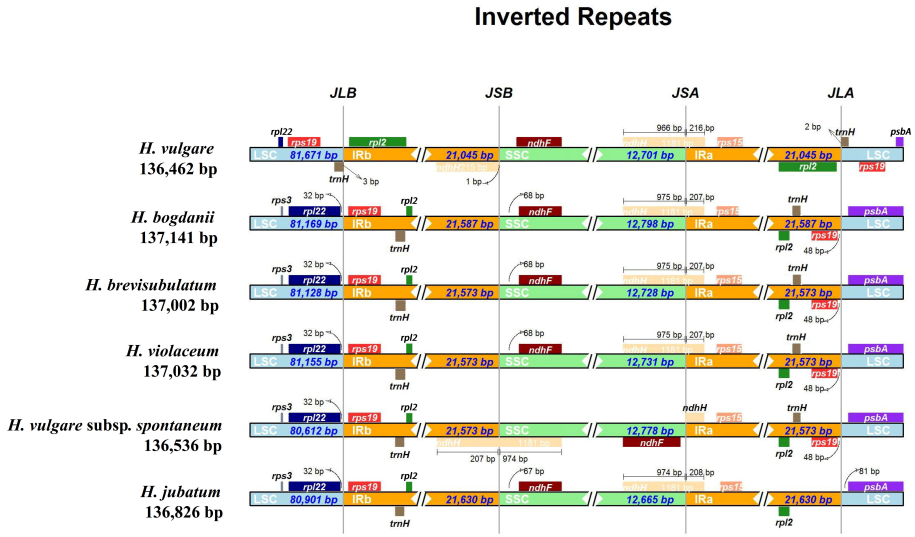



FIGURE 8
IRscope analysis of the six *Hordeum* cp genomes. JLB, JSB, JSA, and JLA represent the junction of LCS and IRb, SSC and IRb, SSC and IRa, and LSC and IRa region, respectively.

other 28 species downloaded from NCBI (Figure 9). The structure of this phylogenetic tree of these species conformed with the classical botanical classification. Twelve *Hordeum* species were divided into six sub-groups, among which *H. brevisubulatum*, and *H. violaceum* were in the same sub-groups, and *H. bogdanii* is further distant from them. Different accessions of the same species are placed in the same subgroup. In addition, genus *Hordeum* was more closely related to the species of *Elymus*, *Aegilops*, *Triticum* than to *Agropyron*.

# 4 Discussion

## 4.1 Characteristics of Cp genomes of *Hordeum* species

The total size and GC content of cp genomes were not significantly different among the three wild perennial *Hordeum* species (*H. bogdanii*, *H. brevisubulatum*, and *H. violaceum*). These results revealed that the cp genome size and GC content of Poaceae

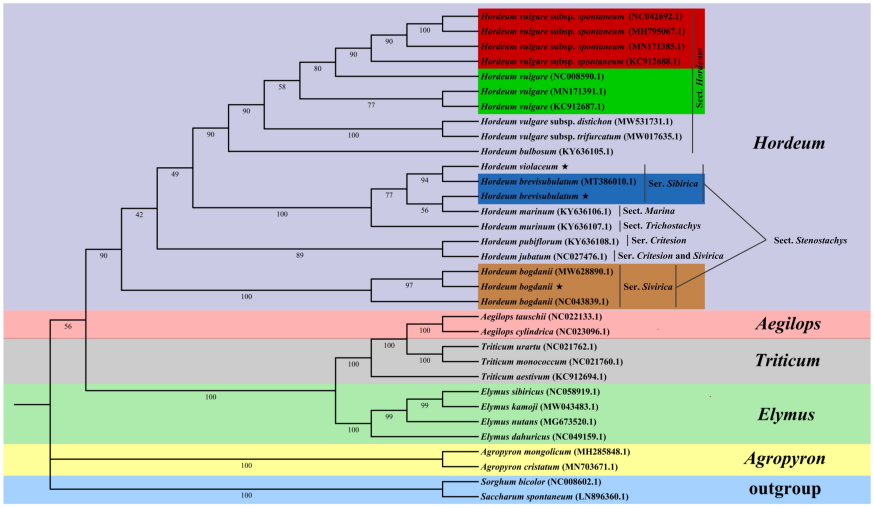



FIGURE 9
ML phylogenetic tree of 31 Poaceae species, with *Saccharum spontaneum* and *Sorghum bicolor* as outgroups. The bootstrap values are shown at the nodes; *H. vulgare* subsp. *spontaneum*, *H. vulgare*, *H. brevisubulatum*, and *H. bogdanii* species of different accession were represented by the base color of red, green, blue, and orange, respectively.

are highly conserved, and the occurrence of variation may help us to better understand the unique variation among species or subspecies (Liu et al., 2019). A total of 129, 131, and 131 genes were detected in the cp genomes of *H. bogdanii*, *H. brevisubulatum*, and *H. violaceum*, respectively. Notably, two mRNA genes, *ycf3* and *ycf4*, which were detected in these transformants and have been shown to contribute to the unstable accumulation of photosystem I complexes in the thylakoid membranes (Boudreau et al., 1997), were not found in *H. bogdanii*. This may be because two genes were transferred from the cp genome of *H. bogdanii* to its nuclear genome during the evolution of the species (Xiong et al., 2020). Two transfer RNA genes (*trnG-UUC* and *trnI-GAU*) and two small subunit of ribosome genes (*rps12* and *rps16*) were found to be unique to only two annual *Hordeum* species, including one wild species (*H. vulgare* subsp. *spontaneum*) and one cultivated species (*H. vulgare*). However, the functions of these four genes require further validation in the future. Genes specific for cultivated species (*H. vulgare*) in this study were not identified. This may be due to genetic changes may not exist in the cp genome but rather in the nuclear genome during plants domestication. Typically, cp genomes of Poaceae species are highly conserved in structure, which is a typical quadripartite (the IR region is separated by LSC and SSC). However, in some plants, cp genomes contain only one IR region (alfalfa) (Tao et al., 2016) or lack the IR region (algae) (Xue et al., 2019). *H. bulbosum*, *H. marinum*, *H. murinum*, and *H. pubiflorum* also fall into this category, with linear cp genomes without the IR region (Bernhardt et al., 2017). Therefore, the cp genome characteristics of these four *Hordeum* species were not analyzed and compared in the current study. However, cp genome characteristics of only two annual species (*H. vulgare*, and *H. vulgare* subsp. *spontaneum*) and four perennial species (*H. bogdanii*, *H. brevisubulatum*, *H. violaceum*, and *H. jubatum*) were analyzed and compared. The result demonstrated that the size and GC content of cp genomes of the six *Hordeum* species ranged from 136,462 to 137,141 bp and 38.23% to 38.32%, respectively, indicating that the cp genome length and GC content of synanthropic species were not significantly different, while the number of genes (139) in cultivated species were more abundant compared with that in wild species. The reason may be that natural selection has led to an accelerated rate of gene loss in wild species (Vishwakarma et al., 2017). It is well known that gene degradation and even loss occur because the cp genome of angiosperms evolves relatively fast (Lei et al., 2016). Our study found no significant difference in the total number of genes among the five wild *Hordeum* species, which ranged from 129 to 131 (Table 1), which was significantly lower than that of *H. vulgare* (139), with a maximum gap of 10 genes and a minimum of 8 genes, such as *rps12*, *rps16*, etc. There is evidence that these genes have been lost in *Ulmus* (Zuo et al., 2017) and *Orchidaceae* (Jing et al., 2014).

Introns, which are located in the non-coding region, typically have higher mutation rates than exons, as their functions are often more restricted (Gan et al., 2018). Nevertheless, it is noteworthy that introns play a crucial role in regulating gene expression (Ma et al., 2016). Nine genes, including *atpF*, *ndhA*, *ndhB*, *tRNA-CGA*, *tRNA-CGU*, *tRNA-UAA*, *tRNA-UAC*, *tRNA-UGC*, and *tRNA-UUU*, are shared by the three wild perennial *Hordeum* species and contain only one intron, while one gene, *ycf3*, contains two introns, which is unique to *H. brevisubulatum* and *H. violaceum* (Supplementary Table 3). In addition, the *ycf3* gene in the cultivated *Hordeum* species contains two introns (Middleton et al., 2013). Therefore, we contemplated that the absence of *ycf3* gene introns in *H. bogdanii* is unusual. Previous research has suggested that a species that a lack of gene introns in a species may indicate that it has taken on additional functions in diverse areas such as protease, RNA polymerase, and ribosomal pathways (Hakobyan et al., 2021).

## 4.2 Repeat sequences, RSCU, and PR2-plot analysis

Cp SSR in population genetics is considered a valuable molecular marker owing to its traits of matrilineal inheritance and low recombination frequency; gene insertion or deletion is also frequent in Cp SSRs (Xiao et al., 2019; Zong et al., 2019). Scattered repetitive sequences (SRS) and SSR of three wild perennial *Hordeum* species were analyzed and compared in the present study. The total number of SRS and cpSSRs in *H. bogdanii*, *H. brevisubulatum*, and *H. violaceum* were 231, 220, 218 and 182, 176, 176, respectively. *H. bogdanii* showed significantly different results from other two species, possibly due to their relatively close phylogenetic relatedness. In addition, the results of the study of *Secale sylvestre* (Skuza et al., 2022) and *Spartina maritima* (Rousseau-Gueutin et al., 2015) suggested that related species usually have similar SSR loci. Remarkably, most of the SSRs of the three *Hordeum* species are mononucleotides repeats dominated by poly-A or poly-T. This SSR phenomenon has not only been reported in the cp genomes of the Poaceae family (*Phalaris arundinacea* and *P. aquatica*) (Xiong et al., 2020) but also in other angiosperm families, such as *Hibiscus rosa-sinensis* (Abdullah et al., 2020), *Firmiana* (Abdullah et al., 2019), and *Taenia* (Yang et al., 2014).

During the translation of mRNA into proteins, there is an uneven frequency of synonymous codon usage called RSCU (Tyagi et al., 2020). In our study, 90.62% of codons with RSCU > 1 preferentially select A/U as the third nucleotide site, which is much higher than those ending with G/C, with similar results in many angiosperms such as *Nicotiana otophora* (Asaf et al., 2016), *Oryza minuta* (Sajjad et al., 2017), and *Medicago sativa* (Tao et al., 2016). The preference for A/U-ending codons is a common feature among most angiosperms and may be associated with certain evolutionary processes (Wang et al., 2023). PR2-plot analysis is essential for exploring codon bias. If the values of $G_3/(G_3+C_3)$ and $A_3/(A_3+T_3)$ are equal to 1, codon bias is completely influenced due to base mutation pressure; if both values are equal to 0, it is entirely because of natural selection (Wen et al., 2016). The majority of genes in our study had $G_3/(G_3+C_3)$ values greater than 0.5 and $A_3/(A_3+T_3)$ values lower than 0.5, indicating a bias towards G and T nucleotides in the third codon position, possibly due to a combination of natural selection and base mutations (Chen et al., 2021).

## 4.3 Sequence divergence

In the process of natural mutation, the probability of point mutation (SNP) is normally greater than that of frameshift (In/Del) (Raes and Van de Peer, 2005). As previously stated, the results of the cp genomes of the three *Hordeum* demonstrated that most mutations supported this conclusion. Interestingly, these mutation sites were concentrated in the intergenic or LSC region. The number of SNPs and In/Dels was significantly higher between *H. bogdanii* vs *H. brevisubulatum* and *H. bogdanii* vs *H. violaceum* compared with *H. brevisubulatum* vs *H. violaceum*. The reason may be that *H. bogdanii* was phylogenetically more distant from *H. brevisubulatum* and *H. violaceum*. Notably, no In/Dels were detected in the IR regions of *H. brevisubulatum* vs *H. violaceum*, suggesting that IR regions were the most conservative in the four-part structure (LSC, SSC, and IRa/IRb) of the cp genome, which warrants further exploration (Ravi et al., 2008). Pi, which is one of the standards that estimate the degree of nucleotide sequence variation and provide greater insight into the genetic variation to reflect complex changeable selection pressures in species and population levels (Namgung et al., 2021). Five genes with relatively high Pi values (Pi ≥ 0.1) were identified in the cp genomes, including *tRNA-UGC*, *tRNA-UAA*, *tRNA-UUU*, *tRNA-UAC*, and *ndhA*. These mutation hotspots can serve as a basis for further development of barcode molecular markers and phylogenetic analysis of the genus *Hordeum*.

The cp genomes of the 10 *Hordeum* species were analyzed for sequence variant and collinearity of using mVISTA and MAUVE procedures, respectively. The results indicated that the cultivated species, *H. vulgare*, were relatively conservative compared with the other wild related species. The wild plants undergo rapid molecular evolution due to which they form hotspot regions more frequently that are mainly located in the non-coding region of the LSC (Peng et al., 2021). Similar observations have been reported with *Morella rubra* (Liu et al., 2017) and three *Cardiocrinum* species (Lu et al., 2016). Notably, a series of hotspots regions were discovered, which mainly concentrated on *tRNA-GGU* ~ *tRNA-GCA*, *tRNA-UGU* ~ *ndhJ*, *psbE* ~ *rps18*, *ndhF* ~ *tRNA-UAG*, etc. Repeated conversions of genes between IRa and IRb regions may be a key factor responsible for generating these hotspots (Park et al., 2019). Collinearity analysis is generally a crucial strategy to determine the degree of cp genome variation (Liu et al., 2018). Collinearity analysis demonstrated that no rearrangement was detected in the cp genomes of the ten *Hordeum* species. However, there were significant differences were observed based on the cp genomes size, genotype, and expansion or contraction of IR boundaries.

As plants continue to evolve, the IR boundary can expand or contract due to the insertion or deletion of certain genes in the IR or SC region, which are the main factors contributing to cp genome size variation (Li et al., 2020). Here, the junction sites of the IR/SC region of the six cp genomes were analyzed using an online IRSCOPE software. In addition to the two annual *Hordeum* species (*H. vulgare* and *H. spontaneum*), no significant gene expansion, contraction, or loss was detected in the LSC/IRs/SSC boundary of the remaining four wild perennial *Hordeum* species (*H. bogdanii*, *H. brevisubulatum*, *H. violaceum*, and *H. jubatum*). This could be related to the fact that annual species have a more rapid evolutionary rate compared to perennial species (Duchene and Bromham, 2013). The length of the SSC region of *H. vulgare* was relatively smaller, mainly because the *ndhH* gene spanned the SSC/IRa region with 966 bp, which was the smallest compared with the other four wild perennial *Hordeum* species, located in the SSC region. Furthermore, the sites of genes *trnH* and *rps19* of *H. vulgare* changed significantly compared with those of the other *Hordeum* species. Besides, the *rpl22* gene only existed in the LSC region of *H. vulgare*, suggesting that it was replicated. This phenomenon may be attributed to the continuous domestication of the cultivated species, *H. vulgare*, leading to genetic changes through natural selection (Suoi et al., 2016). Therefore, the variation of the IR boundary and can be useful for phylogenetic studies of *Hordeum* species.

## 4.4 Phylogenetic relationships

The cp genome is quite conservative in sequence and structure, and the homology of molecular characters is easier to determine, thus it is a useful tool for constructing plant phylogeny (Yang et al., 2022). We conducted a phylogenetic analysis based on 31 Poaceae species (28 have been published and cp genomes of 3 *Hordeum* species were sequenced in the current study), with *Saccharum spontaneum* and *Sorghum bicolor* as the outgroups. The result showed that *H. bogdanii* has a further distance from *H. brevisubulatum* and *H. violaceum*. However, Jonathan et al. (Jonathan and Blattner, 2015) established a phylogenetic tree of these three *Hordeum* species based on the nuclear single-copy genome sequence analysis and demonstrated that they are clustered into a group. There may be two possible reasons for this difference. The first that the maternal ancestor of *H. bogdanii* is quite different from that of *H. brevisubulatum* and *H. violaceum*, and therefore it is hard to determine owing to relatively few reports on their matrilineal inheritance information. Another reason is the difference between the selected outgroups. In addition, although *H. brevisubulatum* (MT386010.1) has been published, the sequenced *H. brevisubulatum* in this study cannot be grouped into an identical subgroup. This may be because the former is a diploid or hexaploidy, while the latter is a tetraploid (Jakob and Blattner, 2006). Our findings provide valuable information for further investigation of the evolution trends of the cp genome in *Hordeum* species.

# 5 Conclusions

In summary, we sequenced and annotated the cp genomes of three *Hordeum* species (*H. bogdanii, H. brevisubulatum*, and *H. violaceum*) that exhibit a typical quadripartite structure. We then compared them to the cp genomes of two annual species, including one cultivated species (*H. vulgare*) and one wild species (*H. vulgare* subsp. *spontaneum*), as well as other five wild *Hordeum* species have been previously published. The results demonstrated that the cp genome of *H. vulgare* was more conserved although it contains a greater number of genes. Two mRNA genes, *ycf3* and *ycf4*, were not identified in *H. bogdanii*, of which *ycf3* contains two introns. Genes *trnG-UUC*, *trnI-GAU*, *rps12*, and *rps16* that are specific to only two annual *Hordeum* (*H. vulgare*, and *H. vulgare* subsp. *spontaneum*) and may be closely

related to the regulation of *Hordeum* growth. Five highly polymorphic genes (*tRNA-UGC*, *tRNA-UAA*, *tRNA-UUU*, *tRNA-UAC*, and *ndhA*) and a series of hotspot regions, which mainly concentrated on *tRNA-GGU ~ tRNA-GCA*, *tRNA-UGU ~ ndhJ*, *psbE ~ rps18*, *ndhF ~ tRNA-UAG*, etc., were identified. These findings lay the foundation for further development of barcode molecular markers and phylogenetic analysis of *Hordeum* L. In addition, based on the phylogenetic tree analysis, *H. brevisubulatum* and *H. violaceum* were classified into the same group and were found to be relatively close phylogenetic relatives as compared with *H. bogdanii*. Finally, the present study highlights the degree of variation between wild and cultivated, as well as annual and perennial *Hordeum* species, providing insights into phylogenetic evolution and population genetics in the genus *Hordeum*.

## 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://db.cngb.org/, CNS0491101, https://db.cngb.org/, CNS0491102, https://db.cngb.org/, CNS0491103.

## Author contributions

SY and CN: Conceptualization, methodology, validation, formal analysis, investigation, data curation, writing – original draft, writing – review and editing, visualization. These authors contributed equally to this work and share the first authorship. YL and XM: Writing – review and editing, supervision, project administration, and funding acquisition. SJ, TL, JZ and JP: Investigation, resources, and writing – review and editing. WK and WL: Formal analysis, investigation, and data curation. YX and YLX: Methodology, software, validation, and formal analysis. XL and QY: Writing – review and editing. All authors contributed to the article and approved the submitted version.

## Funding

This study was supported by the National Natural Science Foundation of China (grant numbers 31570654), the Regional Innovation Cooperation Project of Sichuan Province (2022YFQ0076), the Project of Cooperation between Provincial School and Provincial College (2023YFSY0012). The Forage Position Expert Project of Sichuan Beef Cattle Innovation Team (SCCXTD-2020-13).

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1170004/full#supplementary-material

## References

Abdullah, S. I., Mehmood, F., Ali, Z., and Waheed, M. T. (2019). Comparative analyses of chloroplast genomes among three *Firmiana* species: Identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. *Plant Gene* 19, 100199. doi: 10.1016/j.plgene.2019.100199

Abdullah, S. I., Mehmood, F., Waseem, S., Mirza, B., Ahmed, I., and Waheed, M. T. (2020). Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots. *Genomics* 112, 581–591. doi: 10.1016/j.ygeno.2019.04.010

Asaf, S., Khan, A. L., Khan, A. R., Waqas, M., Kang, S. M., Khan, M. A., et al. (2016). Complete chloroplast genome of *Nicotiana otophora* and its comparison with related species. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00843

Alyr, M. H., Pallu, J., Sambou, A., Nguepjop, J. R., Seye, M., Tossim, H. A., et al. (2020). Fine-mapping of a wild genomic region involved in pod and seed size reduction on chromosome A07 in Peanut (*Arachis hypogaea* L.). *Genes* 11, 1402. doi: 10.3390/genes11121402

Bailey, M., Ivanauskaite, A., Grimmer, J., Akintewe, O., Payne, A. C., Etherington, R., et al. (2020). The *Arabidopsis* NOT4A E3 ligase promotes PGR3 expression and regulates chloroplast translation. *Nat. Commun.* 2020, 21998. doi: 10.1038/s41467-020-20506-4

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Bernhardt, N., Brassac, J., Kilian, B., and Blattner, F. R. (2017). Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evol. Biol.* 17, 141. doi: 10.1186/s12862-017-0989-9

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578–579. doi: 10.1093/bioinformatics/btq683

Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biol.* 13, 1–9. doi: 10.1186/gb-2012-13-6-r56

Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M., and Rochaix, J. D. (1997). The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J.* 16, 6095–6104. doi: 10.1093/emboj/16.20.6095

Brassac, J., and Blattner, F. R. (2015). Species-level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by next-generation sequencing and *In Silico* cloning of multiple nuclear loci. *Syst. Biol.* 64, 792–808. doi: 10.1093/sysbio/syv035

Chen, S. Y., Zhang, H., Wang, X., Zhang, Y. H., Ruan, G. H., and Ma, J.. (2021). Analysis of codon usage bias in the chloroplast genome of *Helianthus annuus* J-01. *IOP Conf. Series: Earth Envir Sci.* 792, 12006–12009. doi: 10.1088/1755-1315/792/1/012009

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

Darling, A., Mau, B., Blattner, F. R., and Perna, A. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Dean, L., and Bjorn, C. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucl. Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152

Duchene, D., and Bromham, L. (2013). Rates of molecular evolution and diversification in plants: Chloroplast substitution rates correlate with species-richness in the Proteaceae. *BMC Evol. Biol.* 13, 65. doi: 10.1186/1471-2148-13-65

Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucl. Acids Res.* 39, 29–37. doi: 10.1093/nar/gkr367

Gan, K. A., Carrasco, P. S., Sewell, J. A., and Fuxman, B. J. I. (2018). Identification of single nucleotide non-coding driver mutations in cancer. *Front. Genet.* 9. doi: 10.3389/fgene.2018.00016

Gumeni, S., Evangelakou, Z., Gorgoulis, V., and Trougakos, I. (2017). Proteome stability as a key factor of genome integrity. *Int. J. Mol. Sci.* 18, 2036. doi: 10.3390/ijms18102036

Hakobyan, S., Loeffler-Wirth, H., Arakelyan, A., Binder, H., and Kunz, M. (2021). A transcriptome-wide isoform landscape of melanocytic nevi and primary melanomas identifies gene isoforms associated with malignancy. *Int. J. Mol. Sci.* 22, 7165. doi: 10.3390/ijms22137165

Hisano, H., Tsujimura, M., Yoshida, H., Terachi, T., and Sato, K. (2016). Mitochondrial genome sequences from wild and cultivated barley (*Hordeum vulgare*). *BMC Genomics* 17, 824. doi: 10.1186/s12864-016-3159-3

Jakob, S. S., and Blattner, F. R. (2006). A chloroplast genealogy of *Hordeum* (Poaceae): Long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Mol. Bio Evol.* 23, 1602–1612. doi: 10.1093/molbev/msl018

Jing, L., Hou, B. W., Niu, Z. T., Liu, W., Xue, Q. Y., and Ding, X. Y. (2014). Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PloS One* 9, e99016. doi: 10.1371/journal.pone.0099016

Jonathan, B., and Blattner, F. R. (2015). Species-level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by next-generation sequencing and in silico cloning of multiple nuclear loci. *Syst. Bio.* 64, 792–808. doi: 10.1093/sysbio/syv035

Kent, W., and Brumbaugh, H. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Bri Bioinf* 9, 299–306. doi: 10.1093/bib/bbn017

Lei, W., Ni, D., Wang, Y., Shao, J., and Liu, C. (2016). Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. *Sci. Rep.* 6, 21669. doi: 10.1038/srep21669

Li, D. M., Zhu, G. F., Xu, Y. C., Ye, Y. J., and Liu, J. M. (2020). Complete chloroplast genomes of three medicinal *Alpinia* species: genome organization, comparative analyses and phylogenetic relationships in family Zingiberaceae. *Plants* 9, 286. doi: 10.3390/plants9020286

Liu, L. X., Li, R., Worth, J. R. P., Li, X., Li, P., et al. (2017). The complete chloroplast genome of Chinese bayberry (Morella rubra, myricaceae): implications for understanding the evolution of fagales. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00968

Liu, X., Li, Y., Yang, H., and Zhou, B. (2018). Chloroplast genome of the folk medicine and vegetable plant *Talinum paniculatum* (Jacq.) *Gaertn.*: gene organization, comparative and phylogenetic analysis. *Molecules* 23, 857. doi: 10.3390/molecules23040857

Liu, H., Su, Z., Yu, S., Liu, J., and Li, B. (2019). Genome comparison reveals mutation hotspots in the chloroplast genome and phylogenetic relationships of *Ormosia* species. *BioMed. Res. Int.* 2019, 1–11. doi: 10.1155/2019/7265030

Lu, R. S., Pan, L., and Qiu, Y. X. (2016). The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: comparative genomic and phylogenetic analyses. *Front. Plant Sci.* 72054. doi: 10.3389/fpls.2016.02054

Ma, J. E., Lang, Q. Q., Qiu, F. F., Li, Z., Li, X. G., Luo, W., et al. (2016). Negative glucocorticoid response-like element from the first intron of the chicken growth hormone gene represses gene expression in the rat pituitary tumor cell line. *Int. J. Mol. Sci.* 17, 1863. doi: 10.3390/ijms17111863

Meng, J., Li, X., Li, H., Yang, J., Wang, H., and He, J. (2018). Comparative analysis of the complete chloroplast genomes of four *Aconitum* medicinal species. *Molecules* 23, 1015. doi: 10.3390/molecules23051015

Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T., et al. (2013). Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the Triticeae tribe. *PloS One* 9, e85761. doi: 10.1371/journal.pone.0085761

Namgung, J., Do, H. D. K., Kim, C., Choi, H. J., and Kim, J. H. (2021). Complete chloroplast genomes shed light on phylogenetic relationships, divergence time, and biogeography of Allioideae (Amaryllidaceae). *Sci. Rep.* 11, 1–3. doi: 10.1038/s41598-021-82692-5

Ogihara, Y., Isono, K., Kojima, T., Endo, A., Hanaoka, M., Shiina, T., et al. (2000). Chinese spring wheat (*Triticum aestivum* L.) chloroplast genome: Complete sequence and contig clones. *Plant Mol. Bio Rep.* 18, 243–253. doi: 10.1007/BF02823995

Park, I., Song, J. H., Yang, S., Kim, W. J., and Moon, B. C. (2019). *Cuscuta* species identification based on the morphology of reproductive organs and complete chloroplast genome sequences. *Int. J. Mol. Sci.* 20, 2726. doi: 10.3390/ijms20112726

Peng, J., Zhao, Y. L., Dong, M., Liu, S. Q., Hu, Z. Y., Zhong, X. F., et al. (2021). Exploring evolution characteristic between cultivated tea and its wild relatives using complete chloroplast genomes. *BMC Ecol. Evo.* 21, 71. doi: 10.1186/s12862-021-01800-1

Raes, J., and Van de Peer, Y. (2005). Functional divergence of proteins through frameshift mutations. *Trends Genet.* 21, 428–431. doi: 10.1016/j.tig.2005.05.013

Ravi, V., Khurana, J. P., Tyagi, A. K., and Khurana, P. (2008). An update on chloroplast genomes. *Plant Syst. Evol.* 271, 101–122. doi: 10.1007/s00606-007-0608-0

Reinert, S., Osthoff, A., Léon, J., and Naz, A. (2019). Population genetics revealed a new locus that underwent positive selection in barley. *Int. J. Mol. Sci.* 20, 202. doi: 10.3390/ijms20010202

Rousseau-Gueutin, M., Bellot, S., Martin, G. E., Boutte, J., Chelaifa, H., Lima, O., et al. (2015). The chloroplast genome of the hexaploid *Spartina maritima* (Poaceae, Chloridoideae): comparative analyses and molecular dating. *Mol. Phyl Evol.* 93, 5–16. doi: 10.1016/j.ympev.2015.06.013

Safonova, Y., Bankevich, A., and Pevzner, P. A. (2014). DipSPAdes: assembler for highly polymorphic diploid genomes. *Int. Conf. Res. Comput. Mol. Biol.* 6, 528–545. doi: 10.1089/cmb.2014.0153

Sajjad, A., Waqas, M., Khan, A. L., Khan, M. A., Kang, S. M., Imran, Q. M., et al. (2017). The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00304

Shen, X. F., Guo, S., Yin, Y., Zhang, J. J., Yin, X. M., Liang, Z. W., et al. (2018). Complete chloroplast genome sequence and phylogenetic analysis of *Aster tataricus*. *Molecules* 23, 2426. doi: 10.3390/molecules23102426

Skuza, L., Gastineau, R., and Sielska, A. (2022). The complete chloroplast genome of *Secale sylvestre* (Poaceae: Triticeae). *J. Appl. Genet.* 63, 115–117. doi: 10.1007/s13353-021-00656-x

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

Suoi, C. K., Gi, C. M., and Seonjoo, P. (2016). The complete chloroplast genome sequences of three *Veroniceae* species (Plantaginaceae): comparative analysis and highly divergent regions. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00355

Tao, X., Ma, L., Zhang, Z., Liu, W., and Liu, Z. (2016). Characterization of the complete chloroplast genome of alfalfa (*Medicago sativa*) (Leguminosae). *Gene Rep.* 6, 67–73. doi: 10.1016/j.genrep.2016.12.006

Tyagi, S., Jung, J. A., Kim, J. S., and Won, S. Y. (2020). Comparative analysis of the complete chloroplast genome of mainland *Aster spathulifolius* and other Aster species. *Plants* 9, 568. doi: 10.3390/plants9050568

Vishwakarma, M. K., Kale, S. M., Manda, S., Talari, N., Yaduru, S., Garg, V., et al. (2017). Genome-wide discovery and deployment of insertions and deletions markers provided greater insights on species, genomes, and sections relationships in the genus *Arachis*. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02064

Wang, Y. Z., Jiang, D. C., Guo, K. G., Zhao, L., Meng, F. F., Xiao, J. L., et al. (2023). Comparative analysis of codon usage patterns in chloroplast genomes of ten *Epimedium* species. *BMC Geno Data* 24, 3. doi: 10.1186/s12863-023-01104-x

Wei, L., He, J., Jia, X., Qi, Q., Liang, Z. S., Zheng, H., et al. (2014). Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evol. Biol.* 14, 1–12. doi: 10.1186/s12862-014-0262-4

Wen, Y., Zou, Z., Li, H., Xiang, Z., and He, N. (2016). Analysis of codon usage patterns in *Morus notabilis* based on genome and transcriptome data. *Genome* 60, 473–484. doi: 10.1139/gen-2016-0129

Wu, L. W., Nie, L. P., Wang, Q., Xu, Z. C., Wang, Y., He, C. N., et al. (2021). Comparative and phylogenetic analyses of the chloroplast genomes of species of Paeoniaceae. *Sci. Rep.* 11, 14643. doi: 10.1038/s41598-021-94137-0

Xiao, C. W., Liu, Y., Wei, Q., Ji, Q. A., Li, K., Pan, L. J., et al. (2019). Inhibitory effects of berberine hydrochloride on *Trichophyton mentagrophytes* and the underlying mechanisms. *Molecules* 24, 742. doi: 10.3390/molecules24040742

Xiong, Y. L., Xiong, Y., He, J., Yu, Q. Q., Zhao, J. M., Lei, X., et al. (2020b). The complete chloroplast genome of two important annual *Clover* species, *Trifolium alexandrinum* and *T. resupinatum*: genome structure, comparative analyses and phylogenetic relationships with relatives in Leguminosae. *Plants* 9, 478. doi: 10.3390/plants9040478

Xiong, Y., Xiong, Y. L., Jia, S. G., and Ma, X. (2020a). The complete chloroplast genome sequencing and comparative analysis of reed canary grass (*Phalaris arundinacea*) and Hardinggrass (*P. aquatica*). *Plants* 9, 748. doi: 10.3390/plants9060748

Xue, S., Shi, T., Luo, W., Ni, X., and Gao, Z. (2019). Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. Armeniaca*, and *P. salicina*. *Horticulture Res.* 6, 89. doi: 10.1038/s41438-019-0171-1

Yang, X., Luo, X., and Cai, X. (2014). Analysis of codon usage pattern in *Taenia saginata* based on a transcriptome dataset. *Par Vect* 7, 1–11. doi: 10.1186/s13071-014-0527-1

Yang, J. P., Zhang, F. W., Ge, Y. J., Yu, W. H., Xue, Q. Q., Wang, M. T., et al. (2022). Effects of geographic isolation on the *Bulbophyllum* chloroplast genomes. *BMC Plant Bio* 22, 1–14. doi: 10.1186/s12870-022-03592-y

Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, K. S., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genom. Prot. Bioinf* 4, 259–263. doi: 10.1016/S1672-0229(07)60007-2

Zong, D., Gan, P., Zhou, A., Li, J., and He, C. (2019). Comparative analysis of the complete chloroplast genomes of seven *Populus* species: insights into alternative female parents of *Populus tomentosa*. *PloS One* 14, e218455. doi: 10.1371/journal.pone.0218455

Zuo, L. H., Shang, A. Q., Zhang, S., Yu, X. Y., and Wang, J. M. (2017). The first complete chloroplast genome sequences of *Ulmus* species by *de novo* sequencing: genome comparative and taxonomic position analysis. *PloS One* 12, e171264. doi: 10.1371/journal.pone.0171264

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 July 2023
DOI 10.3389/fpls.2023.1170448

Check for updates

OPEN ACCESS

EDITED BY
Aliki Xanthopoulou,
Hellenic Agricultural Organisation (HAO), Greece

REVIEWED BY
Prabhakaran Soundararajan,
National Institute of Plant Genome Research (NIPGR), India
Victor Manuel Rodriguez,
Spanish National Research Council (CSIC), Spain
Wagner L. Araújo,
Universidade Federal De Viçosa, Brazil

*CORRESPONDENCE
R. Glen Uhrig
ruhrig@ualberta.ca

RECEIVED 23 February 2023
ACCEPTED 26 June 2023
PUBLISHED 28 July 2023

CITATION
Scandola S, Mehta D, Castillo B, Boyce N and Uhrig RG (2023) Systems-level proteomics and metabolomics reveals the diel molecular landscape of diverse kale cultivars. *Front. Plant Sci.* 14:1170448. doi: 10.3389/fpls.2023.1170448

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

# Systems-level proteomics and metabolomics reveals the diel molecular landscape of diverse kale cultivars

Sabine Scandola, Devang Mehta, Brigo Castillo, Nicholas Boyce and R. Glen Uhrig*

Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada

Kale is a group of diverse *Brassicaceae* species that are nutritious leafy greens consumed for their abundance of vitamins and micronutrients. Typified by their curly, serrated and/or wavy leaves, kale varieties have been primarily defined based on their leaf morphology and geographic origin, despite having complex genetic backgrounds. Kale is a very promising crop for vertical farming due to its high nutritional content; however, being a non-model organism, foundational, systems-level analyses of kale are lacking. Previous studies in kale have shown that time-of-day harvesting can affect its nutritional composition. Therefore, to gain a systems-level diel understanding of kale across its wide-ranging and diverse genetic landscape, we selected nine publicly available and commercially grown kale cultivars for growth under near-sunlight LED light conditions ideal for vertical farming. We then analyzed changes in morphology, growth and nutrition using a combination of plant phenotyping, proteomics and metabolomics. As the diel molecular activities of plants drive their daily growth and development, ultimately determining their productivity as a crop, we harvested kale leaf tissue at both end-of-day (ED) and end-of-night (EN) time-points for all molecular analyses. Our results reveal that diel proteome and metabolome signatures divide the selected kale cultivars into two groups defined by their amino acid and sugar content, along with significant proteome differences involving carbon and nitrogen metabolism, mRNA splicing, protein translation and light harvesting. Together, our multi-cultivar, multi-omic analysis provides new insights into the molecular underpinnings of the diel growth and development landscape of kale, advancing our fundamental understanding of this nutritious leafy green super-food for horticulture/vertical farming applications.

KEYWORDS

**kale, BoxCar DIA proteomics, GC-MS metabolomics, *Brassica oleracea*, diel plant biology**

## Introduction

*Brassica oleracea* and its diverse cultivar groups represent an important food crop for multiple populations across the globe. These include seven major cultivar groups: cauliflower, collard greens, broccoli, kohlrabi, cabbage, brussels sprouts and kale. Together, these crops represented 70.1 million metric tonnes of production in 2019 (https://www.fao.org/faostat/en/#data/QCL). Kale, which encompass several leafy *Brassicaceae* species (*B. oleracea* and *B. napus*) (Reda et al., 2021), is often referred to as a 'super-food' (Samec et al., 2019) as it is rich in numerous antioxidants (carotenoids, flavonoids, glucosinolates) and essentials vitamins (A, K and C), minerals (calcium and iron), dietary fibers (Becerra-Moreno et al., 2014) and low molecular weight carbohydrates (Megias-Perez et al., 2020). Additionally, kale has notable cultivation advantages, including a wide-ranging temperature tolerance that guarantees year-round availability in most climates (Samec et al., 2019). Given these characteristics, kale represents a horticultural crop with the potential to be a source of essential nutrients for multiple global populations (Migliozzi et al., 2015).

To maximize growth and to execute developmental programs, plants require the precise timing of diel (daily) events. Correspondingly, diel events are coordinated by a combination of circadian and light responsive mechanisms, which play a major role in modulating the plant cell environment at all molecular levels (Mehta et al., 2021). For example, in the model plant and related *Brassicaceae*, *Arabidopsis thaliana* (Arabidopsis), it is estimated that the circadian clock controls the diel expression of 1/3 of all genes (Covington et al., 2008). Further, the importance of the circadian clock and diel biology in plants is emphasized by its central role in governing critical agronomic traits such as biomass, flowering time and disease resistance (Creux and Harmer, 2019), suggesting diel biology should be a central facet of next-generation cropping systems (Hotta, 2021; Steed et al., 2021). This is particularly key for kale, which has shown that time-of-day harvesting can affect its nutritional composition (Casajus et al., 2021; Francisco and Rodriguez, 2021), indicating that time-of-day harvesting and postharvest storage is central to enhanced kale nutrition and shelf life when going to market (Reda et al., 2021). Currently, our understanding of kale rests at the production-level, with the diel molecular mechanisms underpinning unique and beneficial morphological and nutritional differences between kale cultivars remaining largely unknown (Megias-Perez et al., 2020).

To date, systemic molecular analyses of kale cultivars *convar. acephala* (Jin et al., 2018; Liu et al., 2020; Liu et al., 2021) var. *sabellica* (Pongrac et al., 2019) and *B. napus* var. *pabularia* (Chiu et al., 2018) have been limited, with no studies examining multiple cultivars or quantifying diel molecular changes. In one transcriptomic study of *B. oleracea convar. acephala* cultivars, glucosinolate, carotenoid and phenylpropanoid biosynthetic pathways were highlighted as critical in defining the differences between green 'manchoo collar' and red 'jeok seol' cultivars (Jeon et al., 2018). To date however, no investigations of *B. oleracea* var. *palmifolia* have been performed, limiting our molecular knowledge of these widely produced and consumed kale cultivars. This set of molecular studies has however demonstrated that kale is a highly dynamic and diverse set of species that requires a systemic, multi-omics investigation of multiple kale cultivars in order to elucidate diel molecular landscape features that can be harnessed for increased production and nutrition.

Light spectra, intensity and photoperiod have each been shown to be important in kale cultivation as modulators of the kale metabolome (Carvalho and Folta, 2014). Light is also essential for plant growth and development as it is a primary entrainment mechanism of the circadian rhythm of plants (Xu et al., 2022). With the circadian clock and diel plant cell regulation governing numerous agronomic traits of interest, including: flowering time, growth and plant defense (Steed et al., 2021), elucidating where changes in the molecular landscape of diverse kale genetics manifests through a diel/circadian lens using quantitative, systems-level omics technologies, represents a critical endeavour for optimizing the growth and nutrient content of kale grown in controlled growth environments.

Therefore, using a combination of gas-chromatography mass spectrometry (GC-MS) and the latest data independent acquisition (DIA) quantitative proteomics workflow called BoxCarDIA (Mehta et al., 2022), we establish the diel metabolome and proteome landscapes of nine widely available, commercial produced kale cultivars grown in controlled growth environments under a natural photoperiod. These growth conditions, combined with multi-omics analyses obtained at end-of-day (ED)/zeitgeber 11 (ZT11), and end-of-night (EN)/zeitgeber 23 (ZT23) time-points, provides critical new insights into how diverse kale cultivars coordinate diel molecular events, while simultaneously generating a proteome resource for further targeted experimentation. Through the development of this resource we create a foundation for uncovering the key proteins and cell processes underpinning critically important growth and development traits in kale, while also revealing the nutritional content and diel metabolic landscape of nine diverse kale cultivars.

## Results

### Diversity in kale growth and morphology under horticultural LED light conditions

To compare the growth of nine publicly available and commercially grown kale cultivars (Table 1), we utilized spectral LED lighting conditions that can be implemented in LED driven horticultural growth systems (Supplemental Figure 1A). In addition, we further implemented twilight conditions at ED and EN to mimic a more natural growth environment (Figure 1A; Supplemental Figure 1B). We employed these parameters in order to comparatively evaluate the growth of all nine cultivars in a controlled growth environment setting that represents conditions comparable to those in future-forward vertical farming and horticultural facilities (Figure 1A; Supplemental Figure 1B). These conditions successfully grew all nine kale cultivars, with each

10.3389/fpls.2023.1170448

TABLE 1 Name and classification of kale cultivars used in the study.

| Species | Convariety | Variety | Cultivars | Types | Paper Abbreviation |
|---|---|---|---|---|---|
| *Brassica oleracea* | *convar. Acephala* | *var. sabellica* | *cv.* Winterbor | Curly Kale | K2 |
| | | | *cv.* Dwarf curled Scotch | | K3 |
| | | | *cv.* Darkibor | | K7 |
| | | | *cv.* Starbor | | K8 |
| | | | *cv.* Scarlet | | K9 |
| | | *var. palmifolia* | *cv.* Lacinato | Italian Kale | K4 |
| | | *var. palmifolia x var. sabellica* (Lacinato x Redbor) | *cv.* Rainbow Lacinato | Hybrids | K6 |
| *Brassica napus (oleracea x rapa)* | *convar. Pabularia* | *var. pabularia* | *cv.* Red Russian (Note: possibility = Siberian Kale x Brassica nigra) | Russian and Siberian Kale | K5 |
| | | *var. pabularia* (Siberian x RedRussian) | *cv.* Red Ursa | Hybrids | K10 |

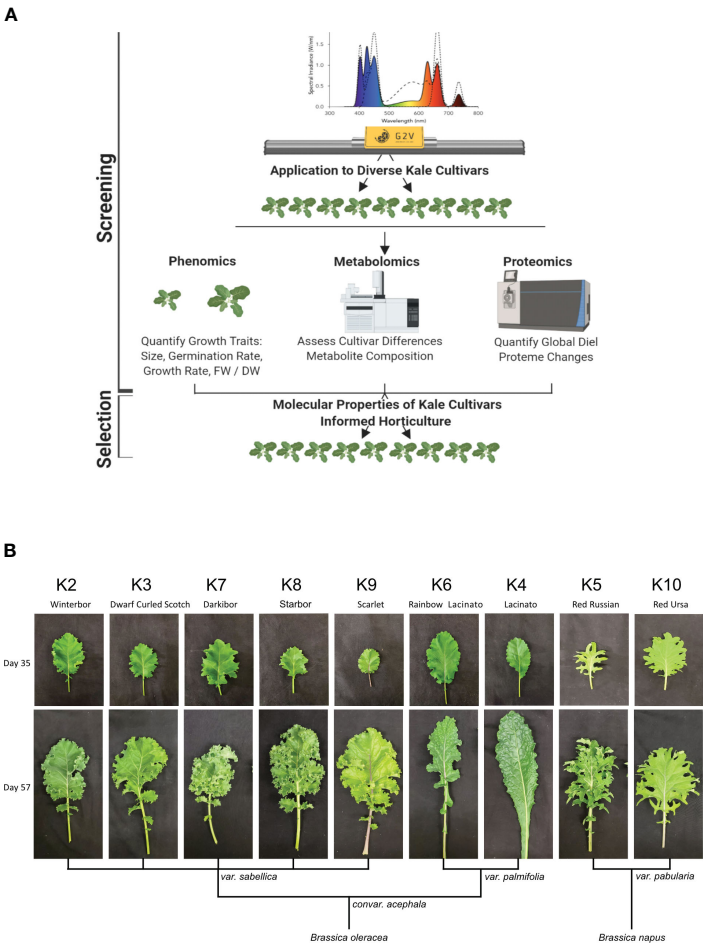



FIGURE 1
Experimental design and kale genetics phenotype. **(A)** Experimental workflow schematic. **(B)** Classification of kale cultivars based on leaf morphology and species.

cultivar presenting unique patterns of leaf size, shape and lobation consistent with previously observed morphologies. However, the leaf reddening typically observed in *B. napus* var. *pabularia* cultivars and the red *B. oleracea* var. *sabellica* 'Scarlet' (Figure 1B; Carvalho and Folta, 2014; Waterland et al., 2019) was not observed. The *B. oleracea* var. *sabellica* cultivars Starbor (K8), Darkibor (K7), Winterbor (K2), Dwarf Curled Scotch (K3) and Scarlet (K9) presented an oval shaped, curly leaf phenotype, while *B. napus* var. *pabularia* cultivars (Red Russian (K5) and Red Ursa (K10) have a notably pronounced, serrated leaf morphology. Alternatively, the Italian cultivars *B. oleracea* var. *palmifolia* (Lacinato and Rainbow Lacinato) present a spear-like leaf phenotype (long and narrow), with a rough surface and darker green coloration relative to the other kale cultivars (Figure 1B). These leaf traits along with leaf size are correlated with each cultivar's origins in Northern Europe, Russia and Italy, respectively (Table 1; Supplemental Figure 2). Lastly, we monitored leaf area over-time using a combination of time-course RGB imaging and PlantCV (https://plantcv.readthedocs.io/), which revealed kale cultivars to differ in their growth rates (Supplemental Figure 2). K10 and K2 varieties demonstrated the largest overall plant area, reaching an area of 13.4 to 12.6 $cm^2$ at 24 days post-imbibition, while K8 and K5 exhibited the slowest growth rate reaching both a plant area of 8 $cm^2$ at 24 days post-imbibition (Supplemental Figure 2). We also found that fresh weight (FW) is correlated with leaf area results, with K10 and K2 having the highest weight average of ~ 4 g and K8 and K5 the lowest, with an average of 1 g and 0.8 g, respectively.

To better characterize each cultivars physiological responses, we next measured relative chlorophyll content or Special Products Analysis Division (SPAD) and a variety of photosynthetic parameters (Phi2, PhiNO, PhiNPQ, LEF, ECSt, gH+ and vH+) using the PhotosynQ platform and the handheld MultispeQ device (Kuhlgert et al., 2016). Across the cultivars, relative chlorophyll amount was not significantly different except for the var. *palmifolia* cultivars K4 and K6, where SPAD was significantly higher (Supplemental Figure 2). This is consistent with the darker phenotype of the two var. *palmifolia* cultivars (Figure 1). Phi2, which measures photosystem II quantum yield, was only significantly higher in K5 compared to K10, while exhibiting no significant differences amongst other cultivars. Alternatively, PhiNO, which is a measurement of the electrons lost to non-regulated processes that can result in cellular damage, was significantly higher in K10, demonstrating that while K10 possesses one of the largest leaf areas, it is less effective at harvesting light energy. Interestingly, we find that linear electron flow (LEF), which estimates photosynthesis, exhibits a trend inversely proportional to FW, with large area cultivars possessing lower LEF and small cultivars possessing higher LEF. However, significant differences were only observed between cultivars K6 and K2.

Lastly, we estimated the energy generating capacity of each cultivar by measuring a series of parameters relating to ATP generation (Supplemental Figure 2). This included: ECSt, gH+ and vH+. ECSt, which describes the magnitude of the electrochromic shift, was higher in the K2 cultivar compared to K4, K8 and K9, suggesting better ATP production for increased growth outcomes while also aligning with their increased growth rate relative to other kale cultivars. Conversely, thylakoid proton conductivity (gH+), which describes steady state proton flux, was highest in the smallest cultivar K9, suggesting more efficient energy generation. However, despite differences in the ECSt and gH+, the initial rate of proton flux through ATP synthase (vH+) was constant across the cultivars. Taken together, these results indicate that photosynthetic parameters LEF and ECSt define important physiological differences in kale cultivars that likely contribute to observed differences in morphology and biomass.

## Diel metabolome analysis

As previous research in the related *Brassicaceae* Arabidopsis has demonstrated the importance of the ED and EN photoperiod transitions (Zeitgeber; ZT11-12 and ZT23-0; respectively) at the molecular-level (Uhrig et al., 2019; Krahmer et al., 2022), we next analyzed the metabolite content of each cultivar at ED (ZT11) and EN (ZT23). Aligning with other diel metabolite studies from the related species Arabidopsis (Flis et al., 2019; Cervela-Cardona et al., 2021) we quantified diel changes in 40 key metabolites across all nine cultivars, which can be grouped into 8 molecule classes: amino acids (12), organic acids (7), sugars (6), fatty acids (4), sterols (1), phenylpropanoid pathway metabolites (4) and vitamins (3) (Figure 2; Supplemental Figure 3; Supplemental Data 1). Hierarchical clustering using Euclidian distance further revealed that based on these 40 metabolites, the kale cultivars analyzed form 2 distinct groups based on their patterns of diel metabolite level fluctuations (Figure 2; Supplemental Data 1). This included a group consisting of

Darkibor (K7), Starbor (K8), Red Russian (K5) and Red Ursa (10), which form Group I and Winterbor (K2), Dwarf Curled Scotch (K3), Scarlet (K9), Rainbow Lacinato (K6) and Lacinato (K4), which form Group II. This grouping now allows us to establish additional kale relationships at the molecular-level (Supplemental Figure 3). Interestingly, Group I kale demonstrate more extensive leaf lobation and serration relative to their Group II counterparts, suggesting a correlation between leaf phenotype and diel metabolite changes. Further, hierarchical clustering analysis revealed two clusters of metabolites, whose relative change in diel abundance seem to be core to the differences between Group I and Group II (Figure 2). This includes: carbohydrates xylose, glucose and fructose; and amino acids aspartic acid, serine and iso-leucine, along with shikimate and α-linolenic acid (Figure 2). Of the compounds that kale produces in larger quantities that are of direct nutritional importance, we detected vitamins niacin (vitamin B3), ascorbic acid (vitamin C) and phytol (vitamin E precursor) in addition to α-linolenic acid (omega-3 fatty acid), with the majority of the kale cultivars possessing increased amounts of these vitamins at EN (Figure 2). Within Group II kale, we also see a sub-cluster consisting of K6 and K9 cultivars, which is largely defined by increased abundance of glyceric acid and glycine at ED.

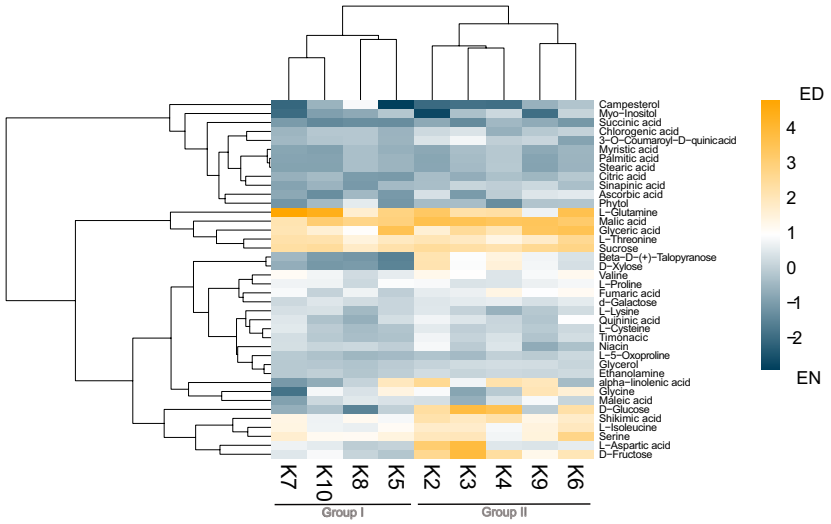



FIGURE 2
Diel changes in the kale metabolic landscape. Eucledian distance clustered heatmap of relative diel metabolite changes with each kale cultivar reveals two clusters of kale based on diel metabolic changes. Scale represents Log2 fold-change (FC); n=4 biological replicates.

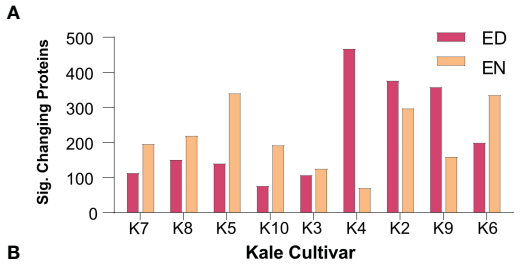
## Diel proteome analysis

To further contextualize Group I and Group II kale, we next performed quantitative proteomic analysis using an advanced data independent acquisition (DIA) workflow called BoxCarDIA which is aimed at better analyzing high complexity, high dynamic range plant samples (Mehta et al., 2022). With the kale analyzed in this study representing a diverse assemblage of cultivars without specifically sequenced genomes, we performed our quantitative proteomic searches using the *B. oleracea var oleracea* proteome. Here, we were able to quantify a total of 2124 protein groups across all nine cultivars (Supplemental Data 2). Of these, a total of 1734 protein groups exhibited a significant change in diel abundance (Bonferonni corrected *p-value* ≤ 0.05 and Log2FC ≥ 0.58) in at least one of the nine kale cultivars examined (Supplemental Data 2). Comparative quantification of the significantly changing proteins at ED and EN supported our metabolite-defined Group I and Group II clusters, with Group I kale possessing more proteins with a significant change in abundance at EN and Group II kale generally possessing more proteins changing at ED (Figure 3A).

Next, we analyzed all significantly changing proteins for enrichment of Gene Ontology (GO) terms relating to biological processes and molecular functions (BH corrected *p-value* ≤ 0.05). Here we found a significant enrichment of biological processes core to plant growth and development. These include significant enrichment of protein translation (GO0006412; GO:0006414), photosynthesis (GO:0009765; GO:0015979), cell redox homeostasis (GO:0045454), glycine metabolism (GO:0019464), lipid transport (GO:0006869) and primary metabolism (GO:0006096), amongst others (Figure 3B). Underpinning these biological processes was the significant enrichment of molecular functions related to translation initiation (GO:0003743) and elongation (GO:0003746), ribosome composition (GO:0003735), chlorophyll binding (GO:0016168) and oxidoreductase activity (GO:0016620), amongst other terms, which relate to protein translation, photosynthesis and cell redox homeostasis, respectively (Figure 3C).

To further elucidate when and where these protein-level changes differentially occur between Group I and II kale, and to increase our resolution of enriched biological processes, we performed an association network analysis using the knowledge database STRING-DB (https://string-db.org/). With *B. oleracea var oleracea* not possessing a STRING-DB dataset, we first identified orthologs from Arabidopsis for all significantly changing proteins using UniProt (https://www.uniprot.org/). Correspondingly, we identified orthologous Arabidopsis gene identifiers for 80.4% (1395/1734) of the significantly changing proteins originally quantified (Supplemental Data 3). Using a highly stringent STRING-DB score of ≥ 0.9, we then mapped an association network for Group I and Group II kale. This revealed diel abundance changes in proteins related to RNA splicing, both cytosolic and plastidial translation, chlorophyll biosynthesis, chaperones, mitochondrial respiration and elements of carbon metabolism, with Group I exhibiting specific changes in the proteasome, protein secretion, fatty acid biosynthesis and methionine metabolism, while Group II maintained specific changes in the phagosome (Figure 4). STRING-DB analyses were further contextualized by subcellular localization data to elucidate where ED and EN changes manifest within the subcellular landscape (Figure 5). Group I predominantly exhibited changes at EN relating to proteins localized to the plastid, cytosol, mitochondria and extracellular compartments, while Group II exhibited most changes in similar compartments (plastid, cytosol and mitochondria), but at ED (Figure 5).

Lastly, using orthologous Arabidopsis gene identifiers for the significantly changing proteomes of Group I and II kale, we performed a metabolic pathway enrichment analysis using the Plant Metabolic Network (PMN; https://plantcyc.org/;




B

| GO Term | Description | Fold Enrichment | *pvalue* | BH FDR |
|---|---|---|---|---|
| GO:0006412 | translation | 2.867301442 | 6.94050992 | 9.71E-20 |
| GO:0006457 | protein folding | 3.532226149 | 2.76203966 | 1.28E-09 |
| GO:0009735 | response to cytokinin | 5.833630563 | 1.13314448 | 2.60E-06 |
| GO:0015979 | photosynthesis | 3.554868624 | 1.27478754 | 6.26E-04 |
| GO:0042026 | protein refolding | 8.294693457 | 0.49575071 | 0.004121373 |
| GO:0006414 | translational elongation | 8.294693457 | 0.49575071 | 0.004121373 |
| GO:0009765 | photosynthesis, light harvesting | 5.078383749 | 0.7082153 | 0.004121373 |
| GO:0019464 | glycine decarboxylation via glycine cleavage system | 11.84956208 | 0.35410765 | 0.013029916 |
| GO:0018298 | protein-chromophore linkage | 4.308931666 | 0.7082153 | 0.013029916 |
| GO:0051017 | actin filament bundle assembly | 7.756076999 | 0.42492918 | 0.01946313 |
| GO:0045454 | cell redox homeostasis | 2.0313535 | 1.84135977 | 0.029462255 |
| GO:0006869 | lipid transport | 3.646019102 | 0.7082153 | 0.035279163 |
| GO:0009658 | chloroplast organization | 4.550231839 | 0.56657224 | 0.035279163 |
| GO:0006096 | glycolytic process | 3.763978544 | 0.63739377 | 0.051183229 |

C

| GO Term | Description | Fold Enrichment | *pvalue* | BH FDR |
|---|---|---|---|---|
| GO:0003735 | structural constituent of ribosome | 3.613022412 | 7.29461756 | 1.73E-28 |
| GO:0000166 | nucleotide binding | 3.376817465 | 4.67422096 | 4.15E-16 |
| GO:0003676 | nucleic acid binding | 2.303738703 | 5.31161473 | 1.78E-09 |
| GO:0003723 | RNA binding | 2.180256549 | 3.11614731 | 1.49E-04 |
| GO:0003743 | translation initiation factor activity | 3.972807771 | 1.20396601 | 2.17E-04 |
| GO:0003746 | translation elongation factor activity | 5.986176415 | 0.63739377 | 0.002948514 |
| GO:0016168 | chlorophyll binding | 5.240423798 | 0.7082153 | 0.002948514 |
| GO:0009055 | electron carrier activity | 2.391640223 | 1.84135977 | 0.002948514 |
| GO:0005509 | calcium ion binding | 2.064883407 | 2.26628895 | 0.005620431 |
| GO:0016620 | oxidoreductase activity | 6.371252091 | 0.49575071 | 0.014811511 |
| GO:0003755 | peptidyl-prolyl cis-trans isomerase activity | 4.446873908 | 0.63739377 | 0.018201474 |
| GO:0004332 | fructose-bisphosphate aldolase activity | 9.607443629 | 0.35410765 | 0.026338488 |

FIGURE 3
Diel changes in the kale proteome. Sampling diverse kale genetics provides a robust depiction of diel plant cell regulation in kale (n=4). **(A)** Number of proteins exhibiting diel changes in protein abundance within each kale cultivar examined at ED and EN. **(B)** Gene Ontology (GO) enrichment analysis of biological processes and **(C)** molecular function among all significantly changing proteins (Bonferonni corrected q-value ≤ 0.05; Log2FC ≥ 0.58 or ≤ -0.58).

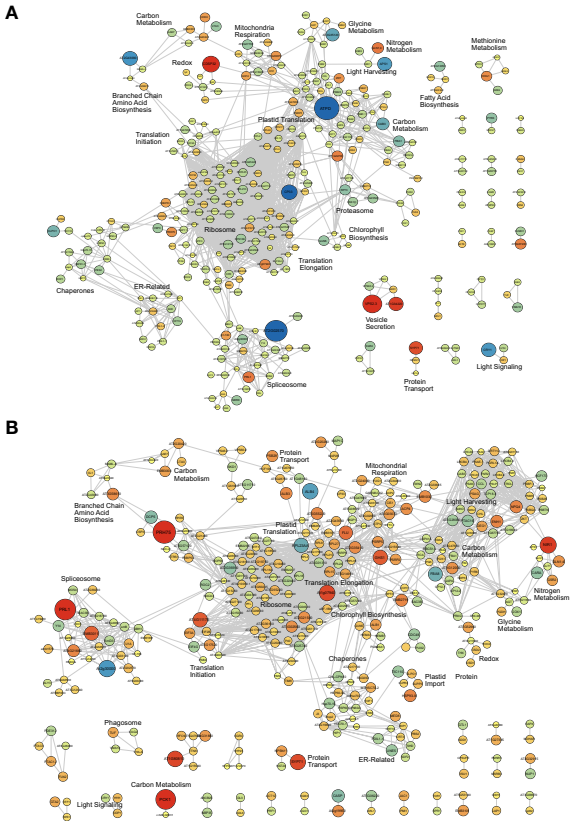



FIGURE 4
Association network analysis of the diel kale proteome. Proteins exhibiting significant diel changes in abundance within Group I **(A)** and Group II **(B)** kale, respectively, were subjected to association network analysis (https://string-db.org/). To maximize the analysis, Brassica oleracea var. oleracea gene identifiers were converted to Arabidopsis thaliana ortholog gene identifiers using UniProt (https://www.uniprot.org/). Proteins exhibiting significant changes in abundance are defined by a Bonferonni corrected q-value ≤ 0.05 and a Log2FC ≥ 0.58 or 5 -0.58). Up- (red) and down- (blue) regulated proteins are shown, with node size related to magnitude of Log2FC.

Supplemental Data 6). This revealed the enrichment of several pathways (Fisher's Exact Test < 0.01) directly related to the metabolites measured by our GC-MS analyses that differ between Group I and II kale, such as multiple pathways related to amino acid and sugar metabolism (Supplemental Data 6). In Group I kale we find an enrichment of serine biosynthesis, which is consistent with our metabolite findings (Figure 2). Specifically, we see serine pathway enzymes (e.g. SERINE HYDROXYMETHYLTRANSFERASE 1; SHM1) exhibiting an average Log2FC change in abundance of 1.45 at ED (Supplemental Data 2, 6). Amongst Group II kale, we see an enrichment of carbohydrate/sugar degradation, which aligns with our observed increase in glucose, fructose and xylose at ED. Here, we find significant increases in the abundance of STARCH-EXCESS 4 (SEX4; Log2FC = 0.68), which degrades starch to affect glucose-levels and FRUCTOKINASE-LIKE 2 (FLN2; Log2FC = 1.22), which phosphorylates fructose to create fructose 6-phosphate, at EN, likely contributing to increased pools of glucose and fructose during the day. We also see an increase of xylan degrading enzymes BETA-XYLOSIDASE 6 (BXL6; Avg Log2FC = 0.162) and ALPHA-XYLOSIDASE 1 (XYL1; Log2FC =0.70) that likely contribute the observed increase in Group II xylose at ED. Lastly, we also find an enrichment of multiple overlapping pathways between Group I and II kale, suggesting that while the timing of amino acid and core carbohydrate metabolism at our sampled ED and EN time-points may differ, there may be additional pathways that have group-specific diel changes at alternative time-points.

# Discussion

## Diverse kale cultivars form two groups based on phenotypic and metabolic signatures

As the functional components of all biological systems and the defining elements of nutrition, the proteome and metabolome represent a reliable means by which to elucidate differences between diverse, but related plant genetics. In the case of kale, it is the underlying molecular differences between cultivars that offer unique opportunities for targeted breeding and growth manipulation to enhance nutrition and biomass production. To

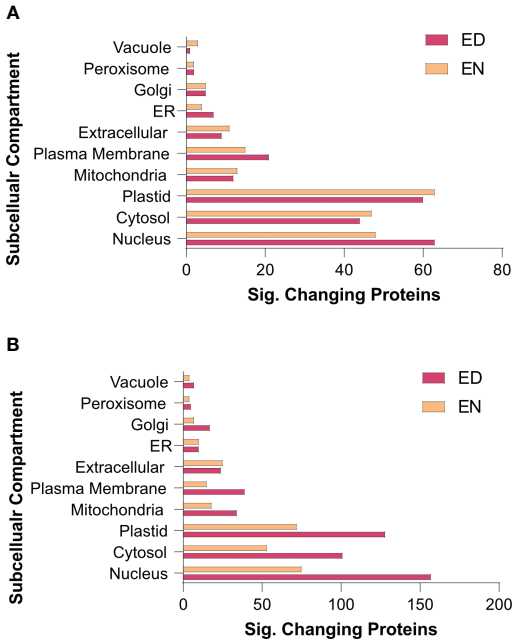



FIGURE 5
Subcellular localization analysis of the kale diel proteome. Subcellular localization analysis of Group I **(A)** and II **(B)** kale proteins exhibiting a significant change in diel abundance, respectively (https://suba.live/). To maximize this analysis, *B oleracea* var. *oleracea* gene identifiers were converted to *Arabidopsis* ortholog gene identifiers using UniProt (https://www.uniprot.org/). Proteins exhibiting significant changes in abundance are defined by a Bonferonni corrected q-value ≤ 0.05 and a Log2FC ≥ 0.58 or ≤ -0.58).

date, differences between kale cultivars has been largely defined phenotypically through leaf morphologies such as coloration, size, shape, lobation and serration (Arias et al., 2021). A few studies have complemented this with transcriptomic analyses of individual cultivars (Chiu et al., 2018; Jeon et al., 2018; Jin et al., 2018; Arias et al., 2021), while others have undertaken metabolomics analyses (Chiu et al., 2018; Jeon et al., 2018). This has predominantly involved targeted metabolomics, examining single kale cultivars for changes in pigmentation (Redbor; var. *sabellica*; (Klopsch et al., 2019), flavonols (Winterbor; var. *sabellica*; (Neugart et al., 2014; Neugart et al., 2016; Neugart and Bumke-Vogt, 2021) and fatty acids (Black Cabbage; *convar. acephala*; (Ayaz et al., 2006), with few studies having pursued global metabolite profiling of an individual kale cultivar (Nemzer et al., 2021). Correspondingly, our systems-level analysis of the leaf proteome and metabolome from nine kale cultivars in the same study using non-targeted GC-metabolomic and LC-proteomic mass spectrometry (MS), respectively, represents a substantial advancement in resolving the broader molecular landscape of kale for future targeted investigations.

Under our growth conditions no red/purple pigmentation in any kale cultivar was observed despite some of the cultivars examined (K9, K5 and K10) being known to have elevated anthocyanin production (Waterland et al., 2019). As a goal of our study was to define the diel molecular landscape of kale across diverse kale cultivars using near-sunlight conditions, our finding of no observable anthocyanin production was very informative (Figure 1; Supplemental Figure 2). It suggests that LED light recipes deployed in controlled growth environments can be utilized to drive substantially different growth outcomes in the same kale variety. This aligns with previous studies, which revealed the application of UV light enhances the profile of flavonoids in kale (Neugart et al., 2014; Neugart and Bumke-Vogt, 2021). Flavonoids are important molecule class in kale, as they represent the molecular precursors for the red/purple coloration some kale cultivars exhibit via anthocyanin (Liu et al., 2021).

Unexpectedly, our analysis of the diel metabolome landscape across nine kale cultivars revealed two groups based on their metabolite signatures. Group I is comprised of both *B. oleracea* (K7 and K8) and *B. napus* (K5 and K10) cultivars, despite notable differences in ploidy between *B. oleracea* (diploid) and *B. napus* (polyploidy) (Gao et al., 2022). This suggests that higher level differences in genomic architecture do not seem to determine baseline growth traits in kale. Group I kale also maintained similarities in leaf lobation architecture, exhibiting a jagged and pronounced lobation morphology relative to the kale cultivars of Group II. Conversely, Group II kale consisted entirely of *B. oleracea* cultivars, which have substantially different and more variable lobation morphologies. The var. *palmifolia* cultivars K6 and K4, have a narrow leaf shape and almost no leaf lobation, while the var. *sabellica* cultivars K2, K3 and K9 have a round and wavy leaf shape with more leaf lobation. No specific photosynthesis or ATP production measurements were found to correlate with these metabolite-based groupings, likely due to the higher order nature of those processes relative to diel metabolite changes; however, Group II kale did possess more light harvesting/photosynthesis and mitochondrial respiration proteins exhibiting a diel change in their abundance relative to Group I. Using un-targeted metabolomics to define groups within a complex species has also proven successful with other crops (*Cucumis melo*; (Moing et al., 2020). Here, a variable alignment between phylogeny and the metabolome was found that parallels our findings in kale. Taken together, our combined phenotypic and diel-metabolomic definition of nine kale cultivars suggests that metabolomic fingerprinting provides a more contextualized understanding of kale cultivars, however, future studies could consider more wide-ranging, untargeted LC-based metabolomics analyses. Overall, our findings provide an initial resource for future research of this leafy-green super food, in addition to offering actionable information for vertical farming and horticultural kale producers.

## Underlying metabolic differences between kale cultivars offers opportunities for production systems

Upon comparing the diel metabolite profiles of our nine kale cultivars, we found a series of core compounds that are critical pre-requisites for the production of nutritionally valued specialized metabolites. Further, many of these metabolomic differences seemed to define Group I and II kale. These compounds include: carbohydrates (xylose, glucose and fructose), amino acids (serine, glycine, aspartic acid and iso-leucine) as well as shikimic acid and

α-linolenic acid (Figure 2; Supplemental Data 1). This analysis successfully defined the molecular potential of each kale cultivar, while also providing information for time-of-day kale harvesting in order to maximize its nutritional content. With the circadian clock and diel plant cell regulation highlighted as a critical consideration for next-generation agriculture [e.g. chronoculture (Steed et al., 2021)], it is important that researchers and vertical farming/horticulture producers have reliable data resources generated using precision LED light systems.

*Carbohydrates* – Carbohydrates represent an important energy source for both plants and humans. Sugars, such as glucose and fructose, are a particularly important component of kale taste, which plays an important role in cultivar attractiveness, as kale is generally characterized as having a bitter and earthy flavor profile (Barker et al., 2022). Like other plants, kale possesses peak sucrose levels at ED, which in combination with transient leaf starch degradation, is vital for driving plant growth at night. At night, leaf starch is degraded to produce glucose, while sucrose is degraded to produce both glucose and fructose (Kim et al., 2017). Despite relatively consistent diel sucrose levels observed across both kale groups, we find a distinct diel pattern of glucose and fructose abundance between Group I and II kale (Figure 2; Supplemental Figure 3; Supplemental Data 1). Group I kale possess more fructose at EN, suggesting they may have better cold-temperature tolerance as fructose enhances cold-induced oxidative stress adaptation by preserving homeostasis under low temperatures (Bogdanovic et al., 2008). This diel difference could also be explained by cultivars originating from northern latitudes (e.g. K5 and K10). Further, fructose can be involved in vitamin C biosynthesis through its conversion to fructose-6-phosphate by fructokinase, which can also aid in cold tolerance (Akram et al., 2017), while also being an essential vitamin for human consumption. Diel changes in vitamin C leading to an accumulation at EN may offer to ease the transition to a high light environment come morning given its antioxidant properties (Paciolla et al., 2019). There is also a positive relationship between sucrose metabolism and anthocyanin production (Shi et al., 2014), which aligns with typically red cultivars K5 and K10 and their enhanced diel accumulation of fructose and glucose at EN. Alternatively, typically green cultivars K4 and K6 of Group II kale possess maximal fructose levels at ED, which aligns with our pathway enrichment analysis and peak abundance changes in key metabolic enzymes related to carbohydrate degradation (Supplemental Data 2, 6).

Additional carbohydrates xylose, myo-inositol and galactose were also present in different quantities across cultivars and throughout the day. Myo-inositol is involved in an array of biological processes such as being a precursor of inositol phosphates (Ips), hormones (auxin), translocation of mRNA into the cytosol, membrane biogenesis, light response germination and abiotic stress response (Munnik et al., 1998). Alternatively, xylose is a major component of the cell wall hemicellulose xylan, which provides plant cell resistance against enzymatic digestion and represents up to 35% of some wood compositions (Rennie and Scheller, 2014). Xylose is derived from UDP-Glucose and is transported to Golgi where xylan is produced, which aligns with the large number of secretion related proteins (e.g. ER, Golgi and extracellular) we see significantly changing in both Group I and II kale. It also specifically aligns with our pathway enrichment analysis for Group II kale, where we find increased abundance of xylan degrading enzymes that would increase the amount of xylose available during the day (Supplemental Data 2, 6). Xylose functions as a dietary fiber with prebiotic properties (Thavarajah et al., 2016). With Group I and II kale having contrasting diel xylose levels, the beneficial properties of xylose may offer value-added properties to a cultivar based on time-of-day harvesting.

*Amino acids* – Our analysis of the diel metabolome found that glycine, serine and aspartic acid have group specific differences in diel abundance. Both serine and glycine predominantly accumulated at EN and ED in Group I and II kale, respectively. Of these amino acids, serine in particular, functions as a key substrate for the biosynthesis of molecules critical for plant growth, including: amino acids glycine, methionine and cysteine (an essential amino acid for human nutrition), nitrogenous bases, proteins, phospholipids and sphingolipids (Stein and Granot, 2019). Specifically, our pathway analysis found an enrichment of serine biosynthesis in Group I kale, with our proteome data indicating that this occurs at ED (Supplemental Data 2, 6). Additionally, we find group specific differences in aspartic acid levels, which is also a precursor for several other amino acids, including: lysine, methionine, threonine and isoleucine (Han et al., 2021), with lysine representing an essential amino acid, whose content in plants is a key nutritional trait for crop improvement (Galili and Amir, 2013). Amongst the kale cultivars examined however, we see lysine consistently produced at EN while we see iso-leucine demonstrating group specific diel abundance changes. Iso-leucine is one of three branched chain amino acids (BCAA; e.g. leucine, isoleucine and valine), but is the only BCAA not built from pyruvate (Joshi et al., 2010). With the BCAA isoleucine an integral to plant defense as a conjugate of jasmonic acid (Armenta-Medina et al., 2021) and being up-regulated in response to drought and cold (Joshi et al., 2010), our data suggest that Group I and II kale may have differences in their stress mitigation capacity.

*Shikimic acid* – Shikimic acid is a precursor of the essential aromatic amino acids tyrosine, phenylalanine and tryptophan and therefore is an important precursor for the phenylpropanoid biosynthetic pathway, which is responsible for the production of a large number of nutritionally valued specialized metabolites (Dong and Lin, 2021). Production of shikimic acid can consume upwards of 30% of the fixed carbon, feeding the production of vitamins K1, B3 (folate), E (tocopherols), in addition to flavonoids, anthocyanins and lignin (Tohge et al., 2013). The shikimic acid pathway is also involved in the color patterning seen in ornamental kale through modulation of anthocyanin content, which is a key trait in kale as a source of antioxidants (Liu et al., 2021).

*α-linolenic acid* – Kale is known to be rich in α-linolenic acid, however, the differences in abundance between cultivars and the diel production landscape of α-linolenic acid have not previously been assessed. α-linolenic acid is a polyunsaturated Omega 3 fatty acid, which is an essential component of a healthy diet (Nemzer et al., 2021), as omega 3 fatty acids are known to decrease the risk of heart disease and lower the blood pressure (Shahidi and Ambigaipalan, 2018). Importantly, α-linolenic acid also forms the

basis of cell membrane components as a precursor of phosphoglycerolipids, cutin and waxes (He and Ding, 2020). With α-linolenic acid contributing broadly to many of the important nutritional properties of kale and our results demonstrating it maintains diel changes in abundance, along with differences in abundance between cultivars, α-linolenic acid offers an array of opportunities for further development of cultivar specific light recipes to maximize its production.

## Differences in the diel proteomes of Group I and II kale indicate they are defined by core elements of plant cell regulation

Elucidating the specific molecular components that underpin the observed metabolomic changes of Group I and II kale revealed a number of plant cell processes that are diel regulated at the protein-level. In particular, we find diel proteome changes across multiple subcellular compartments related to metabolism, RNA processing, protein translation and light harvesting. With diel biology/the circadian clock critical for timing the transitions from day-to-night and night-to-day, understanding where within the cellular environment [e.g. subcellular compartment(s)] molecular changes manifest, can help further resolve each kale cultivars potential agronomic viability from a chronoculture perspective (Steed et al., 2021). From a metabolic perspective, we find diel changes in: carbon, nitrogen, glycine, methionine, BCAA and fatty acid metabolism (Figure 2), which we also see enriched at the pathway level (Supplemental Data 6), while for RNA processing and protein translation, we find: mRNA splicing, cytosolic and plastidial protein translation along with a number of chaperones to exhibit significant diel changes. We also observed differences involving numerous light harvesting and signaling proteins, along with differences in chlorophyll biosynthetic enzymes, which may also be directly related to Group specific productivity differences. At the highest level, our quantitative proteomic analysis resolved Group I kale cultivars to possess significant changes in their diel proteome at EN, while Group II kale exhibit more significant proteome-level changes ED (Figure 3). We also note that many significantly changing metabolic proteins are directly connected to observed diel metabolome changes, reinforcing our multi-omics approach (Supplemental Data 6). Further, significant changes in numerous other, non-metabolic proteins central to proper plant cell regulation demonstrate how quantitative proteome analysis allows us to map diel molecular landscape changes in order to identify how kale breeders and producers can better realize each cultivars genetic potential.

*Metabolism* – Significant changes in the diel proteome related to plant metabolism were wide-ranging, aligning with the metabolite-determined kale groupings. Enzymes involved in carbon and nitrogen metabolism exhibiting a significant diel change in abundance demonstrated coherent time-of-day changes between both kale groups, consistent with their central roles in plant growth and development. Conversely, amino acid metabolic enzymes aligned with the metabolite-defined kale groups, specifically BCAA-related enzymes and glycine metabolic enzymes, likely relating to the specialized metabolites produced by kale that are derived from these amino acids.

From a carbon metabolism perspective, diel changes centered around two protein clusters comprised of mitochondrial enzymes isocitrate dehydrogenase (cICDH), ATP citrate lyase and components of the pyruvate dehydrogenase complex, along with key primary metabolic enzymes fructose-bisphosphate aldolase (FBA), glyceraldehyde 3-phosphate (GAPA) and phosphoglycerate kinase (PGK). With acetyl-CoA representing a critical metabolite involved in multiple biosynthetic pathways, including fatty acid biosynthesis (Xing and Poirier, 2012), it is perhaps not surprising that cICDH and ATP citrate lyase enzymes exhibit consistent diel abundance changes in both Group I and II kale (Figure 4; Supplemental Data 2). Similarly, FBA, GAPA and PGK enzymes consistently exhibited significant diel abundance changes across both Group I and II, offering a consistent benchmark for our systems-level proteome analysis (Supplemental Figure 4A). Intriguingly, different isoforms of PGK and FBA were found to possess significant changes in diel abundance between Group I and II kale, whose activity may be directly related to diel differences in Group I and Group II fructose levels though changes in fructose 1,6-bisphosphate. This possibility is further reinforced by the enrichment of multiple sugar related metabolic pathways amongst both Group I and II kale (Supplemental Data 6). In Arabidopsis, the three encoded FBAs were found to have essential roles in plant metabolism, but with tissue specific expression patterns (Carrera et al., 2021). Here however, it seems that Group I and II kale each utilize a different subset of FBA isozymes, as we analyzed the same leaf tissues across cultivars. Unlike Arabidopsis, the polyploid *B. napus* has been shown to possess twenty-two FBAs, which possess diverse developmental expression patterns across multiple cellular compartments (Zhao et al., 2019). Further, it seems that despite fructose and glucose representing two of the group defining metabolites in our study, very few protein groups related to the generation of fructose and glucose such as starch degradation enzymes, sucrose synthase or invertase, were found to be significantly changing in our proteomics data (Figure 4; Supplemental Data 2). This perhaps indicates that changes in these enzymes, such as their activity are driven by other regulatory mechanisms such as reversible protein phosphorylation, rather than changes in abundance (Hardin et al., 2004).

We also quantified changes in nitrogen assimilating enzymes in both Group I and II kale. Group I kale cultivars possess substantially larger diel changes in nitrite reductase (NiR) levels at ED relative to Group II, while glutamine synthetase isoform 1,4 (GLN1,4), which converts glutamate to glutamine as part of nitrogen assimilation for transport from roots to shoots, exhibited a consistent and significant change in abundance at ED in both Group I and II kale. This aligns with both our metabolite data (Figure 2; Supplemental Data 1) and pathway enrichment analysis (Supplemental Data 6), which finds glutamine levels to peak in nearly all nine kale cultivars at ED. In the model *Brassicaceae* Arabidopsis, it is well known that nitrogen metabolism is a diel regulated process (Flis et al., 2019), with peak transcript and protein abundance of nitrate reductase enzymes occurring early in the day, however here, it seems that the precise time-of-day coordination of these events differs between kale cultivars (Supplemental Figure 4B).

Directly related to both carbon and nitrogen metabolism, our proteomics data and pathway enrichment analysis also resolved extensive amino acid metabolism changes (Figure 4; Supplemental Data 2, 6). In particular, we find significant Group I and II differences in the diel abundance of enzymes related to glycine, serine and BCAA (e.g. iso-leucine) metabolism (Schulze et al., 2016). We also find time-of-day differences in iso-leucine production, which is produced down-stream of aspartate and possesses group-specific time-of-day production differences. Here, both our pathway (Supplemental Data 6) and proteomics (Supplemental Data 2) analyses support the diel differences in BCAA levels (e.g. isoleucine and valine) observed between Group I and II, with BCAA super-pathway enzymes possessing average ED (Log2FC = 0.63) and EN (Log2FC = 0.42) abundance differences, respectively. Interestingly, despite aspartate fueling lysine and methionine production, and our measurements of aspartate and lysine providing no indications of group specific responses (Figure 2; Supplemental Data 1), we find Group I kale to specifically possess a significant change in methionine metabolic enzymes. Collectively, these aligned diel differences in both the proteome and metabolome relating to amino acids makes a case for genetics-based time-of-day harvesting of kale for maximal nutritional content.

*mRNA Processing & Protein Translation* – RNA splicing and protein translation represent two cellular processes carried out by multi-subunit protein complexes; the spliceosome and the ribosome, respectively. However, much remains to be resolved as to how each of these complexes are regulated in plants at the protein level. Previous diel proteome analyses have found both the spliceosome and ribosome to be dynamically regulated at the protein-level by changes in both abundance and phosphorylation (Uhrig et al., 2021). However, resolution of species-specific differences in diel abundance has not previously been resolved. Here we see Group I and II kale defined by smaller abundance changes in mRNA processing and protein translation machinery at EN in Group I kale, coupled with larger abundance changes at ED in Group II kale. Correspondingly, a large abundance of chaperone proteins changing in a similar pattern at the same time-points is observed, likely aiding in effective protein production. Surprisingly, we also see a large abundance of plastidial translational machinery in both Group I and II kale, which parallel the diel changes in abundance found in their cytosolic counterparts, suggesting concerted coordination of global protein production in each kale group. Currently our understanding of diel changes in mRNA splicing and protein translation have largely been defined by transcriptomic sequencing technologies. In particular, use of polysome loading or RiboSeq as a proxy for protein translation, which in Arabidopsis is suggested to negatively correlate with biomass (Ishihara et al., 2017). Similarly, RNAseq profiling of Arabidopsis over a 24 h photoperiod has found diel changes in mRNA spliceforms (Romanowski et al., 2020), however, in both areas of research, direct diel assessment of spliceosome or ribosome complex composition and regulation at the protein-level has remained undefined. In light of the work performed in the related *Brassicaceae* Arabidopsis, our findings here suggest that there are systemic differences between the kale groups in the timing of growth as it relates to protein translation that could be utilized to enhance the productivity kale through the precise adjustment of growth conditions.

*Light Harvesting and Signaling* – In both Group I and II kale we see extensive diel changes in the light harvesting and photosynthetic machinery, with no specific ED or EN changes in either group. The largest changes observed involved the chloroplast ATP synthase delta subunit (ATPD) at EN in Group I kale along with enzymes ENHANCER OF SOS3-1 (ENH1) and NONPHOTOCHEMICAL QUENCHING 4 (NPQ4) at ED in Group II kale. Chloroplast ATP synthase is a critical driver of ATP production in plants in the light, with *Arabidopsis* plants lacking ATPD presenting a lethal phenotype due to the destabilization of the ATP synthase complex (Maiwald et al., 2003). Alternatively, ENH1 is required to main redox balance (Zhu et al., 2007) and NPQ1 is involved in non-photochemical quenching in the presence of excess light energy, which led to increased growth outcomes in tobacco when present in higher abundance (Kromdijk et al., 2016). Interestingly, in both Group I and II kale we also observe a small, but important network of proteins comprised of CRYPTOCHROME 1 (CRY1), ELONGATED HYPOCOTYL 5 HOMOLOG (HYH) and CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1), with CRY1 more abundant at EN in Group I kale. In response to blue light, CRY1 inhibits the degradation of the HY5 transcription factor by COP1 (Wang et al., 2018). Although there is a more limited understanding of HYH, HY5, an HYH ortholog, is a regulator of light-mediated transcription in plants, controlling a wide range of plant cell processes related to growth and development that are of importance to kale production (Xiao et al., 2022). CRY1 is also connected to the circadian clock through detection of blue light fluence (Somers et al., 1998; Sanchez et al., 2020), indicating potential higher-order regulation of timed-metabolism in Group I versus Group II kale. This has particularly intriguing chronoculture implications for Group I kale production in controlled growth environment settings given its larger diel abundance changes.

## Summary

Our integrated, systems-level analysis of nine diverse, commercially produced and readily consumed kale cultivars has substantially advanced our understanding of kale from the phenotypic-level to the underlying molecular-level. In doing this, our dataset reveals new information about the molecular landscapes of these kale cultivars when grown under standardized controlled growth environment conditions, providing new opportunities for vertical farming and/or horticultural growth of kale. Our systems-level analysis has defined diel differences in the molecular landscapes underpinning these diverse kale genetics, elucidating information for time-of-day harvesting considerations to ensure maximal nutritional content. Variations in the diel molecular landscapes of different cultivars or plant accessions have been previously observed in *Arabidopsis* (Rees et al., 2021), tomato (Muller et al., 2016) and soybean (Greenham et al., 2017), with explanations for these differences being related to environmental stimuli and/or geography (seasonal impacts and latitude) (Ruts

et al., 2012; Steppe et al., 2015; Rees et al., 2021). It is also interesting, that many of those differences have been linked to diel plant biology. Here, we observe differences in proteins linked to light perception and the circadian clock (e.g. CRY1), while simultaneously finding group-specific differences in multiple metabolites and pathways connected to diel biology and the circadian clock (Haydon et al., 2013; Cervela-Cardona et al., 2021; Scandola et al., 2022). Overall, our endeavor to define the underlying molecular landscape of diverse, commercially grown kale cultivars opens up new opportunities for horticultural production and targeted research activities moving forward. Further, our results suggest that combined use of plant phenotyping, proteomics and metabolomics represents a powerful approach for characterizing non-model horticultural crops of diverse genetic backgrounds.

# Materials and methods

## Growth conditions

Nine commercially grown cultivars of kale were purchased from OCS Seeds (Table 1; https://www.oscseeds.com/) and West Coast Seeds (https://www.westcoastseeds.com/) and grown for the study. These included: *B. oleracea* var. *sabellica* cultivars Winterbor (K2), Dwarf Curled Scotch (K3), Darkibor (K7), Starbor (K8), Scarlet (K9), *B. oleracea* var. *palmifolia* cultivars Lacinato (K4), Rainbow Lacinato (K6) and *B. napus* var. *pabularia* cultivars Red Russian (K5) and Red Ursa (K10). Seeds were sterilized in 70% ethanol for 2 min followed by a 70% (v/v) bleach (Chlorox 7.5%) treatment for 7 min and 3 washes with distilled water. The seeds were then grown on ½ MS media containing 1% (w/v) sucrose and 7 g/L of agar at pH 5.8. The seeds were cold treated 3 days at 4°C in the dark and exposed to a 12h light and 12h dark photoperiod consisting of 100 PPFD for a week before being transferred to soil (Sun Gro®, Sunshine Mix® #1). At 29 days post-sterilization, entire plants were collected for GC-MS and LC-MS analysis. Growth chambers were equipped with a programmable Perihelion LED fixture (G2V Optics Inc; https://g2voptics.com/) and lined with Reflectix® to ensure a good light distribution. Kales were grown under a 12h light and 12h dark regimen consisting of 100 PPFD and a temperature of 21°C during the day and 19°C at night. All physiological and phenotypic assessments were performed at ZT6 (mid-day). Physiological measurements were taken using a multispecQ device (PhotosynQ; https://www.photosynq.com/product-page/multispeq-v-2-0).

## Metabolite extraction and GC-MS analysis

*Metabolite Extraction and Data Acquisition* - Metabolite extraction and preparation were performed with modifications as previously described (Liu et al., 2016). Leaf tissue was harvested at ZT23 (end-of-night) and ZT11 (end-of-day) and directly flash frozen in liquid nitrogen (n = 4; each biological replicate consists of a pool of leaf tissue from 3 plants). Sample of 100 mg (+/- 1 mg) of pulverized tissue were prepared and homogenized in 700 µl of iced-cold methanol (80% v/v). In each sample, 25 µl of ribitol at 0.4 $mg.ml^{-1}$ in water were added as internal standard. Samples were incubated 2 h at 4 °C with shaking and then 15 min at 70°C at 850 rpm in a Thermomixer. Tubes were centrifuged 30 min at 12000 rpm and the supernatants were transferred in new tubes. Polar and non-polar phases were separated by the addition of 700 µl of water and 350 µl of chloroform, then vortexed thoroughly and centrifuged for 15 min at 5000 rpm. The upper methanol/water phase (150 µl) was transferred to a new tube and dry in a vacuum centrifuge at RT. Samples were derivatized with 100 µl of methoxamine hydrochloride-HCl (20 $mg.ml^{-1}$ in pyridine) for 90 min at 30°C at 850 rpm in thermomixer and followed by incubation with 100 µL of N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) at 80°C during 30 min with shaking at 850 rpm in thermomixer. Finally, samples were injected in split less mode and analyzed using a 7890A gas chromatograph coupled to a 5975C quadrupole mass detector (Agilent Technologies, Palo Alto, CA, USA). In the same manner, 1 µl of retention time standard mixture Supelco C7–C40 saturated alkanes (1,000 $µg.ml^{-1}$ of each component in hexane) diluted 100 fold (10 $µg.ml^{-1}$ final concentration) was injected and analyzed. Alkanes were dissolved in pyridine at 0.22 $mg.ml^{-1}$ final concentration. Chromatic separation was done with a DB-5MS capillary column (30 m × 0.25 mm × 0.25 µm; Agilent J&W Scientific, Folsom, CA, USA). Inlet temperature was set at 280°C. Initial GC Oven temperature was set to 80°C and held for 2 min after injection then GC oven temperature was raised to 300°C at 7°C $min^{-1}$, and finally held at 300°C for 10 min. Injection and ion source temperatures were adjusted to 300°C and 200°C, respectively with a solvent delay of 5 min. The carrier gas (Helium) flow rate was set to 1 $ml.min^{-1}$. The detector was operated in EI mode at 70 eV and in full scan mode (m/z 33–600).

*Metabolites Data Analysis* - Compounds were identified by mass spectral and retention time index matching to the mass spectra of the National Institute of Standards and Technology library (NIST20, https://www.nist.gov/) and the Golm Metabolome Database (GMD, http://gmd.mpimp-golm.mpg.de/). Metabolite quantification was performed using MassHunter Software from Agilent. Peaks were deconvoluted and integrated and were normalized to the internal standard ribitol and by the sample weight.

## Protein extraction and nanoflow LC-MS analysis

*Protein Extraction and Data Acquisition* - Kale leaf tissue was harvested at ZT23 and ZT11, flash frozen and ground to a fine powder under liquid $N_2$ using a mortar and pestle and aliquoted into 400 mg fractions (n = 4; each biological replicate consists of a pool of 3 plants). Samples were then extracted at a 1:2 (w/v) ratio with a solution of 50 mM HEPES-KOH pH 8.0, 50 mM NaCl, and 4% (w/v) SDS. This included vortexing, followed by incubation at 95°C in an Eppendorf microtube table-top shaking incubator shaking at 1100 RPM for 15 mins. This was then followed by an additional 15 mins of shaking at room temperature. All samples

were clarified at 20,000 x g for 5 min at room temperature, with the supernatant retained in fresh Eppendorf microtubes. Sample protein concentrations were measured by bicinchoninic acid (BCA) assay (23225; ThermoScientific), followed by reduction with 10 mM dithiothreitol (DTT) at 95°C for 5 mins. Samples were then cooled and alkylated with 30 mM iodoacetamide (IA) for 30 min in the dark without shaking at room temperature. Subsequently, 10 mM DTT was added to each sample, followed by a quick vortex, and incubation for 10 min at room temperature without shaking. Total proteome peptide pools were generated by sample digestion overnight with 1:100 sequencing grade trypsin (V5113; Promega). Generated peptide pools were quantified by Nanodrop, followed by acidification with formic acid to a final concentration of 5% (v/v) and then dried by vacuum centrifugation. Peptides were then desalted using ZipTip C18 pipette tips (ZTC18S960; Millipore) as previously described [7], dried and dissolved in 3.0% ACN/0.1% FA prior to MS analysis. Digested samples were then analysed using a Fusion Lumos Tribrid Orbitrap mass spectrometer (Thermo Scientific) in a data independent acquisition (DIA) mode using the BoxCarDIA method as previously descried [31]. Dissolved peptides (1 µg) were injected using an Easy-nLC 1200 system (LC140; ThermoScientific) and separated on a 50 cm Easy-Spray PepMap C18 Column (ES903; ThermoScientific). Liquid chromatography and BoxCar DIA acquisition was performed as previously described without deviation (Mehta et al., 2022).

*Proteomic Data Analysis* – All acquired BoxCar DIA data was analyzed in a library-free DIA approach using Spectronaut v14 (Biognosys AG) using default settings. Data were searched using the *B. oleracea* var *oleracea* proteome (Uniprot: https://www.uniprot.org/containing 58,545 proteins). Default search parameters for proteome quantification were used, with specific search parameters including: a protein, peptide and PSM FDR of 1%, trypsin digestion with 1 missed cleavage, fixed modification including carbamidomethylation of cysteine residues and variable modifications including methionine oxidation. Data was Log2 transformed and globally normalized by median subtraction with significantly changing differentially abundant proteins determined and corrected for multiple comparisons (Bonferroni-corrected *p-value* ≤ 0.05; q-value).

## Bioinformatics

Gene Ontology enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID; v 6.8; https://david.ncifcrf.gov/home.jsp). Significance was determined using Benjamini-Hochberg (BH) corrected *p-value* ≤ 0.05. Conversion of *B. oleracea var oleracea* gene identifiers to *Arabidopsis* gene identifiers for STRING association network analysis and SUBA4 subcellular localization information retrieval was performed using UniProt (https://www.uniprot.org/). STRING association network analyses were performed in Cytoscape v3.9.0 (https://cytoscape.org/) using the String DB plugin stringApp, all datatypes and a minimum correlation coefficient setting of 0.9. Predicted subcellular localization information was obtained using SUBA4 and the consensus subcellular localization predictor SUBAcon (https://suba.live/). The Eucledian distance *pheatmap* clustered heatmap analysis of diel metabolites was performed in R (3.6.1 R Core Team, 2019). Pathway enrichment was performed using the Plant Metabolic Network (PMN; https://plantcyc.org/) with enrichment determined by a Fisher's exact test significance threshold of *p-value* < 0.01. Additionally figures were assembled using Affinity Designer software (v1.9.1.179; https://affinity.serif.com/en-us/designer/).

## Data availability statement

All raw data files, mass spectrometry parameters, and Spectronaut search settings have been uploaded to ProteomeXchange (http://www.proteomexchange.org/) via the PRoteomics IDEntification Database (PRIDE; https://www.ebi.ac.uk/pride/). Project Accession: PXD031780.

## Author contributions

Plant growth, phenotyping and harvesting was performed by SS, NB and BC. Metabolite data acquisition and analyses were performed by SS. Proteomics data acquisition and analysis was performed by DM, SS and RGU. The manuscript was written by SS and RGU, with editorial input from DM. All authors contributed to the article and approved the submitted version.

## Funding

Funding for this work was generously provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Alberta Innovates through their Alliance and Campus Alberta Small Business Engagement (CASBE) funding schemes, respectively. This work was also supported by the Canada Foundation for Innovation (CFI).

## Acknowledgments

We would like to thank G2V Optics Inc. for providing their programmable LED lighting system. Thanks to Jack Moore and The Alberta Proteomics and Mass Spectrometry Facility (APM) for providing proteomic support for this work.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1170448/full#supplementary-material

SUPPLEMENTAL DATA SHEET 1
Gas chromatography metabolite data.

SUPPLEMENTAL DATA SHEET 2
Liquid chromatography proteomic data.

SUPPLEMENTAL DATA SHEET 3
Uniprot gene identifier conversion from *Brassica oleracea* to *Arabidopsis*.

SUPPLEMENTAL DATA SHEET 4
PlantCV data and data processing.

SUPPLEMENTAL DATA SHEET 5
MultisynQ phenotype data.

SUPPLEMENTAL DATA SHEET 6
Pathway enrichment analysis using the Plant Metabolic Network (PMN; https://plantcyc.org/.

## References

Akram, N. A., Shafiq, F., and Ashraf, M. (2017). Ascorbic acid-a potential oxidant scavenger and its role in plant development and abiotic stress tolerance. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00613

Arias, T., Niederhuth, C. E., McSteen, P., and Pires, J. C. (2021). The molecular basis of kale domestication: transcriptional profiling of developing leaves provides new insights into the evolution of a brassica oleracea vegetative morphotype. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.637115

Armenta-Medina, A., Gillmor, C. S., Gao, P., Mora-Macias, J., Kochian, L. V., Xiang, D., et al. (2021). Developmental and genomic architecture of plant embryogenesis: from model plant to crops. *Plant Commun.* 2 (1), 100136. doi: 10.1016/j.xplc.2020.100136

Ayaz, F. A., Glew, R. H., Millson, M., Huang, H. S., Chuang, L. T., Sanz, C., et al. (2006). Nutrient contents of kale (Brassica oleraceae l. var. acephala DC.). *Food Chem.* 96 (4), 572–579. doi: 10.1016/j.foodchem.2005.03.011

Barker, S., Moss, R., and McSweeney, M. B. (2022). Identification of sensory properties driving consumers' liking of commercially available kale and arugula. *J. Sci. Food Agric.* 102 (1), 198–205. doi: 10.1002/jsfa.11346

Becerra-Moreno, A., Alanis-Garza, P. A., Mora-Nieves, J. L., Mora-Mora, J. P., and Jacobo-Velazquez, D. A. (2014). Kale: an excellent source of vitamin c, pro-vitamin a, lutein and glucosinolates. *Cyta-J Food* 12 (3), 298–303. doi: 10.1080/19476337.2013.850743

Bogdanovic, J., Mojovic, M., Milosavic, N., Mitrovic, A., Vucinic, Z., and Spasojevic, I. (2008). Role of fructose in the adaptation of plants to cold-induced oxidative stress. *Eur. Biophys. J.* 37 (7), 1241–1246. doi: 10.1007/s00249-008-0260-9

Carrera, D. A., George, G. M., Fischer-Stettler, M., Galbier, F., Eicke, S., Truernit, E., et al. (2021). Distinct plastid fructose bisphosphate aldolases function in photosynthetic and non-photosynthetic metabolism in arabidopsis. *J. Exp. Bot.* 72 (10), 3739–3755. doi: 10.1093/jxb/erab099

Carvalho, S. D., and Folta, K. M. (2014). Sequential light programs shape kale (Brassica napus) sprout appearance and alter metabolic and nutrient content. *Hortic. Res.* 1, 8. doi: 10.1038/hortres.2014.8

Carvalho, S. D., Schwieterman, M. L., Abrahan, C. E., Colquhoun, T. A., and Folta, K. M. (2016). Light quality dependent changes in morphology, antioxidant capacity, and volatile production in sweet basil (Ocimum basilicum). *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01328

Casajus, V., Perini, M., Ramos, R., Lourenco, A. B., Salinas, C., Sanchez, E., et al. (2021). Harvesting at the end of the day extends postharvest life of kale (Brassica oleracea var. sabellica). *Sci. Hortic-Amsterdam* 276 (109757), 1-7. doi: 10.1016/j.scienta.2020.109757

Cervela-Cardona, L., Yoshida, T., Zhang, Y., Okada, M., Fernie, A., and Mas, P. (2021). Circadian control of metabolism by the clock component TOC1. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.683516

Chiu, Y. C., Juvik, J. A., and Ku, K. M. (2018). Targeted metabolomic and transcriptomic analyses of "Red russian" kale (Brassicae napus var. pabularia) following methyl jasmonate treatment and larval infestation by the cabbage looper (Trichoplusia ni hubner). *Int. J. Mol. Sci.* 19, (4). doi: 10.3390/ijms19041058

Covington, M. F., Maloof, J. N., Straume, M., Kay, S. A., and Harmer, S. L. (2008). Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol.* 9 (8), R130. doi: 10.1186/gb-2008-9-8-r130

Creux, N., and Harmer, S. (2019). Circadian rhythms in plants. *Cold Spring Harb. Perspect. Biol.* 11, (9). doi: 10.1101/cshperspect.a034611

Dong, N. Q., and Lin, H. X. (2021). Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. *J. Integr. Plant Biol.* 63 (1), 180–209. doi: 10.1111/jipb.13054

Flis, A., Mengin, V., Ivakov, A. A., Mugford, S. T., Hubberten, H. M., Encke, B., et al. (2019). Multiple circadian clock outputs regulate diel turnover of carbon and nitrogen reserves. *Plant Cell Environ.* 42 (2), 549–573. doi: 10.1111/pce.13440

Francisco, M., and Rodriguez, V. M. (2021). Importance of daily rhythms on brassicaceae phytochemicals. *Agronomy-Basel* 11, (4). doi: 10.3390/agronomy11040639

Galili, G., and Amir, R. (2013). Fortifying plants with the essential amino acids lysine and methionine to improve nutritional quality. *Plant Biotechnol. J.* 11 (2), 211–222. doi: 10.1111/pbi.12025

Gao, P., Quilichini, T. D., Yang, H., Li, Q., Nilsen, K. T., Qin, L., et al. (2022). Evolutionary divergence in embryo and seed coat development of u's triangle brassica species illustrated by a spatiotemporal transcriptome atlas. *New Phytol.* 233 (1), 30–51. doi: 10.1111/nph.17759

Greenham, K., Lou, P., Puzey, J. R., Kumar, G., Arnevik, C., Farid, H., et al. (2017). Geographic variation of plant circadian clock function in natural and agricultural settings. *J. Biol. Rhythms* 32 (1), 26–34. doi: 10.1177/0748730416679307

Han, M., Zhang, C., Suglo, P., Sun, S., Wang, M., and Su, T. (2021). L-aspartate: an essential metabolite for plant growth and stress acclimation. *Molecules* 26, (7). doi: 10.3390/molecules26071887

Hardin, S. C., Winter, H., and Huber, S. C. (2004). Phosphorylation of the amino terminus of maize sucrose synthase in relation to membrane association and enzyme activity. *Plant Physiol.* 134 (4), 1427–1438. doi: 10.1104/pp.103.036780

Haydon, M. J., Mielczarek, O., Robertson, F. C., Hubbard, K. E., and Webb, A. A. (2013). Photosynthetic entrainment of the arabidopsis thaliana circadian clock. *Nature* 502 (7473), 689–692. doi: 10.1038/nature12603

He, M., and Ding, N. Z. (2020). Plant unsaturated fatty acids: multiple roles in stress response. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.562785

Hotta, C. T. (2021). From crops to shops: how agriculture can use circadian clocks. *J. Exp. Bot.* 72 (22), 7668–7679. doi: 10.1093/jxb/erab371

Ishihara, H., Moraes, T. A., Pyl, E. T., Schulze, W. X., Obata, T., Scheffel, A., et al. (2017). Growth rate correlates negatively with protein turnover in arabidopsis accessions. *Plant J.* 91 (3), 416–429. doi: 10.1111/tpj.13576

Jeon, J., Kim, J. K., Kim, H., Kim, Y. J., Park, Y. J., Kim, S. J., et al. (2018). Transcriptome analysis and metabolic profiling of green and red kale (Brassica oleracea var. acephala) seedlings. *Food Chem.* 241, 7–13. doi: 10.1016/j.foodchem.2017.08.067

Jin, S. W., Rahim, M. A., Afrin, K. S., Park, J. I., Kang, J. G., and Nou, I. S. (2018). Transcriptome profiling of two contrasting ornamental cabbage (Brassica oleracea var. acephala) lines provides insights into purple and white inner leaf pigmentation. *BMC Genomics* 19 (1), 797. doi: 10.1186/s12864-018-5199-3

Joshi, V., Joung, J. G., Fei, Z., and Jander, G. (2010). Interdependence of threonine, methionine and isoleucine metabolism in plants: accumulation and transcriptional regulation under abiotic stress. *Amino Acids* 39 (4), 933–947. doi: 10.1007/s00726-010-0505-7

Kim, J. A., Kim, H. S., Choi, S. H., Jang, J. Y., Jeong, M. J., and Lee, S. I. (2017). The importance of the circadian clock in regulating plant metabolism. *Int. J. Mol. Sci.* 18, (12). doi: 10.3390/ijms18122680

Klopsch, R., Baldermann, S., Voss, A., Rohn, S., Schreiner, M., and Neugart, S. (2019). Narrow-banded UVB affects the stability of secondary plant metabolites in kale (Brassica oleracea var. sabellica) and pea (Pisum sativum) leaves being added to lentil flour fortified bread: a novel approach for producing functional foods. *Foods* 8, (10). doi: 10.3390/foods8100427

Krahmer, J., Hindle, M., Perby, L. K., Mogensen, H. K., Nielsen, T. H., Halliday, K. J., et al. (2022). The circadian clock gene circuit controls protein and phosphoprotein rhythms in arabidopsis thaliana. *Mol. Cell Proteomics* 21 (1), 100172. doi: 10.1016/j.mcpro.2021.100172

Kromdijk, J., Glowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K., et al. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* 354 (6314), 857–861. doi: 10.1126/science.aai8878

Kuhlgert, S., Austic, G., Zegarac, R., Osei-Bonsu, I., Hoh, D., Chilvers, M. I., et al. (2016). MultispeQ beta: a tool for large-scale plant phenotyping connected to the open PhotosynQ network. *R Soc. Open Sci.* 3 (10), 160592. doi: 10.1098/rsos.160592

Liu, Y., Feng, X., Zhang, Y., Zhou, F., and Zhu, P. (2021). Simultaneous changes in anthocyanin, chlorophyll, and carotenoid contents produce green variegation in pink-leaved ornamental kale. *BMC Genomics* 22 (1), 455. doi: 10.1186/s12864-021-07785-x

Liu, X. S., Yang, X., Wang, L. M., Duan, Q. Q., and Huang, D. F. (2016). Comparative analysis of metabolites profile in spinach (Spinacia oleracea l.) affected by different concentrations of gly and nitrate. *Sci. Hortic-Amsterdam* 204, 8–15. doi: 10.1016/j.scienta.2016.02.037

Liu, X., Zhang, B., Wu, J., Li, Z., Han, F., Fang, Z., et al. (2020). Pigment variation and transcriptional response of the pigment synthesis pathway in the S2309 triple-color ornamental kale (Brassica oleracea l. var. acephala) line. *Genomics* 112 (3), 2658–2665. doi: 10.1016/j.ygeno.2020.02.019

Maiwald, D., Dietzmann, A., Jahns, P., Pesaresi, P., Joliot, P., Joliot, A., et al. (2003). Knock-out of the genes coding for the rieske protein and the ATP-synthase delta-subunit of arabidopsis. effects on photosynthesis, thylakoid protein composition, and nuclear chloroplast gene expression. *Plant Physiol.* 133 (1), 191–202. doi: 10.1104/pp.103.024190

Megias-Perez, R., Hahn, C., Ruiz-Matute, A. I., Behrends, B., Albach, D. C., and Kuhnert, N. (2020). Changes in low molecular weight carbohydrates in kale during development and acclimation to cold temperatures determined by chromatographic techniques coupled to mass spectrometry. *Food Res. Int.* 127, 108727. doi: 10.1016/j.foodres.2019.108727

Mehta, D., Krahmer, J., and Uhrig, R. G. (2021). Closing the protein gap in plant chronobiology. *Plant J.* 106 (6), 1509–1522. doi: 10.1111/tpj.15254

Mehta, D., Scandola, S., and Uhrig, R. G. (2022). BoxCar and library-free data-independent acquisition substantially improve the depth, range, and completeness of label-free quantitative proteomics. *Anal. Chem.* 94 (2), 793–802. doi: 10.1021/acs.analchem.1c03338

Migliozzi, M., Thavarajah, D., Thavarajah, P., and Smith, P. (2015). Lentil and kale: complementary nutrient-rich whole food sources to combat micronutrient and calorie malnutrition. *Nutrients* 7 (11), 9285–9298. doi: 10.3390/nu7115471

Moing, A., Allwood, J. W., Aharoni, A., Baker, J., Beale, M. H., Ben-Dor, S., et al. (2020). Comparative metabolomics and molecular phylogenetics of melon (Cucumis melo, cucurbitaceae) biodiversity. *Metabolites* 10, (3). doi: 10.3390/metabo10030121

Muller, N. A., Wijnen, C. L., Srinivasan, A., Ryngajllo, M., Ofner, I., Lin, T., et al. (2016). Domestication selected for deceleration of the circadian clock in cultivated tomato. *Nat. Genet.* 48 (1), 89–93. doi: 10.1038/ng.3447

Munnik, T., Irvine, R. F., and Musgrave, A. (1998). Phospholipid signalling in plants. *Biochim. Biophys. Acta* 1389 (3), 222–272. doi: 10.1016/s0005-2760(97)00158-6

Nemzer, B., Al-Taher, F., and Abshiru, N. (2021). Extraction and natural bioactive molecules characterization in spinach, kale and purslane: a comparative study. *Molecules* 26 (9), 2515. doi: 10.3390/molecules26092515

Neugart, S., and Bumke-Vogt, C. (2021). Flavonoid glycosides in brassica species respond to UV-b depending on exposure time and adaptation time. *Molecules* 26 (9), 494. doi: 10.3390/molecules26020494

Neugart, S., Fiol, M., Schreiner, M., Rohn, S., Zrenner, R., Kroh, L. W., et al. (2014). Interaction of moderate UV-b exposure and temperature on the formation of structurally different flavonol glycosides and hydroxycinnamic acid derivatives in kale (Brassica oleracea var. sabellica). *J. Agric. Food Chem.* 62 (18), 4054–4062. doi: 10.1021/jf4054066

Neugart, S., Krumbein, A., and Zrenner, R. (2016). Influence of light and temperature on gene expression leading to accumulation of specific flavonol glycosides and hydroxycinnamic acid derivatives in kale (Brassica oleracea var. *sabellica). Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00326

Paciolla, C., Fortunato, S., Dipierro, N., Paradiso, A., De Leonardis, S., Mastropasqua, L., et al. (2019). Vitamin c in plants: from functions to biofortification. *Antioxidants-Basel* 8 (11), 519. doi: 10.3390/antiox8110519

Pongrac, P., Fischer, S., Thompson, J. A., Wright, G., and White, P. J. (2019). Early responses of brassica oleracea roots to zinc supply under sufficient and Sub-optimal phosphorus supply. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01645

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Reda, T., Thavarajah, P., Polomski, R., Bridges, W., Shipe, E., and Thavarajah, D. (2021). Reaching the highest shelf: a review of organic production, nutritional quality, and shelf life of kale (Brassica oleracea var. acephala). *Plants People Planet* 3 (4), 308–318. doi: 10.1002/ppp3.10183

Rees, H., Joynson, R., Brown, J. K. M., and Hall, A. (2021). Naturally occurring circadian rhythm variation associated with clock gene loci in Swedish arabidopsis accessions. *Plant Cell Environ.* 44 (3), 807–820. doi: 10.1111/pce.13941

Rennie, E. A., and Scheller, H. V. (2014). Xylan biosynthesis. *Curr. Opin. Biotechnol.* 26, 100–107. doi: 10.1016/j.copbio.2013.11.013

Romanowski, A., Schlaen, R. G., Perez-Santangelo, S., Mancini, E., and Yanovsky, M. J. (2020). Global transcriptome analysis reveals circadian control of splicing events in arabidopsis thaliana. *Plant J.* 103 (2), 889–902. doi: 10.1111/tpj.14776

Ruts, T., Matsubara, S., Wiese-Klinkenberg, A., and Walter, A. (2012). Diel patterns of leaf and root growth: endogenous rhythmicity or environmental response? *J. Exp. Bot.* 63 (9), 3339–3351. doi: 10.1093/jxb/err334

Samec, D., Urlic, B., and Salopek-Sondi, B. (2019). Kale (Brassica oleracea var. acephala) as a superfood: review of the scientific evidence behind the statement. *Crit. Rev. Food Sci.* 59 (15), 2411–2422. doi: 10.1080/10408398.2018.1454400

Sanchez, S. E., Rugnone, M. L., and Kay, S. A. (2020). Light perception: a matter of time. *Mol. Plant* 13 (3), 363–385. doi: 10.1016/j.molp.2020.02.006

Scandola, S., Mehta, D., Li, Q., Rodriguez Gallo, M. C., Castillo, B., and Uhrig, R. G. (2022). Multi-omic analysis shows REVEILLE clock genes are involved in carbohydrate metabolism and proteasome function. *Plant Physiol.* 190 (2), 1005-1023. doi: 10.1093/plphys/kiac269

Schulze, S., Westhoff, P., and Gowik, U. (2016). Glycine decarboxylase in C3, C4 and C3-C4 intermediate species. *Curr. Opin. Plant Biol.* 31, 29–35. doi: 10.1016/j.pbi.2016.03.011

Shahidi, F., and Ambigaipalan, P. (2018). Omega-3 polyunsaturated fatty acids and their health benefits. *Annu. Rev. Food Sci. T* 9, 345–381. doi: 10.1146/annurev-food-111317-095850

Shi, L. Y., Cao, S. F., Shao, J. R., Chen, W., Zheng, Y. H., Jiang, Y. M., et al. (2014). Relationship between sucrose metabolism and anthocyanin biosynthesis during ripening in Chinese bayberry fruit. *J. Agr Food Chem.* 62 (43), 10522–10528. doi: 10.1021/jf503317k

Somers, D. E., Devlin, P. F., and Kay, S. A. (1998). Phytochromes and cryptochromes in the entrainment of the arabidopsis circadian clock. *Science* 282 (5393), 1488–1490. doi: 10.1126/science.282.5393.1488

Steed, G., Ramirez, D. C., Hannah, M. A., and Webb, A. A. R. (2021). Chronoculture, harnessing the circadian clock to improve crop yield and sustainability. *Science* 372 (6541), eabc9141. doi: 10.1126/science.abc9141

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

Steppe, K., Sterck, F., and Deslauriers, A. (2015). Diel growth dynamics in tree stems: linking anatomy and ecophysiology. *Trends Plant Sci.* 20 (6), 335–343. doi: 10.1016/j.tplants.2015.03.015

Thavarajah, D., Thavarajah, P., Abare, A., Basnagala, S., Lacher, C., Smith, P., et al. (2016). Mineral micronutrient and prebiotic carbohydrate profiles of USA -grown kale (Brassica oleracea l. var. acephala). *J. Food Compos Anal.* 52, 9–15. doi: 10.1016/j.jfca.2016.07.003

Tohge, T., Watanabe, M., Hoefgen, R., and Fernie, A. R. (2013). Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00062

Uhrig, R. G., Echevarria-Zomeno, S., Schlapfer, P., Grossmann, J., Roschitzki, B., Koerber, N., et al. (2021). Diurnal dynamics of the arabidopsis rosette proteome and phosphoproteome. *Plant Cell Environ.* 44 (3), 821–841. doi: 10.1111/pce.13969

Uhrig, R. G., Schlapfer, P., Roschitzki, B., Hirsch-Hoffmann, M., and Gruissem, W. (2019). Diurnal changes in concerted plant protein phosphorylation and acetylation in arabidopsis organs and seedlings. *Plant J.* 99 (1), 176–194. doi: 10.1111/tpj.14315

Wang, Q., Zuo, Z., Wang, X., Liu, Q., Gu, L., Oka, Y., et al. (2018). Beyond the photocycle-how cryptochromes regulate photoresponses in plants? *Curr. Opin. Plant Biol.* 45 (Pt A), 120–126. doi: 10.1016/j.pbi.2018.05.014

Waterland, N. L., Moon, Y., Tou, J. C., Kopsell, D. A., Kim, M. J., and Park, S. (2019). Differences in leaf color and stage of development at harvest influenced phytochemical content in three cultivars of kale (Brassica oleracea l. and b. napus). *J. Agric. Sci.* 11 (3). doi: 10.5539/jas.v11n3p14

Xiao, Y., Chu, L., Zhang , Y., Bian, Y., Xiao, J., Xu, D., et al. (2002). HY5: A Pivotal Regulator of Light-Dependent Development in Higher Plants. *Front Plant Sci.* 12, 800989. doi: 10.3389/fpls.2021.800989

Xing, S., and Poirier, Y. (2012). The protein acetylome and the regulation of metabolism. *Trends Plant Sci.* 17 (7), 423–430. doi: 10.1016/j.tplants.2012.03.008

Xu, X., Yuan, L., Yang, X., Zhang, X., Wang, L., and Xie, Q. (2022). Circadian clock in plants: linking timing to fitness. *J. Integr. Plant Biol.* 64 (4), 792–811. doi: 10.1111/jipb.13230

Zhao, W., Liu, H., Zhang, L., Hu, Z., Liu, J., Hua, W., et al. (2019). Genome-wide identification and characterization of FBA gene family in polyploid crop brassica napus. *Int. J. Mol. Sci.* 20 (22), 5749. doi: 10.3390/ijms20225749

Zhu, J. H., Fu, X. M., Koo, Y. D., Zhu, J. K., Jenney, F. E., Adams, M. W. W., et al. (2007). An enhancer mutant of arabidopsis salt overly sensitive 3 mediates both ion homeostasis and the oxidative stress response. *Mol. Cell Biol.* 27 (14), 5214–5224. doi: 10.1128/Mcb.01989-06

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 August 2023
DOI 10.3389/fpls.2023.1252777

Check for updates

OPEN ACCESS

EDITED BY
Christos Bazakos,
Max Planck Institute for Plant Breeding Research, Germany

REVIEWED BY
Hao Li,
Henan University, China
Christos Noutsos,
State University of New York at Old Westbury, United States

*CORRESPONDENCE
Pasquale Tripodi
pasquale.tripodi@crea.gov.it
Sandra Goritschnig
s.goritschnig@cgiar.org

RECEIVED 04 July 2023
ACCEPTED 26 July 2023
PUBLISHED 18 August 2023

CITATION
Tripodi P, Beretta M, Peltier D, Kalfas I, Vasilikiotis C, Laidet A, Briand G, Aichholz C, Zollinger T, Treuren Rv, Scaglione D and Goritschnig S (2023) Development and application of Single Primer Enrichment Technology (SPET) SNP assay for population genomics analysis and candidate gene discovery in lettuce. *Front. Plant Sci.* 14:1252777. doi: 10.3389/fpls.2023.1252777

COPYRIGHT
© 2023 Tripodi, Beretta, Peltier, Kalfas, Vasilikiotis, Laidet, Briand, Aichholz, Zollinger, Treuren, Scaglione and Goritschnig. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Development and application of Single Primer Enrichment Technology (SPET) SNP assay for population genomics analysis and candidate gene discovery in lettuce

Pasquale Tripodi[1]*, Massimiliano Beretta[2], Damien Peltier[3], Ilias Kalfas[4], Christos Vasilikiotis[5], Anthony Laidet[6], Gael Briand[6], Charlotte Aichholz[7], Tizian Zollinger[8], Rob van Treuren[9], Davide Scaglione[10] and Sandra Goritschnig[11]*

[1]Council for Agricultural Research and Economics (CREA), Research Centre for Vegetable and Ornamental Crops, Pontecagnano Faiano, SA, Italy, [2]ISI Sementi SpA, Fidenza (PR), Italy, [3]Limagrain - Vilmorin-Mikado, La Ménitré, France, [4]American Farm School, Thessaloniki, Greece, [5]Perrotis College, American Farm School, Thessaloniki, Greece, [6]Gautier Semences Route d'Avignon 13630, Eyragues, France, [7]Sativa Rheinau AG, Rheinau, Switzerland, [8]Zollinger Conseilles Sarl, Les Evouettes, Switzerland, [9]Centre for Genetic Resources, the Netherlands (CGN), Wageningen University and Research, Wageningen, Netherlands, [10]IGA Technology Services Srl, Udine, Italy, [11]European Cooperative Programme for Plant Genetic Resources (ECPGR) Secretariat c/o Alliance of Bioversity International and CIAT, Rome, Italy

Single primer enrichment technology (SPET) is a novel high-throughput genotyping method based on short-read sequencing of specific genomic regions harboring polymorphisms. SPET provides an efficient and reproducible method for genotyping target loci, overcoming the limits associated with other reduced representation library sequencing methods that are based on a random sampling of genomic loci. The possibility to sequence regions surrounding a target SNP allows the discovery of thousands of closely linked, novel SNPs. In this work, we report the design and application of the first SPET panel in lettuce, consisting of 41,547 probes spanning the whole genome and designed to target both coding (~96%) and intergenic (~4%) regions. A total of 81,531 SNPs were surveyed in 160 lettuce accessions originating from a total of 10 countries in Europe, America, and Asia and representing 10 horticultural types. Model ancestry population structure clearly separated the cultivated accessions (*Lactuca sativa*) from accessions of its presumed wild progenitor (*L. serriola*), revealing a total of six genetic subgroups that reflected a differentiation based on cultivar typology. Phylogenetic relationships and principal component analysis revealed a clustering of butterhead types and a general differentiation between germplasm originating from Western and Eastern Europe. To determine the potentiality of SPET for gene discovery, we performed genome-wide association analysis for main agricultural traits in *L. sativa* using six models (GLM naive, MLM, MLMM, CMLM, FarmCPU, and BLINK) to compare their strength and power for association detection. Robust associations were detected for seed color on chromosome 7 at 50 Mbp.

Colocalization of association signals was found for outer leaf color and leaf anthocyanin content on chromosome 9 at 152 Mbp and on chromosome 5 at 86 Mbp. The association for bolting time was detected with the GLM, BLINK, and FarmCPU models on chromosome 7 at 164 Mbp. Associations were detected in chromosomal regions previously reported to harbor candidate genes for these traits, thus confirming the effectiveness of SPET for GWAS. Our findings illustrated the strength of SPET for discovering thousands of variable sites toward the dissection of the genomic diversity of germplasm collections, thus allowing a better characterization of lettuce collections.

KEYWORDS

**lettuce, SPET, high-throughput genotyping, genomic diversity, phenotyping, GWAS, candidate genes**

## 1 Introduction

Recent years witnessed astonishing advancements in the development of cutting-edge technologies for next-generation sequencing (NGS), opening new frontiers for investigating the genomic diversity of crops (Van Treuren and van Hintum, 2014; Onda and Mochida, 2016). The availability of reference genome sequences and the progress in the field of bioinformatics made it possible to implement high-throughput genotyping methods capable of massively detecting single-nucleotide polymorphisms (SNPs). Being highly abundant across the genome and given their biallelic nature (Wendt and Novroski, 2019), SNPs offer the opportunity to be processed in automated pipelines providing a high resolution in the analysis of population structure and genetic ancestry, enabling furthermore a high-density scan of variants underlying complex traits. Different techniques for the identification of polymorphisms either in specific sites or randomly have therefore been developed. Among these, arrays based on customized oligonucleotide (allele-specific) probes hybridized on solid supports (Tripodi, 2022) offer an efficient technology combining a robust allele calling rate with lower investments in terms of library preparation and downstream bioinformatic analyses. However, arrays are affected by ascertainment bias due to the non-arbitrary sampling of polymorphisms and to the low representativeness of samples used to design the SNP panel leading to the exclusion of rare alleles (You et al., 2018). Furthermore, they are not flexible in terms of upgrades, requiring significant costs to increase the throughput.

The possibility to curtail the complexity of genomes and apply NGS, increasing read depth in determined genomic regions, enabled the development of reduced-representation library based-methods (RRL) (Van Tassell et al., 2008). Among these, genotyping by sequencing (GBS) and restriction site-associated DNA sequencing (RAD-seq) have been the most attractive and affordable options for genome-wide SNP discovery and genotyping (Poland and Rife, 2012; Pante et al., 2015). These methods rely on the use of endonucleases to produce short restriction fragments that, after various steps including adaptor ligation, size selection, and amplification, are sequenced providing the frame for SNP discovery (Deschamps et al., 2012; Kim et al., 2016a; Kim et al., 2016b). Despite the potentialities for developing numerous SNPs in comparison to other genotyping methods (e.g., microsatellites and arrays) and the advantage of a minor ascertainment bias, the main drawback of both GBS and RAD-seq is the uneven distribution of endonuclease cutter sites in the genome (Peterson et al., 2014). The untargeted detection reduces the possibility to identify polymorphisms within functionally relevant chromosomal regions. Indeed, single genes, gene families, promoters and enhancers, gene clusters, and non-coding genes are the genomic fractions that probably contain polymorphisms that are causative of, or tightly associated with, phenotypic variability.

To enable a more targeted approach on functional diversity, NuGEN Inc. (San Carlos, CA, USA) developed single primer enrichment technology (SPET, Patent US9650628B2) (Amorese et al., 2013), a novel customized and cost-effective technology based on Allegro Targeted Genotyping (Lovci et al., 2018). SPET offers the possibility to perform targeted genotyping of known polymorphisms and to discover new random polymorphic loci, thus combining the benefits of both arrays and RRLs (Scaglione et al., 2019). The technology relies on the previous identification of the sites to be sequenced holding the polymorphisms. Based on information gathered from reference genomes or transcriptomes, the target sites are selected, and short DNA probes of ~40 bases long are designed in the adjacent regions. In addition to sequencing of target sites, the probes enable the detection of closely linked novel polymorphisms within the area surrounding the target. Because it uses single primers, the panel design is straightforward, thus enabling a high capability of multiplexing. The tailored design allows SPET to have superior reproducibility and transferability when compared to the other RRL genotyping methods. In plants, SPET has been applied in maize (*Zea mays* L.), black poplar (*Populus nigra* L.) (Scaglione et al., 2019), oil palm (*Elaeis guineensis* Jacq.) (Herrero et al., 2020), cultivated and wild species of tomato and eggplant (*Solanum* spp.) (Barchi et al., 2019), and peach (*Prunus armeniaca* L.) (Baccichet et al., 2022), showing the power of this method for genotyping germplasm collections and crossing populations. Applications included population structure

analyses, phylogenetic investigations, high-density linkage map development, and association mapping analysis.

Cultivated lettuce (*Lactuca sativa* L.) is a commercially important crop belonging to the Compositae (Asteraceae), one of the largest angiosperm families comprising over 1,800 genera and 24,000 species (WFO Plant List, 2023). It is considered a main leafy vegetable, widely appreciated by consumers for the content of fibers and the low-calorie intake (Kim et al., 2016). It also represents a good source of vitamin C, iron, folate, and different health-beneficial bioactive compounds (Kim et al., 2016). Its production in 2020 was estimated to be 27.6 million tons on an area of 1.2 million hectares (FAOSTAT, 2023). The genus *Lactuca* comprises approximately 100 species, of which *L. sativa* and its wild progenitor *L. serriola*, both part of the primary gene pool, represent over 90% of the accessions held in genebanks (van Treuren et al., 2012). Cultivated accessions can be classified into diverse horticultural types based on the morphological characteristics of leaves and stems (Simko, 2009). Lettuce germplasm diversity has been explored using different molecular tools including microsatellites (Simko, 2009; Rauscher and Simko, 2013), anonymous and targeted PCR-based markers (van Treuren and van Hintum, 2009), arrays (Stoffel and van Leeuwen, 2012), and RRLs (Seki et al., 2020; Park et al., 2021; Park et al., 2022) to study genetic relationships within and among horticultural types. In the past few years, several genomic resources have been released including the first draft of the lettuce genome (cv. Salinas) (Reyes-Chin-Wo et al., 2017) and the resequencing of 445 accessions including cultivated lettuce and 12 wild *Lactuca* species (Wei et al., 2021), providing a useful source for assessing and exploiting germplasm diversity through novel marker discovery. The possibility to implement both genomic and phenotypic information in genome-wide association studies (GWAS) paves the way to dissect the genetic basis of complex traits. GWAS enable the identification of genomic regions underlying the variation of traits exploiting the ancient recombination events occurring in unrelated individuals (Huang and Han, 2014). The rapid advances of NGS technologies and computational pipelines make GWAS a powerful approach for candidate gene detection in crops. In lettuce, GWAS using different genotyping platforms for SNP discovery investigated agronomic traits (Kwon et al., 2013), resistances (Lu et al., 2014), and quality-related traits (Sthapit Kandel et al., 2020; Park et al., 2021).

In the present work, we describe the development of the first SPET panel in lettuce and its application for analyzing genomic diversity and population structure. A heterogeneous collection of 160 accessions of *L. sativa* and *L. serriola* was used as a proof of concept to validate the SPET assay. We further investigated the potentiality of SPET for candidate gene identification through GWAS in four main lettuce horticultural traits. The obtained results showed the strength of SPET for lettuce genomics.

# 2 Materials and methods

## 2.1 Plant material

Plant materials consisted of 155 accessions of *L. sativa* and 5 of the closely related wild species *L. serriola*, which were part of the germplasm panel established in the frame of the ECPGR European Evaluation (EVA) Lettuce Network (ECPGR, 2023). Plant materials originated from the germplasm collections of four institutions: the Institute for Plant Genetic Resources "K.Malkov" (Sadovo, Plovdiv district, Bulgaria), the Centre for Genetic Resources, the Netherlands (CGN, Wageningen, Netherlands), the Unité de Génétique et Amélioration des Fruits et Légumes, Plant Biology and Breeding, INRAe (GAFL, Avignon, Montfavet Cedex, France), and the Nordic Genetic Resource Center (Nordgen, Alnarp, Sweden). Genotypes encompassed cultivars, breeding materials, and landraces originating from a total of 10 different countries in Europe, America, and Asia. Different horticultural types were represented (Figure 1), including Butterhead (54), Iceberg (46), Cos or Romaine (17), Batavia or Summer/French Crisp (11), Crisp (10), Loose leaf (9), Oak leaf (4), Latin (3), and Lollo (1), as well as wild *L. serriola* (5) (also known as prickly lettuce). A detailed list with all available information on the assayed accessions is provided in Supplementary Table 1.

## 2.2 Single primer enrichment technology panel design

For probe design, a dataset including whole-genome resequencing data of 131 *L. sativa* accessions (Wei et al., 2021) was considered (Supplementary Table 2). Raw sequence data were retrieved from the FTP site of the China National Gene Bank Sequence Archive (CNSA) repository (Guo et al., 2020). Variants (SNP and INDEL separately) were selected by filtering those present in the dataset with a minimum allele count of 3 (i.e., one homozygous accession and one heterozygous or three heterozygous accessions). The lettuce reference genome (*L. sativa* cv Salinas V8) and its annotation were retrieved from https://lgr.genomecenter.ucdavis.edu/Home.php and all gene coordinates were extended by 5,000 bp upstream and 1,000 bp downstream. All selected genomic variants were intersected with these gene coordinates and labeled as gene-space variants. A panel of 50k target sites was then built by imposing a minimum distance of 3,000 bp for variants on the gene-space and a minimum distance of 200,000 bp in the intergenic regions. After two rounds of design, a final panel of 41,547 targets were successfully identified by unique probes. Each probe consisted of a 40-bp sequence. SNP calling was enabled 460 bp downstream of the probe.

## 2.3 DNA extraction, library preparation, and sequencing

Genomic DNA was isolated from young leaves of a single individual per accession using a NucleoSpin Plant II Mini kit (Macherey-Nagel GmbH & Co. KG., Düren, Germany. DNA concentration was measured using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Libraries were prepared using the "Allegro Targeted Genotyping" protocol from NuGEN Technologies (San Carlos, CA), using 10 ng/µl of DNA as input and following the manufacturer's instructions. Libraries were


FIGURE 1
*Lactuca sativa* horticultural types considered in this study. **(A)** EVA_Lsa_00156, Butterhead; **(B)** EVA_Lsa_00166, Batavia; **(C)** EVA_Lsa_00094, Cos; **(D)** EVA_Lsa_00150, Crisp; **(E)** EVA_Lsa_00114, Iceberg; **(F)** EVA_Lsa_00184, Latin; **(G)** EVA_Lsa_00196, Lollo; **(H)** EVA_Lsa_00206, Loose leaf; **(I)** EVA_Lsa_00174, Oak Leaf. Photos provided by Charlotte Aichholz and Tizian Zollinger.

quantified using the Qubit 2.0 Fluorometer, and their size was checked using the High-Sensitivity DNA assay from Bioanalyzer (Agilent technologies, Santa Clara, CA) or the High-Sensitivity DNA assay from Caliper LabChip GX (Caliper Life Sciences, Alameda CA). Libraries were quantified through qPCR using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA) and sequenced on the Illumina NovaSeq 6000 (Illumina, San Carlos, CA).

## 2.4 Sequence analysis and SNP detection

Demultiplexing of raw sequencing data and base calling (BCL files into FASTQ files) were performed with the Illumina bcl2fastq2 Conversion Software v2.20 (Illumina, San Carlos, CA). Read quality check and adapter trimming were carried out using ERNE v1.4.6 (Del Fabbro et al., 2013) and Cutadapt (Martin, 2011), both with default parameters. Alignment to the reference genome *L. sativa* cv Salinas V8 (Reyes-Chin-Wo et al., 2017) was done using the Burrows–Wheeler Aligner BWA-MEM v0.7.17 (Li and Durbin, 2009) with default parameters and selection of uniquely aligned reads (i.e., reads with a mapping quality >10). SNP calling was obtained using gatk-4.0 (DePristo et al., 2011) following the software best practices for germline short variant discovery. SNP calling was limited to the regions (460 bp) that were previously defined as downstream of each enrichment probe.

All analyses were implemented in GATK Best Practices v4.1.2.0 (Van der Auwera and O'Connor, 2020) and included the following steps: (i) per-sample variants calling on target regions using HaplotypeCaller with default parameters to create a GVCFs file for each sample; (ii) GVCFs consolidation across multiple samples using GenomicsDBImport with default parameters and target intervals in order to improve scalability and speed for further joint genotyping; (iii) joint genotyping using GenotypeGVCFs with default parameters to produce a set of joint-called variants; (iv) Selection of SNPs using SelectVariants and quality filtering of SNPs using VariantFiltration (filter expression used: QD< 2.0 || MQ< 40.0 || MQRankSum< −12.5). A 1,911,467 biallelic SNPs matrix was obtained. The extra filtration of the VCF was performed with bcftools by setting all data points with fewer than five reads in coverage to a missing data genotype (./.) and retaining only records where a minimum of 96 samples reported a coverage above 10 reads. In total, 835,426 SNPs were obtained. For downstream analysis, 81,531 SNP sites were retained with minor allele count =

3, max missing 0.5, minQ = 30, and minor allele frequency = 5%. VCFtools version 0.1.17 (Danecek et al., 2011) was used. Pattern of nucleotide diversity (p) was estimated in non-overlapping sliding windows with a size of 1 kbp in VCFtools. Functional annotation of the identified variants associated genes was performed using SnpEff (version 3.1) (Cingolani, 2022).

## 2.5 Genomic diversity analysis

Genetic diversity summary of the SNP matrix was performed by the Geno Summary tool implemented in Tassel v5.2.15 (Bradbury et al., 2007). Considering the biallelic nature of SNPs, expected heterozygosity according to Hardy–Weinberg equilibrium (*H*) was calculated according to the formula

$$H = 1 - \rho^2 - q^2$$

where $p$ and $q$ each represent the frequency of the different alleles for each SNP.

The polymorphic information content (PIC) was calculated according to the formula (Shete et al., 2000)

$$PIC = H - 2 \times \rho^2 \times q^2$$

Population structure was determined using the model-based ancestry estimation obtained with ADMIXTURE software (Alexander et al., 2015) with K ranging from 1 to 15. One thousand bootstrap replicates were run to estimate parameter standard errors. Tenfold cross-validation (CV) procedure with five iterations was performed, and CV scores were used to determine the best *K* value. Individuals were considered to belong to a specific *K* population if its membership coefficient (qi) was ≥0.5, whereas the genotypes with qi lower than 0.5 at each assigned *K* were considered as admixed. A neighbor-joining phylogenetic tree was built using the Jones–Taylor–Thornton (JTT) model with 1,000 bootstraps. Analyses were conducted in MEGA X software (Kumar et al., 2018). Principal component analysis (PCA) was performed in Tassel v5.2.15 and the biplot was drawn using the ggplot2 R package (Wickham, 2016).

## 2.6 Phenotypic evaluation

The phenotypic traits were surveyed across five locations (Eyragues, Avignon, France; La Ménitré, France; Les Evouettes, Port-Valais, Switzerland; Rheinau, Switzerland; and Thessaloniki, Greece) during the 2020–2022 spring seasons. Plants were grown in a randomized block design with three replicates. Field trials were conducted using the standard agricultural practices for the local area of cultivation. Four traits were assayed including (i) seed color (1 = white/cream, 2 = yellow, 3 = brown, 4 = black), (ii) outer leaf color before bolting stage (1 = yellow green, 2 = green, 3 = gray green, 4 = blue green, 5 = red green), (iii) leaf anthocyanin content before bolting stage (0 = absent, 3 = weak, 5 = medium, 7 = strong), and (iv) bolting time (number of days from sowing to bolting).

## 2.7 Genome-wide association analysis

Genome-wide association analysis was performed in 155 *L. sativa* genotypes. Six models were used including the general linear model (GLM) (Loley et al., 2013), the mixed linear model (MLM) (Zhang et al., 2010), the multi-locus mixed linear model (MLMM) (Segura et al., 2012), the compressed mixed linear model (CMLM) with population parameters previously defined (P3D) (Zhang et al. in 2010), the fixed and random model circulating probability unification model (FarmCPU) (Liu et al., 2016), and the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway model (BLINK) (Huang et al., 2019). All models included the population structure as a covariate. The kinship was estimated using the identity by state (IBS) for accounting relationships among individuals. Phenotypic data from independent experiments were implemented. The significance threshold for marker–trait association was determined after Bonferroni multiple test correction with genome-wide $\alpha = 0.05$. Considering 81,531 SNPs, the marker was considered significant when the *p*-value was less than 6.212 ($-\log_{10}P = 6.133 \times 10^{-7}$). GLM and CMLM were computed in Tassel v 5.2.82 (Bradbury et al., 2007). MLM, MLMM, FarmCPU, and BLINK were calculated with the GAPIT R package (Wang and Zhang, 2021). Manhattan and quantile–quantile (Q–Q) plots for GWAS results were produced using the R package CMplot. The chromosomal location of the genome-wide significantly associated SNPs was displayed using PhenoGram (https://ritchielab.org/software/phenogram). Significant association signals were checked for their physical position on the *L. sativa* (cv. Salinas) V8 genome. The information about predicted genes was downloaded from the Lettuce genome browser v8.0 (https://phytozome-next.jgi.doe.gov/jbrowse/). Underlying genes and their functions were determined according to Reyes-Chin-Wo et al. (2017).

# 3 Results

## 3.1 SPET array

Based on SNP data retrieved from 131 lettuce raw sequences (Wei et al., 2021) and on the alignment to reference genome *L. sativa* cv Salinas V8, 41,547 probes were designed, of which 1,707 (4.1%) were localized in intergenic regions and 39,840 (95.9%) within genes (Supplementary Table 3). The average coverage of the total set of probes was 77.1×; for those located within intergenic regions, 93.4×; and for those within genes, 76.6× (Supplementary Figure 1). The SPET panel showed an average distribution of one probe per 55.5 kilobase pair (kbp). Regarding inter-probe distance, 27% of the probes were more than 50 kbp apart, while the largest gap was 3.2 mega base pair (Mbp) on chromosome 3 (Figure 2). The sequencing of SPET libraries in the 160 study samples produced a total of 668,695,867 paired end raw reads corresponding to an average of 4,179,349 read pairs per sample ranging from 2,000 to 21 million and a mean depth of 79.7× (Supplementary Table 4). The mapping rate on the whole genome was on average 88%, and only eight samples had an average below 70% (Supplementary Figure 2).


FIGURE 2
Distribution of 41,547 SPET probes on the nine lettuce chromosomes. The number of SNPs is represented within 1 Mb window size. The horizontal axis shows the chromosome (Chr) length (Mb); each bar represents a chromosome, with Chr 1 at the top and Chr 9 at the bottom. The different colors depict SNP density following the gradient in the legend on the right.

By applying stringent filtering criteria, we identified 81,531 SNPs ranging from 5,291 on chromosome 6 to 13,920 on chromosome 4 (Table 1). SNPs were predominantly located within transcript regions, covering over 65% of the gene space in all chromosomes. SNP effect analysis showed that the majority of SNPs (88.08%) have a possible modifier effect, while the rest exhibited low (6.96%), moderate (4.82%), and high (0.14%) impacts (Supplementary Table 5). Within gene space, SNPs were mostly localized in upstream and downstream gene regions (27.45% and 16.27%, respectively). SNPs in exons and introns were 11.37% and 7.47%, respectively (Supplementary Figure 3). The average density corresponded to one SNP every 28.99 kbp across the nine chromosomes, ranging from 21.48 kbp on chromosome 1 to 36.35 kbp on Chr 6. Across the whole set, PIC values ranged from 0.033 to 0.375 (data not shown) with a mean of 0.240. The minimum average PIC value was encountered on Chr 6 (0.226), while the maximum value was found in Chr 2 (0.258). Chr 6 and Chr 2 exhibited the lowest and highest nucleotide diversity with $4.681^{e-4}$ and $6.459^{e-4}$, respectively. On average, heterozygosity was 0.292, reaching values above 0.300 only on chromosomes 2 and 5. The observed transitions/transversions ratio was 2.12 (Supplementary Figure 4A). In particular, among transition events, C > T and G > A were the most abundant (18.697% and 18.225%, respectively), whereas C > A and A > T abounded within transversion events (4.826% and 4.506%, respectively). The allele content of the SNP matrix was balanced, being on average represented for 70% by the four nucleotide bases in homozygosity state (Supplementary Figure 4B).

## 3.2 Genomic diversity and population structure

An admixture-based clustering model implemented in the software ADMIXTURE (Alexander et al., 2015) was used to infer the genetic structure of the studied germplasm. Using the entire SNP dataset, results of CV error suggested six different clusters (Supplementary Figure 5) representing the most likely number of subpopulations (*K*) (Figure 3).

The subpopulations reflected to some extent a differentiation based on cultivar typology rather than country of provenance (Figure 3; Supplementary Table 6). The first cluster (K1) grouped 21 accessions, mostly Iceberg and Cos lettuce types from Bulgaria. Butterhead were mostly grouped in clusters 2 (K2) and 6 (K6) and represented 62% and 67% of the total individuals within each cluster, respectively. The subpopulation 3 (K3) included several Batavia and Crisp types whereas Oak leaf types were included in cluster 4 (K4) together with Iceberg and Loose leaf types. Among the different cultivar types, Iceberg accessions were clustered in several subpopulations. *Lactuca serriola* accessions were grouped separately from the rest in a distinct group (K = 5). Thirty-two accessions belonging to 8 out of the 10 considered cultivar types were classified as admixed, as they showed values for the highest cluster membership coefficient (qi) lower than 0.5. The Fixation Index ($F_{ST}$) values, measuring the population (*K*) differentiation based on SNP data, are reported in Table 2.

The highest $F_{ST}$ values were found between K5 and the other subpopulations, thus confirming the differentiation of the wild *L. serriola* from the cultivated *L. sativa*. The lowest divergence was found between clusters 1 and 3 ($F_{ST}$ = 0.265) mostly comprising the same type of cultivars. Considering the average *q*-value at $K = 6$ (Figure 4), the analysis showed how among the most represented cultivars, iceberg types were included in five out of the six detected clusters while butterheads were included in clusters 2, 3, and 6. Batavia, Crisp, and Lollo as well as Loose and Oak leaf types were mostly represented by clusters 3 and 4, respectively. The average heterozygosity of the accessions was on average lower than 4% in all cultivated variety groups (Figure 5). Prickly lettuce accessions showed

TABLE 1 SNP number, distribution in intergenic and genic regions, average distance for each chromosome, polymorphic information content (PIC), nucleotide diversity ($\pi$), and heterozygosity ($H$).

| Chromosome | Chromosome length (bp) | Total SNPs | SNP in genic regions | SNP in intergenic regions | % genic SNP | Average SNP interdistance (kb)a | Max SNP interdistance (kb) | Average PIC | Average p | Average *H* |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 214,780,997 | 9,995 | 9,651 | 344 | 0.97 | 21.486 | 2,434.756 | 0.245 | 5.834E-04 | 0.299 |
| 2 | 217,124,359 | 9,560 | 9,149 | 411 | 0.96 | 22.714 | 1,961.823 | 0.258 | 6.459E-04 | 0.317 |
| 3 | 256,900,232 | 7,295 | 6,622 | 673 | 0.91 | 35.220 | 4,241.704 | 0.230 | 5.122E-04 | 0.279 |
| 4 | 377,162,472 | 13,920 | 12,904 | 1,016 | 0.93 | 27.092 | 1,945.504 | 0.237 | 5.407E-04 | 0.286 |
| 5 | 339,292,695 | 11,255 | 10,593 | 662 | 0.94 | 30.148 | 3,567.503 | 0.246 | 5.294E-04 | 0.301 |
| 6 | 192,650,122 | 5,291 | 4,934 | 357 | 0.93 | 36.347 | 4,532.669 | 0.226 | 4.681E-04 | 0.272 |
| 7 | 195,410,018 | 6,797 | 6,470 | 327 | 0.95 | 28.753 | 2,749.737 | 0.244 | 5.048E-04 | 0.297 |
| 8 | 309,580,090 | 10,614 | 10,055 | 559 | 0.95 | 29.164 | 2,714.974 | 0.237 | 5.069E-04 | 0.287 |
| 9 | 203,529,833 | 6,804 | 6,476 | 328 | 0.95 | 29.889 | 2,562.293 | 0.235 | 4.697E-04 | 0.285 |

an average heterozygosity of 4.69% with values ranging from 4.64% to 4.99%. The same trend was observed among different subpopulations based on admixture analysis (data not shown). Twelve accessions belonging to Butterhead (7) and Iceberg (5) exhibited heterozygosity higher than 5% with values up to 6.61% (Butterhead) and 10.08% (Iceberg). Only a single accession, representing the Cos horticultural type, showed a relatively high heterozygosity of 16.11%.

## 3.3 Genetic relationships among accessions

Phylogenetic clustering and PCA were performed to find patterns of genetic variation among accessions. The phylogenetic network using the neighbor-joining method was generally in agreement with Admixture analysis. Two main subpopulations were detected. Group I mostly included butterhead types (Figure 6A) from the clusters K2 and K6 (Figure 6B).

Group II consisted of several icebergs, loose leaf, and cos types from clusters K1, K3, and K4 (Figures 6A, B). All prickly lettuce (*L. serriola*) genotypes were grouped closely together according to the cluster K5. The distribution of the accessions in the PCA bi-plot graph corroborated population structure analysis highlighting, among *L. sativa* accessions, a clustering of butterhead types compared to the rest (Figure 7A). Prickly lettuce genotypes were grouped apart on the second component, thus confirming the observed subpopulation in the ancestry analysis. A slight differentiation between French and Bulgarian germplasm was observed. Interestingly, several close relationships were found between Italian and Bulgarian accessions. Although more admixtures were found when the geographical provenance was considered, a general differentiation was observed between germplasm retrieved from Western and Eastern Europe (Figure 7B).

## 3.4 Genome-wide association analysis

Genome-wide association scans using six models detected a total of 306 significant SNP–trait associations (STA) (Supplementary Table 7) distributed across all chromosomes except for chromosome 6. The majority of STA were detected for seed color and leaf anthocyanin content: 133 and 117, respectively. Fifty-eight percent of associations were identified with the GLM, whereas among the five multi-locus models used, MLM and CMLM highlighted the highest number of association signals. Only for bolting time was no association found with multivariate models, except FarmCPU. Considering all models, chromosomes 5, 7, and 9 held over 94% of the STA, showing further several colocalizations. Manhattan plots showing the associations, their chromosomal positions, and Bonferroni threshold are shown in Figure 8, and Q–Q plots for multi-model GWAS and physical position of STA are shown in Figure 9. Two main clusters were found for leaf anthocyanin content and outer leaf color in a 160-kb region at 86 Mbp on chromosome 5 as well as in a 2-Mbp region at 150–152 Mbp on chromosome 9. Furthermore, different significant SNPs were detected on chromosome 7 for seed color in a 3-Mbp region at 49–52 Mbp position and for bolting time at 164 Mbp.

10.3389/fpls.2023.1252777

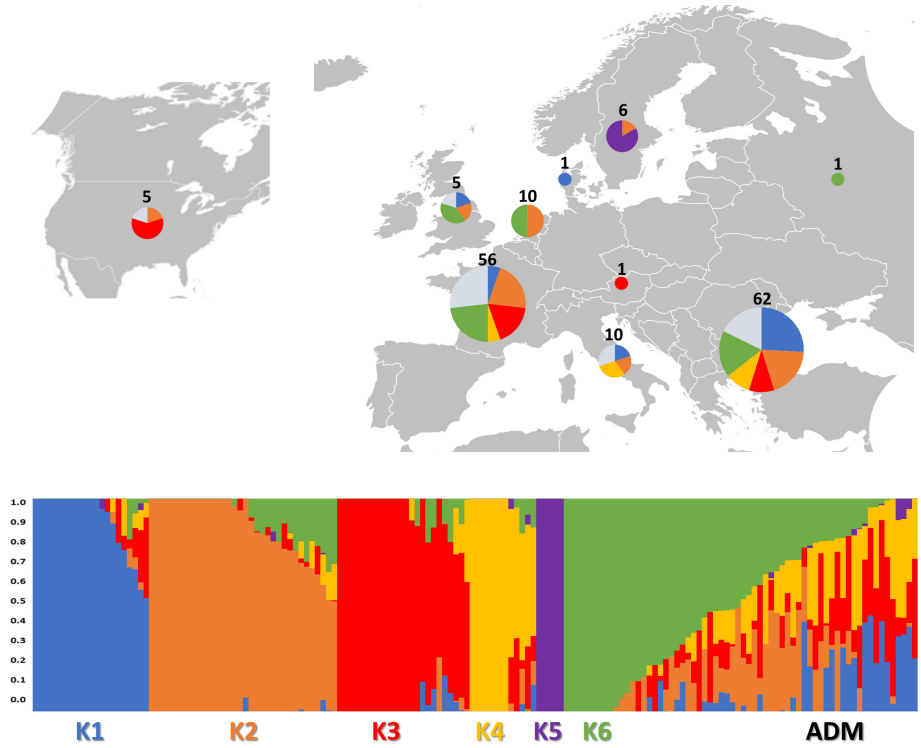



FIGURE 3
Genetic structure of the 160 study samples using 81,531 SNPs from SPET analysis. On the top: the geographical distribution of the germplasm studied and its subdivision according to the observed K groups; in the pie chart, the proportion of admixed accessions is indicated in ice blue color. On the bottom: bar plot describing the population admixture by the Bayesian approach. Each individual is represented by a thin vertical line, which is partitioned into K-colored segments whose length is proportional to the estimated membership coefficient (*q*). The population was divided into six (*K* = 6) groups according to the most informative *K* value.

TABLE 2 $F_{ST}$ values between populations inferred from a model-based ancestry estimation through the ADMIXTURE analysis.

| | K1 | K2 | K3 | K4 | K5 |
|---|---|---|---|---|---|
| **K2** | 0.409 | | | | |
| **K3** | 0.265 | 0.428 | | | |
| **K4** | 0.313 | 0.356 | 0.33 | | |
| **K5** | 0.676 | 0.684 | 0.686 | 0.592 | |
| **K6** | 0.38 | 0.334 | 0.397 | 0.34 | 0.682 |

In order to narrow down to potential GWAS hotspots, we considered the top-ranked SNPs within each model (Table 3). For seed color, five out of the six models detected the strongest signal at 50.40 Mbp on chromosome 7 in an intergenic region at 19.55 kb to an *Atp-dependent rna helicase DEAH5*. The percentage of phenotypic variation explained (PVE%) by each locus ranged from 0.03% to 45.26%. Only with the CMLM model was the highest peak found 147 kb downstream to the previous one (chromosome 7, 50.54 Mbp) and in correspondence to *Cytokinin dehydrogenase 3*. For leaf anthocyanin content, five models detected a robust association on chromosome 5 at 86.12 Mbp in correspondence to *Phototropin-2* with a PVE% ranging from 4.55% to 15.26%. In addition, all models detected the strongest STA on chromosome 9 at 152.91 Mbp within an *MLO like protein 11*. Also, for outer leaf color, the strongest associations were in both chromosome 5 and 9, at ~27 kbp distance from those identified for leaf anthocyanin content. For leaf color, a *Signal peptidase complex subunit 3B* was the candidate gene identified with GLM, CMLM, and FarmCPU on chromosome 5 at 86.15 Mbp. The three models exhibited a PVE% ranging from 2.78 to 20.32. Furthermore, all models detected the strong STA at 152.88 Mbp on chromosome 9 in correspondence to a *General transcription factor 3C polypeptide 6*, with a PVE% ranging from 10.46% to 47.65%.

For bolting time, only GLM, BLINK, and FarmCPU revealed the strongest STA in an intergenic region at 1.77 kbp from *FAR1-related sequence 10* located on chromosome 7 at 164.43 Mbp. The three models exhibited a PVE% ranging from 18.71% to 48.65%.

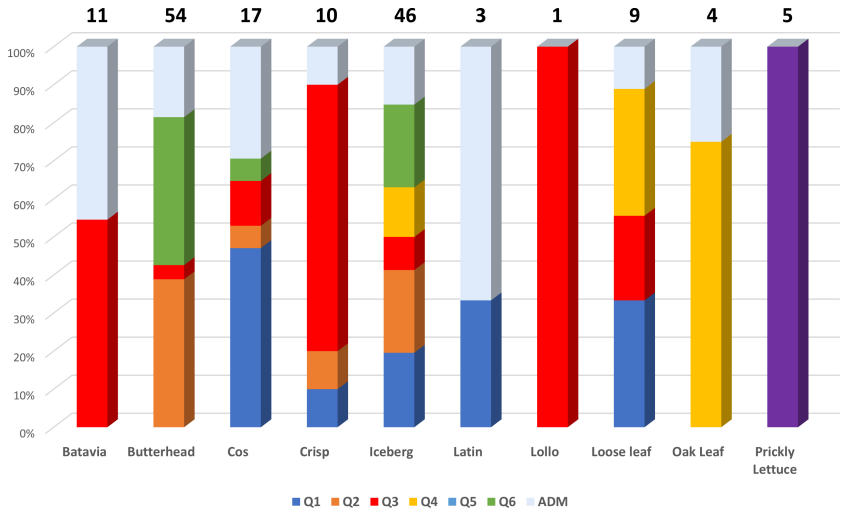



FIGURE 4
Stacked bar chart of the allele frequency based on Q membership coefficient at $K$ = 6. For each cultivar group, the number of accessions is indicated above each bar.

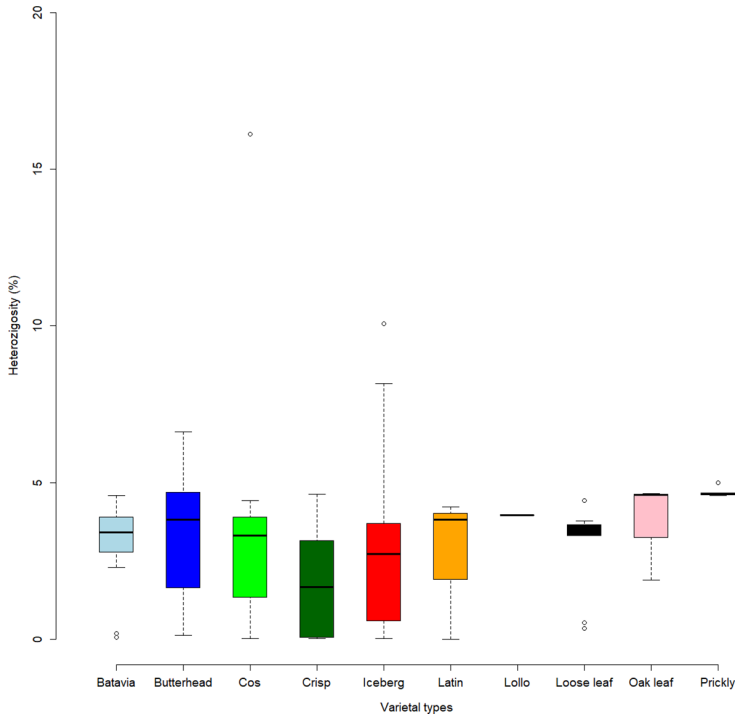



FIGURE 5
Heterozygosity level (in percentage) of the lettuce accessions. Box plots show median values and quartiles (first and third) of accessions considering the different varietal types.

# 4 Discussion

## 4.1 SPET development and genomic diversity

In this work, we investigated the effectiveness of SPET as a tool for high-throughput genotyping in lettuce. This method has been developed recently, but so far, very little information about how well it performs in plants is reported. To that end, we developed and validated a novel SNP panel enriched of intraspecific SNPs from 131 resequenced genomes and consisting of over 40,000 probes across the lettuce genome. The potentialities of SPET rely on the high-efficiency enrichment of targeted loci and the high scalability of up to thousands of probes in a single reaction (Scaglione et al., 2019).

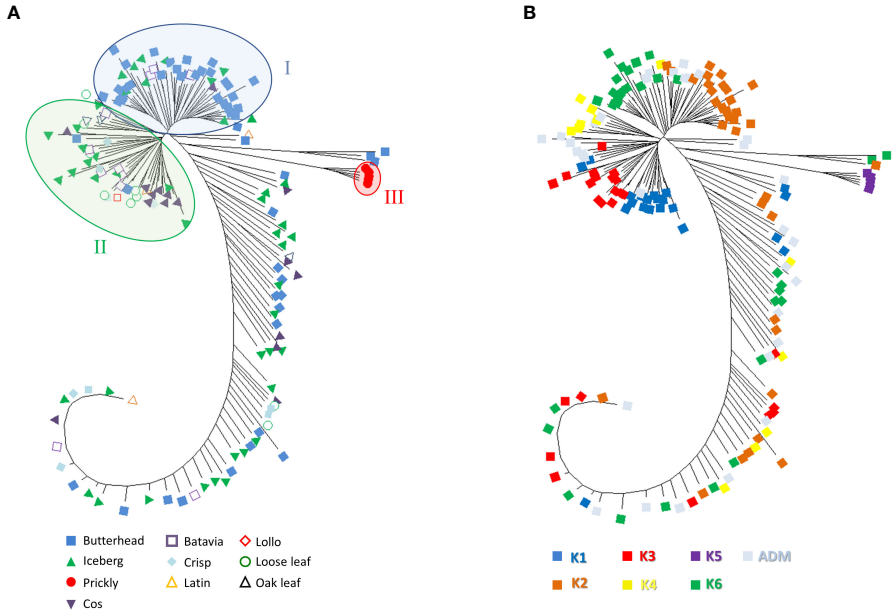



FIGURE 6
Neighbor-joining phylogenetic tree (radiation style) using 81,531 SNPs from SPET analysis. The evolutionary distances were computed using the Jones–Taylor–Thornton (JTT) model with 1,000 bootstraps. **(A)** Tree with annotated species and horticultural type. **(B)** Tree with annotated grouping revealed by the population structure analysis.

In addition, it offers the possibility of discovering novel SNPs by sequencing the genomic regions surrounding the target SNPs. Compared to other genotyping strategies for reducing genome complexity, this method offers full control of target sites, thus broadening the investigation of variation within genomic regions with a functional role. Furthermore, the possibility to detect SNPs within probe-defined regions improves reproducibility, thus enabling one to implement and/or compare genomic information from different genotyping experiments. Our main goal was to determine the applicability of SPET for assessing the diversity of a heterogeneous germplasm collection of lettuce including genotypes belonging to different horticultural types with diverse geographic origins. This work was done as part of the ECPGR European Evaluation Network (EVA) with the goal of improving the knowledge of crop genetic diversity and exploiting it to breed more resilient crops that can meet the major problems facing agriculture in the upcoming years (FAO, 2021; ECPGR, 2023). A more efficient use of crop diversity is essential for genetic improvement, management, and conservation of germplasm resources. The sequenced dataset comprised an average of 4 million SNPs per sample, which has been indicated to be adequate for processing several thousands of probes (Scaglione et al., 2019). Compared to the 25K SPET panel reported in peach (Baccichet et al., 2022) and the 5K SPET panel described for tomato, eggplant, and oil palm (Barchi et al., 2019; Herrero et al., 2020), the 40K SPET assay designed in lettuce provides a higher number of

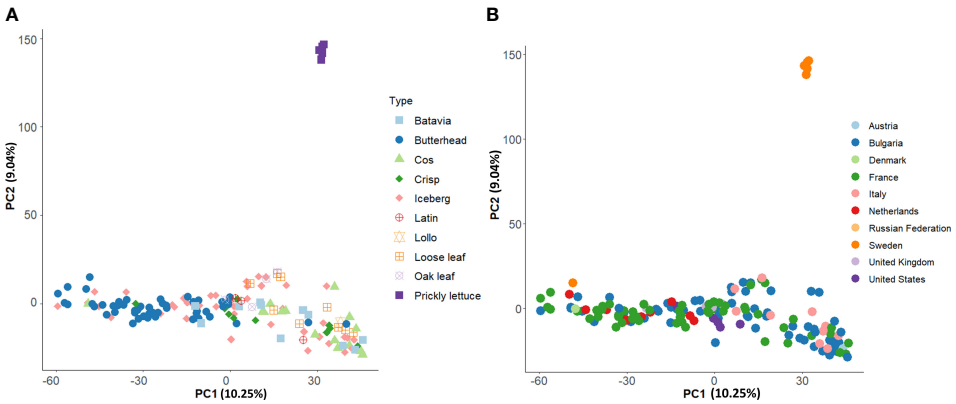



FIGURE 7
Loading plot in the first two components, showing the genomic diversity of the 160 studied accessions. The PCA was computed with 81,531 SNPs. **(A)** PCA with annotated species and horticultural types. **(B)** PCA with annotated country of origin.

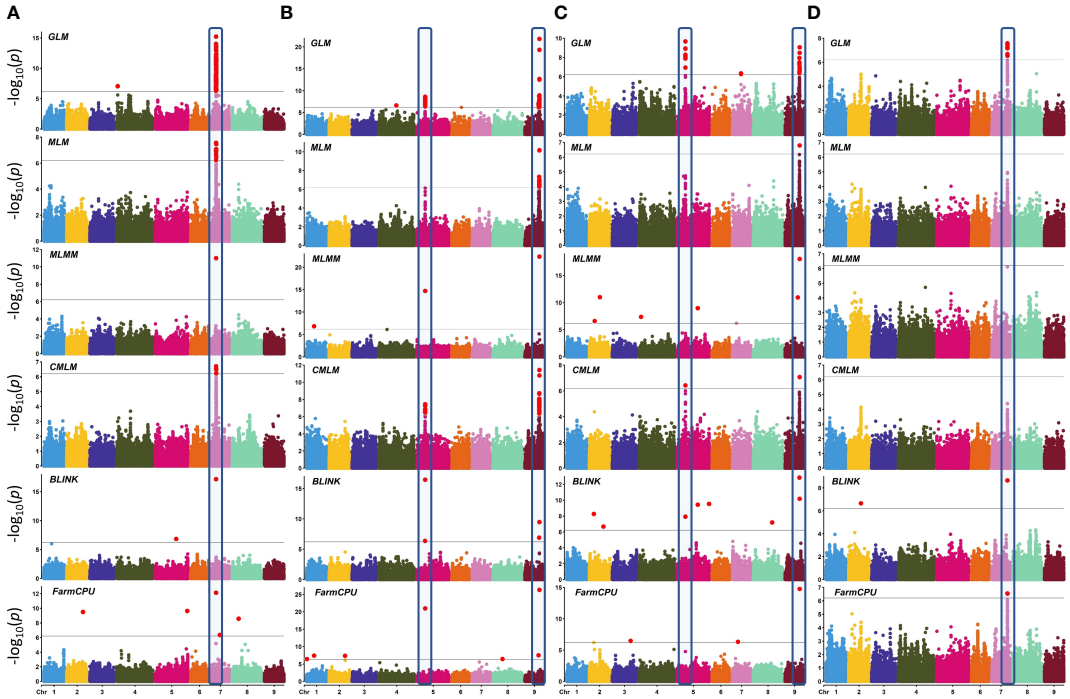



FIGURE 8
Manhattan plots showing SNP–trait associations (STA) in *L. sativa* using six multi-locus GWAS models. Four horticultural traits are shown: **(A)** seed color, **(B)** leaf anthocyanin content, **(C)** outer leaf color, and **(D)** time of beginning of bolting (bolting time). Analysis has been performed considering 81,531 SNPs on 155 accessions. The black horizontal line indicates a significant threshold (–log10 *p*-value) according to Bonferroni. The *X*-axis indicates the chromosome position. The STA repeatedly identified by three or more GWAS models are highlighted.

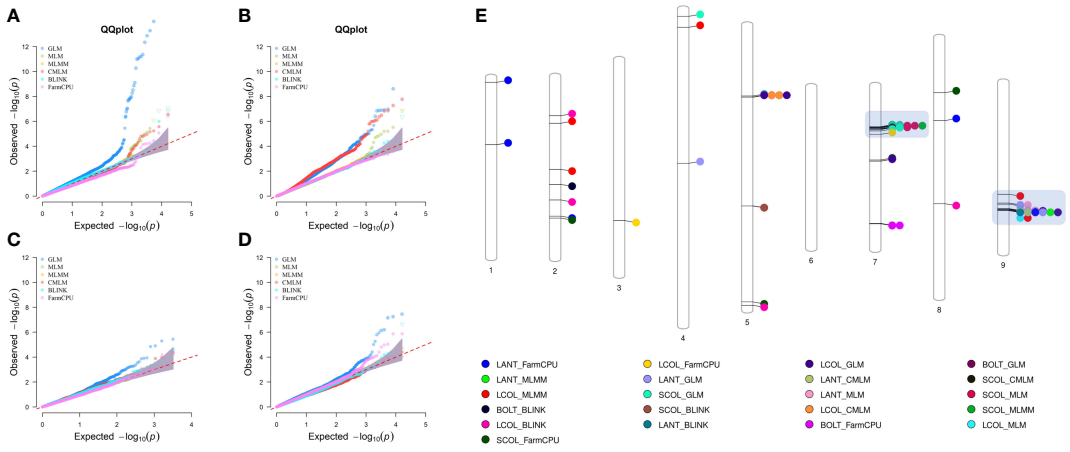



FIGURE 9
Quantile–quantile plots for six multi-locus GWAS models **(A–D)** and physical position of SNP–trait associations **(E)**. For QQ plots, the order of traits are as follows: **(A)** seed color, **(B)** leaf anthocyanin content, **(C)** outer leaf color, and **(D)** time of beginning of bolting. For chromosomes 7 and 9, the clusters of regions with most STA are highlighted.

SNPs covering up to 96% of gene-rich regions. Our findings demonstrated how effective SPET is compared to other genotyping techniques for detecting SNPs within coding regions. In fact, prior studies in lettuce utilizing genotyping by sequencing revealed that the proportion of SNP loci within genic regions ranged from 0.94% to 27.6% (Park et al., 2021; Park et al., 2022).

We detected over 80,000 high-quality polymorphisms that were analyzed to determine population ancestry, phylogenetic relationships, and principal components among the EVA lettuce accessions. The three approaches were complementary, thus supporting the interpretation of results. In agreement with earlier findings (Park et al., 2021; Park et al., 2022; Simko, 2009; Stoffel and van Leeuwen, 2012), no admixture was

TABLE 3 Robust associations detected with a multimodel GWAS for four horticultural traits in a germplasm collection of 155 cultivated lettuce accessions.

| Trait | Chromosome | Model* | Position^ | Major/ Minor allele | MAF+ | Minor allele effect | PVE# | Nearest candidate gene | Candidate gene annotation |
|---|---|---|---|---|---|---|---|---|---|
| Seed color | 7 | a,b,c,e, f | 50,400,650 | C/T | 0.25 | 0.48–0.80 | 0.03–45.26 | +19.55 kb | *Atp-dependent rna helicase DEAH5* |
| | 7 | d | 50,547,653 | G/T | 0.28 | 3.33 e-08 | 0.73 | 0.0 kb | *Cytokinin dehydrogenase 3* |
| Leaf anthocyanin content | 5 | a,c,d,e,f | 86,123,750 | T/A | 0.27 | 0.30–0.55 | 4.55–15.26 | 0.0 kb | *Phototropin-2* |
| | 9 | a,b,c,d,e,f | 152,909,707 | G/A | 0.22 | −1.12 to −0.5 | 3.99–23.70 | 0.0 kb | *MLO like protein 11* |
| Outer leaf color | 5 | a,d,f | 86,150,826 | T/A | 0.21 | 0.10–0.30 | 2.78–20.32 | 0.0 kb | *Signal peptidase complex subunit 3B* |
| | 9 | a,b,c,d,e,f | 152,883,490 | A/G | 0.26 | 0.45–0.70 | 10.46–47.65 | 0.0 kb | *General transcription factor 3C polypeptide 6* |
| Bolting time | 7 | a,e,f | 164,434,052 | A/G | 0.49 | −1.67 to −1.47 | 18.71–48.65 | −1.77 kb | *FAR1-related sequence 10* |

*a, GLM; b, MLM; c, MLMM; d, CMLM; e, BLINK; f, FarmCPU.
^ Position in base pair (bp) based on the v8 version of the reference genome assembly for L. sativa (cv. Salinas) (Reyes-Chin-Wo et al., 2017).
+ MAF, Minor frequency allele (range).
# PVE, Range of percentage variance explained.

found between *L. sativa* and *L. serriola*. Indeed, the accessions of the two species were clearly separated. This evidence promotes the potentiality of SPET for phylogenetic studies, as already observed in aubergine and tomato (Barchi et al., 2019). Population structure and phylogenetic analysis revealed the presence of five distinct subpopulations within *L. sativa* with a variable degree of mixture across cultivar groups, confirming previous studies using both short-read genotyping-based techniques (Park et al., 2021; Park et al., 2022; Stoffel and van Leeuwen, 2012) and microsatellites (Rauscher and Simko, 2013). This could be related to the fact that in lettuce breeding, different horticultural types may be used in the pedigree scheme. In the collection assayed, we found a major clustering of butterhead genotypes when compared to the rest. This tendency contrasted with Park and colleagues (2021), who reported instead a greater separation of iceberg accessions from the other types in a collection of 441 individuals. Despite finding a slight differentiation according to geographical provenance, the effect due to the composition of the diversity panel assayed in terms of horticultural types and represented countries must be considered. Furthermore, for breeding and research materials, the reported origin often matches the places where the selection is carried out, thus providing an additional confounding effect. Several factors could affect the subpopulations enclosed in germplasm collections, such as the management practices occurring in the holding genebanks (e.g., level of heterozygosity retained and duplications), the areas of sampling of materials, or the biological status of accessions. Iceberg types investigated by Park et al. (2021) were mostly patented lines from the USDA, whereas we assayed mostly breeding materials, thus suggesting the presence of accessions still under development. The possibility to discover *de novo* polymorphisms free from any sequencing ascertainment bias and at affordable costs commensurable to other next-generation genotyping methodologies designates SPET as an efficient tool for population genomic analysis in lettuce.

## 4.2 Genome-wide association analysis

The advances in genomics and cutting-edge genotyping technology have contributed to the growing availability of large-scale genotypic data of germplasm resources for various crops. The analysis of the genetic underpinnings of complex traits used in GWAS has benefited greatly from the ability to link phenotypic data to genomic sequence data. GWAS has proven to be an effective method for finding genetic variations that are significantly more common for a specific phenotype in unrelated individuals (Xiao et al., 2022). Owing to the greater number of recombination events occurring in natural populations, the advantage over bi-parental mapping populations depends on a larger genetic base to exploit and on higher map resolution (Han et al., 2020). Over the past years, the GWAS computing efficiency has been improved by developing different multivariate models that consider the family kinship inference and population structure covariates to enhance the power of associations and decrease the rate of false positives (Wang and Zhang, 2021). GWAS has been performed with the aim of investigating the potentiality of the SPET panel for candidate gene detection. To that end, we focused on four main agronomic traits driving the selection of cultivated lettuce cultivars and underlying market and consumer preferences. To test the most likely candidate regions underpinning the variation of the considered traits, different models were implemented. As expected, the GLM detected the highest number of STA in all traits, although this model accumulates several false positives, which are eliminated by incorporating additional correcting factors involving a multi-dimensional genome scan able to simultaneously estimate all marker effects (Wang et al., 2014: Chaurasia et al., 2021). By combining multivariate models, we identified seven candidate regions across chromosomes 5, 7, and 9 for the assayed traits. For

seed color, the STA found on chromosome 7 confirmed a previous investigation reporting three associations in a 12-Mbp region spanning 69.87 Mbp to 80.63 Mbp (Kwon et al., 2013). We better refined the position at 50.40 Mbp near *DEAH5*, an ATP-dependent RNA helicase involved in abscisic acid and stress responses in the acquisition of embryogenic competence (Almeida et al., 2020). The *CYTOKININ DEHYDROGENASE 3* detected within the association may regulate cell division as well as a large number of developmental events in plants (Schmülling et al., 2003). The two candidates may therefore play a role in seed coat development and color.

The position of STA located on chromosomes 5 and 9 for leaf color traits agreed with previous studies (Zhang et al., 2017; Su et al., 2020; Wei et al., 2021). On chromosome 5, Zhang et al. (2017) reported the lead SNPs for leaf color at less than 150 bp (86,123,627, 86,123,633, and 86,123,651) from the top association for leaf anthocyanin color. In the same study, the association on chromosome 9 was in the same region at 17.45 kb (152,892,248) from the top- ranked STA found in this study. These regions are reported to harbor two genes *RLL2* (Red Lettuce Leaves 2) and *ANS* (Anthocyanin Synthase) that encode key enzymes for anthocyanin biosynthesis. We found four main candidate genes. *PHOTOTROPIN 2* and *MLO LIKE PROTEIN 11* both play a key role in leaf development and physiology. *PHOTOTROPIN 2* is primarily involved in the reception of light direction in the blade and has been demonstrated to promote leaf expansion and flattening (Legris et al., 2021). In *Pistacia chinensis*, *PHOTOTROPIN 2* has been reported to be involved in the signal transduction for anthocyanin accumulation during leaf coloration in autumn (Song et al., 2021), whereas in octaploid strawberry, it was involved in anthocyanin accumulation in strawberry fruits (Kadomura-Ishikawa et al., 2013).

The *MLO LIKE PROTEIN 11* is part of the large family of proteins that regulates pathogen defense and leaf cell death (Pozharskiy et al., 2022). No previous report indicates any function of *MLO LIKE PROTEIN 11* in leaf color. A general transcription factor (*3C POLYPEPTIDE 6*) was found to be involved in outer leaf color on chromosome 9. In plants, transcription factors regulate secondary metabolism (Vom Endt et al., 2002) and are potential candidates for plant organ pigmentation (Ban et al., 2007; Zhou et al., 2014; Su et al., 2020).

The strongest signals found for bolting time at 164.43 Mb on chromosome 7 confirmed previous evidence. Indeed, several studies consistently supported the importance of chromosome 7 for lettuce flowering control (Kwon et al., 2013; Sthapit Kandel et al., 2020; Lee et al., 2021; Rosental et al., 2021). Despite the exact comparisons of the candidate region not always being possible, owing to the different marker system used (Han et al., 2021), our study supports whole-genome resequencing data findings (Wei et al., 2021), which detected a strong association at 164.5 Mbp in correspondence to *PHYTOCHROME C* involved in delaying of flowering. With the same effect, the strong STA found in the present study was near *FAR1* (FAR-RED IMPAIRED RESPONSE 1), a component of the phytochrome A and putatively involved in regulating light control during the developmental stage (Siddiqui et al., 2016; Liu et al., 2020). *FAR1* directly activates the expression of the evening gene *ELF4* that plays a key role in the circadian flowering clock. In Arabidopsis, it negatively regulates flowering time in synergy with other FRS (FAR-Related Sequence) and FRF (FRS-Related Factor) genes (Ma and Li, 2018). The variation of *FAR1* expression has also been reported to regulate shoot growth and flowering time in roses.

The creation of a novel SPET assay in lettuce was described in this work, and its potential for genetic diversity and GWAS research was demonstrated by comparing the results with earlier discoveries using different genotyping technologies. Additional research could weigh the benefits and drawbacks of SPET in comparison to whole-genome short and long read sequencing.

## 5 Conclusion

Here, we presented SPET as an efficient method combining the properties of random complexity reduction techniques and arrays, allowing us to choose a set of gene-associated targeted markers for the accurate characterization of lettuce germplasm. The combination of population ancestry and phylogenetic approaches proved to be effective to better understand the genomic structure of lettuce genotypes. It is evident that the observed diversity patterns reflect the varietal composition of the collection and, to a minor extent, the geographical origin, which can be assumed primary factors underlying the diversification. Given the high marker density, the SPET panel has been used as a proof of concept for genome-wide association analysis to identify genomic regions underpinning the variation of main agronomic traits in lettuce. We confirmed previous findings, refined the genomic position of trait loci, and demonstrated the power of SPET for GWAS. These results will be useful for breeding and selection in lettuce. Further applications may include analysis of genetic relationships among species, management of genebank collections, and genetic fingerprinting for plant variety protection as well as GWAS for other additional important traits in lettuce.

## 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 authors.

## Author contributions

SG and MB conceived and coordinated the project. PT analyzed genomic data and prepared the draft of the manuscript. MB, DP, IK, CV, AL, GB, CA, and TZ performed phenotyping trials. DS, MB, RvT, and SG jointly designed the SPET array. DS and PT performed bioinformatic analysis. All authors contributed to the article and approved the submitted version.

## Funding

The authors are grateful for the financial support for this work by the German Federal Ministry of Food and Agriculture, grant

GenRes 2019-2 to ECPGR, which allowed the implementation of the EVA networks.

## Conflict of interest

DS was employed by the company IGA Technology Services Srl. MB was employed by the company ISI Sementi SpA. DP was employed by the company Limagrain - Vilmorin-Mikado. AL and GB were employed by the company Gautier Semences. CA was employed by the company Sativa Rheinau AG. TZ was employed by the company Zollinger Conseilles Sarl.

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1252777/full#supplementary-material

## References

Alexander, D. H., Shringarpure, S. S., Novembre, J., and Lange, K. L. (2015). *Admixture 1.3 software manual* (Los Angeles: UCLA Human Genetics Software Distribution).

Almeida, F. A., Passamani, L. Z., Santa-Catarina, C., Mooney, B. P., Thelen, J. J., and Silveira, V. (2020). Label-free quantitative phosphoproteomics reveals signaling dynamics involved in embryogenic competence acquisition in sugarcane. *J. Proteome Res.* 19 (10), 4145–4157. doi: 10.1021/acs.jproteome.0c00652

Amorese, D., Armour, C., and Kurn, N. (2013). Compositions and methods for targeted nucleic acid sequence enrichment and high efficiency library regeneration *US Patent* US9650628B2.

Baccichet, I., Chiozzotto, R., Scaglione, D., Bassi, D., Rossini, L., and Cirilli, M. (2022). Genetic dissection of fruit maturity date in apricot *(P. Armeniaca* L.) through a Single Primer Enrichment Technology (SPET) approach. *BMC Genomics* 23 (1), 1–16. doi: 10.1186/s12864-022-08901-1

Ban, Y., Honda, C., Hatsuyama, Y., Igarashi, M., Bessho, H., and Moriguchi, T. (2007). Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. *Plant Cell Physiol.* 48, 958–970. doi: 10.1093/pcp/pcm066

Barchi, L., Acquadro, A., Alonso, D., Aprea, G., Bassolino, L., Demurtas, O. C., et al. (2019). Single Primer Enrichment Technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01005

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

Chaurasia, S., Singh, A. K., Kumar, A., Songachan, L. S., Yadav, M. C., Kumar, S., et al. (2021). Genome-wide Association Mapping Reveals Key Genomic Regions for Physiological and Yield-Related Traits under Salinity Stress in Wheat (*Triticum aestivum* L.). *Genomics* 113 (5), 3198–3215. doi: 10.1016/j.ygeno.2021.07.014

Cingolani, P. (2022). Variant annotation and functional prediction: SnpEff. *Methods Mol. Biol. Clifton NJ* 2493, 289–314. doi: 10.1007/978-1-0716-2293-3_19

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

Del Fabbro, C., Scalabrin, S., Morgante, M., and Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PloS One* 8, e85024. doi: 10.1371/journal.pone.0085024

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806

Deschamps, S., Llaca, V., and May, G. D. (2012). Genotyping-by-sequencing in plants. *Biol.* 1, 460–483. doi: 10.3390/biology1030460

ECPGR (2023). Available at: https://www.ecpgr.cgiar.org/european-evaluation-network-eva/eva-networks/lettuce (Accessed July 1, 2023).

FAO (2021) *FAO's Strategic Framework 2022-31*. Available at: https://www.fao.org/3/cb7099en/cb7099en.pdf (Accessed July 1, 2023).

FAOSTAT (2023) *FAOSTAT*. Available at: http://www.fao.org/faostat/en/ (Accessed July 1, 2023).

Guo, X., Chen, F., Gao, F., Li, L., Liu, K., You, L., et al. (2020). CNSA: a data repository for archiving omics data. *Database (Oxford)* 2020, baaa055. doi: 10.1093/database/baaa055

Han, Z., Hu, G., Liu, H., Liang, F., Yang, L., Zhao, H., et al. (2020). Bin-based genome-wide association analyses improve power and resolution in QTL mapping and identify favorable alleles from multiple parents in a four-way MAGIC rice population. *Theor. Appl. Genet.* 133, 59–71. doi: 10.1007/s00122-019-03440-y

Han, R., Truco, M. J., and Lavelle, D. O. (2021). Michelmore, R.W. @ a composite analysis of flowering time regulation in lettuce. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.632708

Herrero, J., Santika, B., Herrán, A., Erika, P., Sarimana, U., Wendra, F., et al. (2020). Construction of a high density linkage map in Oil Palm using SPET markers. *Sci. Rep.* 10, 1–9. doi: 10.1038/s41598-020-67118-y

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8. doi: 10.1093/gigascience/giy154

Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65, 531–551. doi: 10.1146/annurev-arplant-050213-035715

Kadomura-Ishikawa, Y., Miyawaki, K., Noji, S., and Takahashi, A. (2013). Phototropin 2 is involved in blue light-induced anthocyanin accumulation in Fragaria × ananassa fruits. *J. Plant Res.* 126, 847–857. doi: 10.1007/s10265-013-0582-2

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L.-S., and Paterson, A. H. (2016a). Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* 242, 14–22. doi: 10.1016/j.plantsci.2015.04.016

Kim, M. J., Moon, Y., Tou, J. C., Mou, B., and Waterland, N. L. (2016b). Nutritional value, bioactive compounds and health benefits of lettuce (*Lactuca sativa* L.). *J. Food Compos. Anal.* 49, 19–34. doi: 10.1016/j.jfca.2016.03.004

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kwon, S., Simko, I., Hellier, B., Mou, B., and Hu, J. (2013). Genome-wide association of 10 horticultural traits with expressed sequence tag-derived SNP markers in a collection of lettuce lines. *Crop J.* 1, 25–33. doi: 10.1016/j.cj.2013.07.014

Lee, N., Fukushima, K., Park, H. Y., and Kawabata, S. (2021). QTL analysis of stem elongation and flowering time in lettuce using genotyping-by-sequencing. *Genes* 12, 947. doi: 10.3390/genes12060947

Legris, M., Szarzynska-Erden, B. M., Trevisan, M., Allenbach Petrolati, L., and Fankhauser, C. (2021). Phototropin-mediated perception of light direction in leaves regulates blade flattening. *Plant Physiol.* 187 (3), 1235–1249. doi: 10.1093/plphys/kiab410

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

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet* 12 (2), e1005767. doi: 10.1371/journal.pgen.1005767

Liu, Y., Ma, M., Li, G., Yuan, L., Xie, Y., Wei, H., et al. (2020). Transcription factors FHY3 and FAR1 regulate light-induced CIRCADIAN CLOCK ASSOCIATED1 gene expression in Arabidopsis. *Plant Cell* 32, 1464–1478. doi: 10.1105/tpc.19.00981

Loley, C., König, I. R., Hothorn, L., and Ziegler, A. (2013). A unifying framework for robust association testing, estimation, and genetic model selection using the generalized linear model. *Eur. J. Hum. Gen.* 21 (12), 1442–1448. doi: 10.1038/ejhg.2013.62

Lovci, M. T., Bruns, S. C., Eide, M., Sherlin, L., and Heath, J. D. (2018). "Nugen's allegro™ Targeted genotyping: an accurate and cost-effective sequencing workflow for any genome," in *Plant and Animal Genome XXVI Conference* (PAG, San Diego).

Lu, H., Hu, J., and Kwon, S. J. (2014). Association analysis of bacterial leaf spot resistance and SNP markers derived from expressed sequence tags (ESTs) in lettuce (*Lactuca sativa* L). *Mol. Breed* 34, 997–1006. doi: 10.1007/s11032-014-0092-5

Ma, L., and Li, G. (2018). FAR1-RELATED SEQUENCE (FRS) and FRS-RELATED FACTOR (FRF) family proteins in arabidopsis growth and development. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00692

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Onda, Y., and Mochida, K. (2016). Exploring genetic diversity in plants using high-throughput sequencing techniques. *Curr. Genomics* 17, 356–365. doi: 10.2174/1389202917666160331202742

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S. C., Boisselier, M. C., et al. (2015). Use of RAD sequencing for delimiting species. *Heredity* 114, 450–459. doi: 10.1038/hdy.2014.105

Park, J. S., Kang, M. Y., Shim, E. J., Oh, J. H., Seo, K. I., Kim, K. S., et al. (2022). Genome-wide core sets of SNP markers and Fluidigm assays for rapid and effective genotypic identification of Korean cultivars of lettuce (*Lactuca sativa* L.). *Hortic. Res.* 9, 1–15. doi: 10.1093/hr/uhac119

Park, S., Kumar, P., Shi, A., and Mou, B. (2021). Population genetics and genome-wide association studies provide insights into the influence of selective breeding on genetic variation in lettuce. *Plant Genome* 14, 20086. doi: 10.1002/tpg2.20086

Peterson, G. W., Dong, Y., Horbach, C., and Fu, Y. B. (2014). Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity* 6, 665–680. doi: 10.3390/d6040665

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005

Pozharskiy, A., Kostyukova, V., Nizamdinova, G., Kalendar, R., and Gritsenko, D. (2022). MLO proteins from tomato (*Solanum lycopersicum* L.) and related species in the broad phylogenetic context. *Plants* 11 (12), 1588. doi: 10.3390/plants11121588

Rauscher, G., and Simko, I. (2013). Development of genomic SSR markers for fingerprinting lettuce (Lactuca sativaL.) cultivars and mapping genes. *BMC Plant Biol.* 13, 11. doi: 10.1186/1471-2229-13-11

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., et al. (2017). Genome assembly with in *vitro* proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* 8, 14953. doi: 10.1038/ncomms14953

Rosental, L., Still, D. W., You, Y., Hayes, R. J., and Simko, I.. (2021). Mapping and identification of genetic loci affecting earliness of bolting and flowering in lettuce. *Theor. Appl. Genet.* 134, 3319–3337. doi: 10.1007/s00122-021-03898-9

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830. doi: 10.1038/ng.2314

Scaglione, D., Pinosio, S., Marroni, F., Centa, E., Di Fornasiero, A., Magris, G., et al. (2019). Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. *Ann. Bot.* 124 (4), 543–551. doi: 10.1093/aob/mcz054

Schmülling, T., Werner, T., Riefler, M., Krupková, E., and Bartrina y Manns, I. (2003). Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, Arabidopsis and other species. *J. Plant Res.* 116, 241–252. doi: 10.1007/s10265-003-0096-4

Seki, K., Komatsu, K., Hiraga, M., Tanaka, K., Uno, Y., and Mastumura, H. (2020). Identification of two QTLs for resistance to Fusarium wilt race 1 in lettuce (*Lactuca sativa* L.). *Euphytica* 216, 174. doi: 10.1007/s10681-020-02713-8

Shete, S., Tiwari, H., and Elston, R. C. (2000). On estimating the heterozygosity and polymorphism information content value. *Theor. Popul. Biol.* 57, 265–271. doi: 10.1006/tpbi.2000.1452

Siddiqui, H., Khan, S., Rhodes, B. M., and Devlin, P. F. (2016). FHY3 and FAR1 act downstream of light stable phytochromes. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00175

Simko, I. (2009). Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). *J. Heredity* 100 (2), 256–262. doi: 10.1093/jhered/esn072

Song, X., Duan, X., Chang, X., Xian, L., Yang, Q., and Liu, Y. (2021). Molecular and metabolic insights into anthocyanin biosynthesis during leaf coloration in autumn. *Env. Exp. Bot.* 190, 104584. doi: 10.1016/j.envexpbot.2021.104584

Sthapit Kandel, J., Peng, H., Hayes, R. J., Mou, B., and Simko, I. (2020). Genome-wide association mapping reveals loci for shelf life and developmental rate of lettuce. *Theor. Appl. Genet.* 133, 1947–1966. doi: 10.1007/s00122-020-03568-2

Stoffel, K., and van Leeuwen, H. (2012). Kozik, A. et al.(2012) Development and application of a 6.5 million feature Affymetrix Genechip® for massively parallel discovery of single position polymorphisms in lettuce (*Lactuca* spp.). *BMC Genomics* 13, 185. doi: 10.1186/1471-2164-13-185

Su, W., Tao, R., Liu, W., Yu, C., Yue, Z., He, S., et al. (2020). Characterization of four polymorphic genes controlling red leaf colour in lettuce that have undergone disruptive selection since domestication. *Plant Biotechnol. J.* 18, 479–490. doi: 10.1111/pbi.13213

Tripodi, P. (2022). Next generation sequencing technologies to explore the diversity of germplasm resources: Achievements and trends in tomato. *Comput. Struct. Biotechnol. J.* 6250–6258. doi: 10.1016/j.csbj.2022.11.028

Van der Auwera, G. A., and O'Connor, B. D. (2020). Genomics in the cloud: using Docker, GATK, and WDL in Terra. *O'Reilly Media*. pp 1–440.

Van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., et al. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* 5 (3), 247–252. doi: 10.1038/nmeth.1185

van Treuren, R., Coquin, P., and Lohwasser, U. (2012). Genetic resources collections of leafy vegetables (lettuce, spinach, chicory, artichoke, asparagus, lamb's lettuce, rhubarb and rocket salad): composition and gaps. *Genet. Resour Crop Evol.* 59, 981–997. doi: 10.1007/s10722-011-9738-x

van Treuren, R., and van Hintum, T. J. (2009). Comparison of anonymous and targeted molecular markers for the estimation of genetic diversity in *ex situ* conserved Lactuca. *Theor. Appl. Genet.* 119, 1265–1279. doi: 10.1007/s00122-009-1131-1

Van Treuren, R., and van Hintum, T. (2014). Next-generation genebanking: plant genetic resources management and utilization in the sequencing era. *Plant Genet. Resour.* 12, 298–307. doi: 10.1017/S1479262114000082

Vom Endt, D., Kijne, J. W., and Memelink, J. (2002). Transcription factors controlling plant secondary metabolism: what regulates the regulators? *Phytochemistry* 61 (2), 107–114. doi: 10.1016/S0031-9422(02)00185-1

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., and Zhang, Z. (2014). A SUPER powerful method for genome wide association study. *PloS One* 9, e107684. doi: 10.1371/journal.pone.0107684

Wang, J., and Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genom. Proteomics Bioinf.* 19, 629–640. doi: 10.1016/j.gpb.2021.08.005

Wei, T., van Treuren, R., Liu, X., Zhang, Z, Chen, J, Liu, Y, et al. (2021). Whole-genome resequencing of 445 Lactuca accessions reveals the domestication history of cultivated lettuce. *Nat. Genet.* 53, 752–760. doi: 10.1038/s41588-021-00831-0

Wendt, F. R., and Novroski, N. M. (2019). Identity informative SNP associations in the UK Biobank. Forensic Science International. *Genetics* 42, 45–48. doi: 10.1016/j.fsigen.2019.06.007

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis* (NY, USA: Springer: New York).

World Flora Online Plant List (2023). Available at: https://wfoplantlist.org/plant-list/taxon/wfo-7000000146-2022-12 (Accessed July, 1, 2023).

Xiao, Q., Bai, X., Zhang, C., and He, Y. (2022). Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *J. Adv. Res.* 35, 215–230. doi: 10.1016/j.jare.2021.05.002

You, Q., Yang, X., Peng, Z., Xu, L., and Wang, J. (2018). Development and applications of a high throughput genotyping tool for polyploid crops: single nucleotide polymorphism (SNP) array. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00104

Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., et al. (2017). RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nat. Commun.* 8 (1), 2264–2312. doi: 10.1038/s41467-017-02445-9

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Zhou, Y., Zhou, H., Lin-Wang, K., Vimolmangkang, S., Espley, R. V., Wang, L., et al. (2014). Transcriptome analysis and transient transformation suggest an ancient duplicated MYB transcription factor as a candidate gene for leaf red coloration in peach. *BMC Plant Biol.* 14, 388. doi: 10.1186/s12870-014-0388-y

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 September 2023
DOI 10.3389/fpls.2023.1222414

Check for updates

OPEN ACCESS

EDITED BY
Eleni Tani,
Agricultural University of Athens, Greece

REVIEWED BY
Sofia D. Carvalho,
Independent Researcher, Laramie, WY, United States
Georgios Liakopoulos,
Agricultural University of Athens, Greece

*CORRESPONDENCE
Jiajia Luo
luojiajia@catas.cn;
13909481919@126.com

†These authors have contributed equally to this work

RECEIVED 14 May 2023
ACCEPTED 22 August 2023
PUBLISHED 07 September 2023

CITATION
Cai Z, Wang X, Xie Z, Wen Z, Yu X, Xu S, Su X and Luo J (2023) Light response of gametophyte in *Adiantum flabellulatum*: transcriptome analysis and identification of key genes and pathways.
*Front. Plant Sci.* 14:1222414.
doi: 10.3389/fpls.2023.1222414

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

# Light response of gametophyte in *Adiantum flabellulatum*: transcriptome analysis and identification of key genes and pathways

Zeping Cai[1†], Xiaochen Wang[1†], Zhenyu Xie[1], Zhenyi Wen[2], Xudong Yu[3], Shitao Xu[3], Xinyu Su[1] and Jiajia Luo[4*]

[1]Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education, College of Forestry, Hainan University, Haikou, Hainan, China, [2]College of Ecology and Environment, Hainan University, Haikou, Hainan, China, [3]Key Laboratory of Germplasm Resources Biology of Tropical Special Ornamental Plants of Hainan Province, College of Forestry, Hainan University, Haikou, China, [4]Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

Light serves not only as a signaling cue perceived by plant photoreceptors but also as an essential energy source captured by chloroplasts. However, excessive light can impose stress on plants. Fern gametophytes possess the unique ability to survive independently and play a critical role in the alternation of generations. Due to their predominantly shaded distribution under canopies, light availability becomes a limiting factor for gametophyte survival, making it imperative to investigate their response to light. Previous research on fern gametophytes' light response has been limited to the physiological level. In this study, we examined the light response of *Adiantum flabellulatum* gametophytes under different photosynthetic photon flux density (PPFD) levels and identified their high sensitivity to low light. We thereby determined optimal and stress-inducing light conditions. By employing transcriptome sequencing, weighted gene co-expression network analysis, and Gene Ontology and Kyoto Encyclopedia of Genes and Genomes analyses, we identified 10,995 differentially expressed genes (DEGs). Notably, 3 *PHYBs* and 5 Type 1 *CRYs* (*CRY1s*) were significantly down-regulated at low PPFD (0.1 μmol $m^{-2}$ $s^{-1}$). Furthermore, we annotated 927 DEGs to pathways related to photosynthesis and 210 to the flavonoid biosynthesis pathway involved in photoprotection. Additionally, we predicted 34 transcription factor families and identified a close correlation between *mTERFs* and photosynthesis, as well as a strong co-expression relationship between *MYBs* and *bHLHs* and genes encoding flavonoid synthesis enzymes. This comprehensive analysis enhances our understanding of the light response of fern gametophytes and provides novel insights into the mechanisms governing their responses to light.

KEYWORDS

**fern gametophyte, light signal, photosynthesis, photoprotection, weighted gene co-expression network analysis**

# 1 Introduction

Ferns, an ancient lineage of land plants, have thrived on Earth for approximately 380 million years and are a fundamental component of the planet's plant community (Schneider et al., 2004). Their gametophytes exhibit the remarkable ability to survive independently and serve as the site for the production and occurrence of sexual reproductive cells. Additionally, they play a critical role in the alternation of generations (also known as metagenesis) by nourishing the young sporophyte (Krieg and Chambers, 2022).

Light, a pivotal environmental factor for plant growth, serves both as a signal perceived by photoreceptors and as energy captured by chloroplasts to facilitate photosynthesis (de Wit et al., 2016; Poorter et al., 2019). However, excessive light can induce photostress, challenging photoprotective mechanisms (Takahashi and Badger, 2011). Fern gametophytes predominantly inhabit shaded habitats, where light conditions pose a limiting factor for their survival. Previous studies have revealed that gametophytes of *Trichomanes speciosum* can effectively conduct photosynthesis under weak light conditions, with 85% of the daytime light intensity being less than 1 μmol $m^{-2}$ $s^{-1}$ (equivalent to approximately 0.01% to 0.1% of full daylight exposure). Nevertheless, prolonged exposure to higher light levels (>50 μmol $m^{-2}$ $s^{-1}$) can lead to the demise of these gametophytes (Johnson et al., 2000). The growth and/or photosynthetic rates of gametophytes from *Onoclea sensibilis* and *Cibotium glaucum* exhibit an increasing trend within a certain range of light intensity, but excessive light inhibits their development (Miller and Miller, 1961; Friend, 1974). In contrast, fern gametophytes are adapted to lower light conditions than sporophytes (Nitta et al., 2021). However, research on the light response of fern gametophytes has thus far been confined to the physiological level.

Flavonoids are secondary metabolites that are widely present in bryophytes, pteridophytes (including lycophytes, horsetails, and ferns), and spermatophytes (Santos et al., 2017). These compounds have photoprotective properties and can absorb ultraviolet and some visible light, as well as function as antioxidants to scavenge reactive oxygen species generated by strong light exposure (Ferreyra et al., 2021). Previous studies have shown that environmental factors such as light can regulate the expression of flavonoid biosynthetic enzyme genes in the sporophytes of ferns. However, it is unknown whether the expression of flavonoid biosynthetic enzyme genes in the gametophytes of ferns is regulated by light. For example, research has found that the expression of flavonoid biosynthetic enzyme genes in fronds of *Azolla filiculoides*, such as *CHI*, *C4H*, *CHS*, and *DFR1*, is regulated by environmental factors such as light, which affects flavonoid synthesis (Costarelli et al., 2021). Similarly, studies on cultured spores of *Dryopteris fragrans* have shown that UV-B stress can increase the expression of the *DfCHS* gene and the total flavonoid content (Sun et al., 2014). Therefore, further research on the expression of flavonoid biosynthetic enzyme genes and their regulation by light in the gametophytes of ferns is of great significance for a deeper understanding of the photoprotective mechanism of fern gametophytes.

Transcription factors/transcriptional regulators are proteins that regulate gene expression (Strader et al., 2022). In the genomes of the model plant *Arabidopsis thaliana* and sorghum (*Sorghum bicolor*), the proportion of transcription factors/transcriptional regulators is 7.51% and 7.69%, respectively (Arabidopsis Genome, 2000; Tian et al., 2016). Despite their large genome size, the proportion of transcription factors/transcriptional regulators in ferns is relatively low (Sessa and Der, 2016). For example, in the transcriptomes of ferns *Huperzia serrata* and *Monachosorum maximowiczii*, the proportion of transcription factors/transcriptional regulators is only 1.87% and 1.11%, respectively (Liu et al., 2016; Yang et al., 2017). In the transcriptome of the gametophyte of *Adiantum flabellulatum* that we previously reported, the proportion of transcription factors/transcriptional regulators is 2.25% (Cai et al., 2022a). The genome of *Salvinia cucullata*, the smallest known genome among ferns, has only 983 transcription factors/transcriptional regulators, accounting for 4.94% of all genes, which is also lower than that in gymnosperms or angiosperms (Li et al., 2018). Currently, research on the regulatory role of transcription factors/transcriptional regulators in ferns is very limited, and we still do not understand how they respond to light, and whether there is a potential regulatory relationship with genes related to photoprotection.

*A. flabellulatum*, which typically grows under the canopy, is an acid-soil indicator plant (Liao et al., 2017) that has significant value in both ornamental and medicinal uses (Shahriar and Kabir, 2011). In this study, we used transcriptome sequencing, weighted gene co-expression network analysis, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, and real-time quantitative polymerase chain reaction (qPCR) to explore the genes responsive to light and potential regulatory relationships between transcription factors and genes related to flavonoid biosynthesis. Furthermore, by identifying key genes and pathways, this study enhances our understanding of the light response mechanism of *A. flabellulatum* gametophytes and provides a foundation for uncovering their molecular mechanism.

# 2 Materials and methods

## 2.1 Cultivation of *Adiantum flabellulatum* gametophytes

Spores of *A. flabellulatum* were collected from the rubber forest at Ma'an Mountain (109.519°E, 19.501°N) in Danzhou, Hainan Province, China, as described by Cai et al. (2022b). The spores were sterilized with 0.1% $HgCl_2$ for 5 min, washed with sterile water five times, and then plated on 1/4 MS solid medium (3% w/v sucrose and 0.7% w/v agar). The plated spores were exposed to meticulously controlled illumination, with a photosynthetic photon flux density (PPFD) level of 7.1 μmol $m^{-2}$ $s^{-1}$, and maintained under an uninterrupted 24-hour photoperiod. Upon spore germination, young gametophytes with diameters ranging from 1 to 2 mm were transferred to fresh 1/4 MS solid medium, also containing 3% w/v sucrose and 0.7% w/v agar. Subsequently, the gametophytes were divided into seven distinct experimental groups, each cultivated under varying PPFD conditions: 0, 0.1, 1.4, 7.1, 14.4, 73.6, and 145.3 μmol $m^{-2}$ $s^{-1}$, respectively (Supplementary Table 1, Supplementary Figure 1). These cultivations were carried out over a continuous 40-day period under an uninterrupted 24-hour

photoperiod. The perimeter and projected area of the gametophytes were quantified using Image J software (v1.8.0). Each treatment was replicated at least eight times, and the cultivation temperature was maintained at 25°C. Statistical analysis was conducted using the least significant difference (LSD) test.

## 2.2 RNA extraction, library construction and sequencing

The gametophytes of *A. flabellulatum* were cultivated under diverse light conditions: 0, 0.1, 7.1, and 145.3 μmol $m^{-2}$ $s^{-1}$, for a period of 12 days, maintaining an uninterrupted 24-hour photoperiod. Subsequently, the four sample groups, denoted as Af0, Af0.1, Af7.1, and Af145.3, were collected for transcriptome sequencing. Three biological replicates were established for each treatment, resulting in a total of 12 samples. Approximately 200 mg of gametophytes were collected from each sample, and total RNA was extracted using the CTAB method. mRNA was then enriched using oligo dT magnetic beads, fragmented, and reverse-transcribed into cDNA for library construction. Sequencing was performed on the DNBSEQ platform. The raw reads were filtered using SOAPnuke software (v1.5.2) to remove reads with adapters, N content exceeding 10%, and more than 50% of bases with a quality score below 15. The clean reads were aligned to the full-length transcriptome of *A. flabellulatum* gametophytes using Bowtie2 (v2.2.5), and gene expression levels were calculated using RSEM (v1.2.8). These steps were executed by the Beijing Genomics Institute (BGI). Finally, we conducted sequencing saturation analysis, correlation analysis, and principal component analysis using the method described by Cai et al. (2021).

## 2.3 Selection of differentially expressed genes and acquisition of co-expression gene modules

The study involved three comparison groups: Af0 vs Af0.1, Af0.1 vs Af7.1, and Af7.1 vs Af145.3. Genes with average expression levels less than 0.5 in each comparison group were removed, and differentially expressed genes (DEGs) were screened using the criteria of $|\log_2 \text{FoldChange}| \geq 1$ and Q-Values < 0.001. The expression levels of all DEGs in the 12 samples were used to construct gene co-expression modules using the WGCNA package in R software, with "sft$powerEstimate=13" and "mergeCutHeight=0.25". Modules that were strongly correlated with the phenotypes and light conditions of *A. flabellulatum* gametophyte were selected for further analysis.

## 2.4 GO and KEGG analysis

The module genes were annotated and classified using Blast2GO (v2.5.0) for Gene Ontology (GO, http://www.geneontology.org/) and Blastx (v2.2.23) and Diamond (v0.8.31) tools for KEGG (http://www.genome.jp/kegg/) annotation and classification, respectively (Cai et al., 2022b). Enrichment analysis was performed using the Phyper function package in the R software. BGI completed the GO and KEGG annotations. Enrichment pathways related to the gametophyte phenotype and light conditions of *A. flabellulatum* were selected for further analysis and discussion.

## 2.5 Prediction of transcription factor families

In this study, the identification of open reading frames (ORFs) from gene sequences was conducted using getorf (EMBOSS: 6.5.7.0). The resulting ORFs were aligned to transcription factor protein domains (http://plantregmap.gao-lab.org/) using hmmsearch (v3.0) (Tian et al., 2020). The genes' potential for encoding transcription factors were then identified based on the transcription factor family characteristics outlined in PlantTFDB (http://planttfdb.gao-lab.org/) (Jin et al., 2017) and classified according to their respective transcription factor families.

## 2.6 Visualization of gene co-expression network

We used Cytoscape software (v3.9.1) to visualize the co-expression networks of flavonoid biosynthesis genes, as well as *MYBs* and *bHLHs* in the green, brown, yellow, and turquoise modules, following the method described by Shannon et al. (2003). In the visualization, nodes represent genes, and the size of each node is positively correlated with its degree. Lines represent the weight of co-expression between genes, with thicker and darker lines indicating higher weights.

## 2.7 Quantitative real-time polymerase chain reaction

Fifteen genes were chosen for validation by quantitative real-time polymerase chain reaction (qRT-PCR) (Supplementary Table 2). A total of 12 samples were used, with three technical replicates per sample. To achieve relative quantification, we selected isoform_78364 (*Elongation Factor Ts*), which exhibited stable expression across all samples, as the internal control gene, following the method described by Pfaffl (2001). Relative quantification was carried out using the $2^{-\Delta\Delta Ct}$ method, and significance analysis was performed using the LSD method.

# 3 Results

## 3.1 Effects of different PPFD levels on gametophyte growth of *Adiantum flabellulatum*

The perimeters and projected areas of gametophytes grown under different PPFD levels were measured. The results showed that gametophytes are highly sensitive to low light, with their perimeter

increasing significantly under a PPFD as low as 0.1 μmol $m^{-2}$ $s^{-1}$. Below 7.1 μmol $m^{-2}$ $s^{-1}$, the perimeter and/or projected area of gametophytes increased with increasing PPFD levels and reached a maximum at 7.1 μmol $m^{-2}$ $s^{-1}$. When the PPFD levels ranged between 7.1 μmol $m^{-2}$ $s^{-1}$ and 73.6 μmol $m^{-2}$ $s^{-1}$, the average value of the perimeter and projected area of gametophytes decreased slightly, but no significant differences were observed. However, when the PPFD reached 145.3 μmol $m^{-2}$ $s^{-1}$, the perimeter and projected area of gametophytes decreased significantly (Figure 1). Therefore, we conclude that 7.1 μmol $m^{-2}$ $s^{-1}$ is the optimal PPFD for the growth of *A. flabellulatum* gametophytes, while both PPFD levels below 7.1 μmol $m^{-2}$ $s^{-1}$ and at 145.3 μmol $m^{-2}$ $s^{-1}$ can induce light stress on the gametophytes.

## 3.2 Quality evaluation of transcriptome sequencing

We selected four treatments with PPFD levels of 0, 0.1, 7.1, and 145.3 μmol $m^{-2}$ $s^{-1}$ (represented by Af0, Af0.1, Af7.1, and Af145.3, respectively) to explore the genes involved in the response of *A. flabellulatum* gametophytes to light using transcriptome sequencing, based on previous experimental results. Three biological replicates were performed for each treatment, resulting in a total of 12 samples that were sequenced using the DNBSEQ platform. The full-length transcript sequences of *A. flabellulatum* gametophytes reported by Cai et al. (2022a) were used as reference genes for relative quantification.

The raw read counts of each sample were greater than 42 M, and after filtering, each sample had at least 41 M clean reads, with a ratio greater than 96.17% and Q20 and Q30 values higher than 94% and 87%, respectively (Supplementary Table 3). Sequencing saturation analysis indicated that the gene identification ratio plateaued when the reads number exceeded 50×100 K, suggesting that the sequencing data had reached saturation (Supplementary Figure 2A). The gene expression level distributions of the 12 samples were analyzed, and the median gene expression levels were all ≥0.97 (Supplementary Figure 2B). Pearson correlation coefficients between biological replicates were ≥0.91 (Figure 2A),

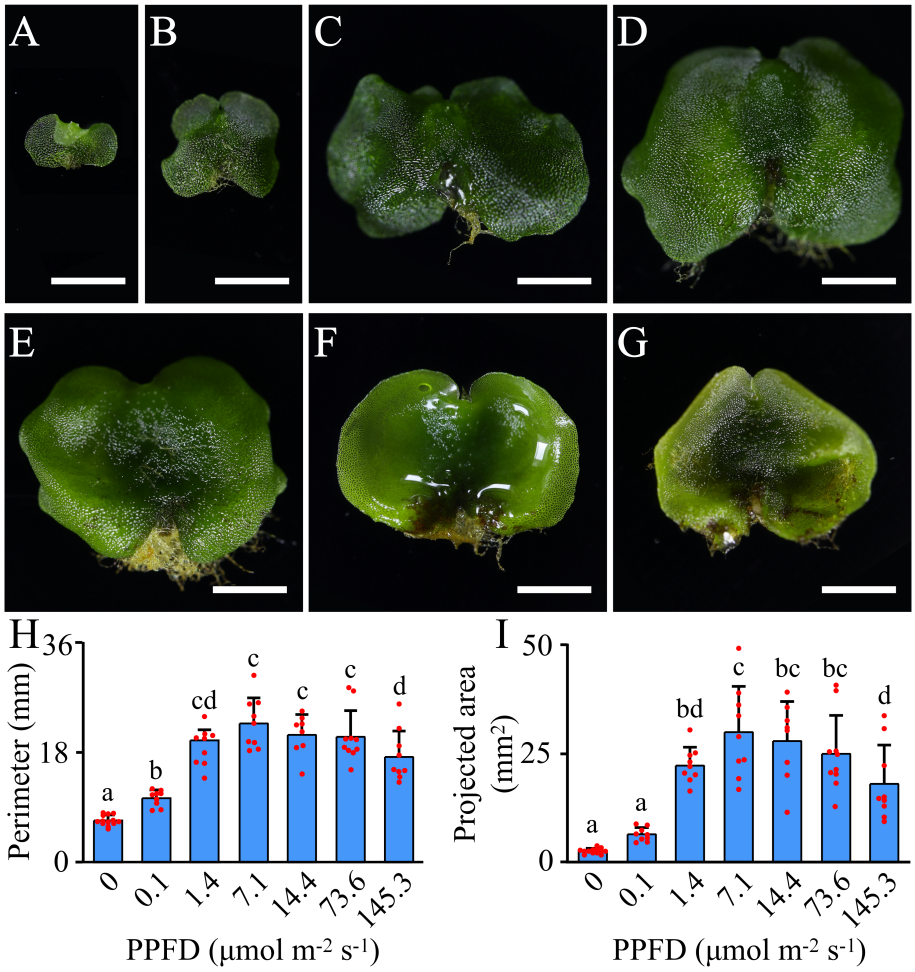



FIGURE 1
Effects of different PPFD levels on the growth of *A. flabellulatum* gametophytes. **(A–G)**: *A. flabellulatum* gametophytes were cultured for 40 days under PPFD levels of 0, 0.1, 1.4, 7.1, 14.4, 73.6, and 145.3 μmol $m^{-2}$ $s^{-1}$. **(H, I)**: The gametophyte perimeter **(H)** and projected area **(I)** were measured after 40 days of cultivation under different PPFD levels. Lowercase letters indicate significant differences ($p < 0.05$) between different PPFD levels. Scale bar = 2 mm.

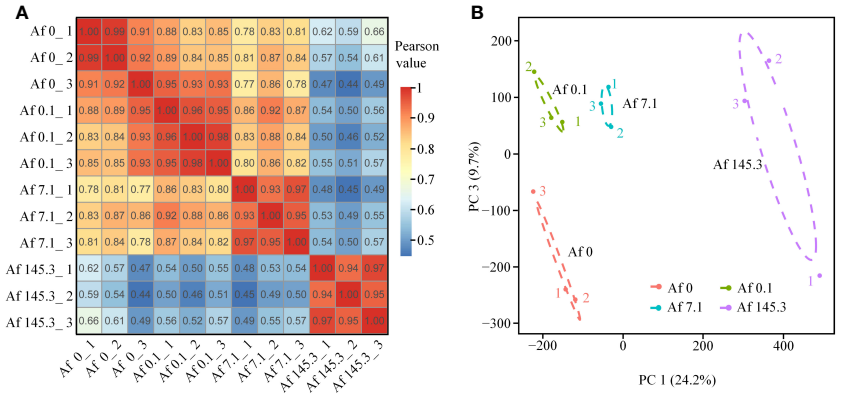



FIGURE 2
Pearson correlation coefficient analysis and principal component analysis of 12 samples. **(A, B)**: Pearson correlation coefficient analysis **(A)** and principal component analysis (PCA) **(B)** based on the expression data of all genes in 12 samples from the four treatments. The PCA results show that the four treatments are separated on a plane constructed by PC1 and PC3, indicating that these two principal components account for most of the variation among the treatments.

and the four treatments were separated from each other on the plane composed of PC1 and PC3 in the principal component analysis (Figure 2B). These results confirm the reliability and high quality of the sequencing data, which can be used for subsequent DEGs screening and analysis. The sequencing data have been deposited in the NCBI SRA database under accession number PRJNA868202.

## 3.3 Selection of differentially expressed genes

A total of 248,028 genes were detected among the 12 samples. Three comparison groups were set: Af0 vs Af0.1, Af0.1 vs Af7.1, and Af7.1 vs Af145.3. In each comparison group, we filtered out genes with an average expression level of less than 0.5 for both treatments and then selected DEGs with $|\log_2 \text{FoldChange}| \geq 1$ and *Q*-Values < 0.001. This resulted in 1,059, 5,536, and 10,152 DEGs, respectively. The Af7.1 vs Af145.3 comparison had the most DEGs, with 7,692 upregulated and 2,460 downregulated genes, followed by Af0.1 vs Af7.1, which had 1,864 upregulated and 3,672 downregulated genes. The Af0 vs Af0.1 comparison had the fewest DEGs (Supplementary Tables 4–6; Figures 3A–C). After removing duplicates, a total of 15,874 DEGs were obtained (Figure 3D).

For these 15,874 DEGs, we performed weighted gene co-expression analysis and obtained 12 gene co-expression modules (Figure 3E). The turquoise module contained 8,796 DEGs and was positively correlated with "PPFD levels," while the yellow (746 DEGs) and brown (875 DEGs) modules were positively correlated with "gametophyte area." The green module (387 DEGs) was positively correlated with "presence or absence of light" (Supplementary Table 7). In contrast, the black module (191 DEGs) was negatively correlated with "presence or absence of light" (Supplementary Table 7, Supplementary Figure 3). Based on these results, we selected DEGs from these five modules for further analysis.

## 3.4 Gene ontology analysis of module genes

We performed GO annotation on a total of 10,995 DEGs from the five gene modules mentioned earlier. Of these, 8,033 were successfully annotated, accounting for 73.06% of the total. These annotations covered three major functional categories: "Biological process," "Cellular component," and "Molecular function," which contained 4,911, 5,451, and 6,424 genes, respectively. In the "Biological process" category, all five modules had the highest number of genes annotated in the "cellular process" term, while in the "Cellular component" category, it was the "cellular anatomical entity." In the "Molecular function" category, the black and brown modules had the highest number of genes annotated in the "binding" term, while the green, yellow, and turquoise modules were most enriched in the "catalytic activity" term (Supplementary Tables 8, 9; Figure 4A).

We present the top 21 enriched GO terms with *Q*-values < 0.01. Firstly, the green module was enriched in at least 8 terms related to light signaling (8/21). Among them, "photoreceptor activity" had the smallest *Q*-value (*Q*=1.67E-07) and the most genes (8) (Supplementary Table 10, Supplementary Figure 4A). This indicates that the DEGs in the green module play an important role in regulating light signaling during the development of *A. flabellulatum*'s gametophyte. Secondly, most of the terms in the black and brown modules were related to photosynthesis. For example, "thylakoid" had the smallest *Q*-value (*Q*=1.63E-28) and the most genes (52) in the black module, while "regulation of photosynthesis, dark reaction" had the highest enrichment rate (0.11) (Supplementary Table 10, Supplementary Figure 4B). In the brown module, "photosystem II" had the smallest *Q*-value (*Q*=2.35E-115) and the most genes (166), while "photosynthetic acclimation" had the highest enrichment rate (0.125) (Supplementary Table 10, Supplementary Figure 4C). This indicates that the DEGs in the black and brown modules mainly participate in regulating photosynthesis during the development of

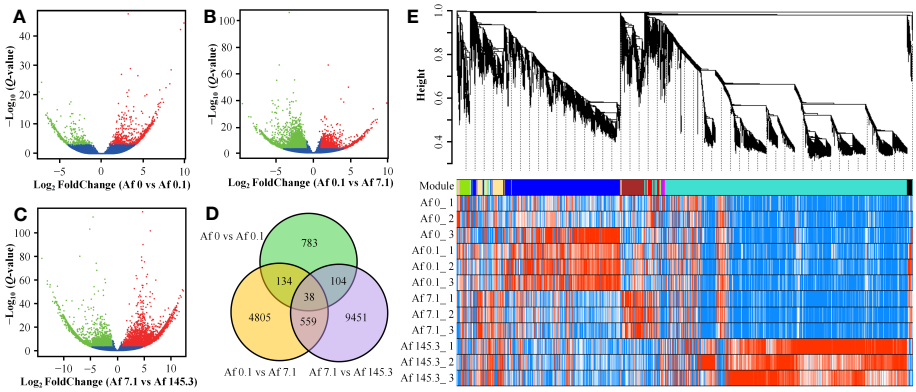



FIGURE 3
Identification of DEGs and weighted gene co-expression network analysis. **(A–C)**: DEGs were identified using criteria of $|\log_2$ FoldChange$| \geq 1$ and *Q*-Values < 0.001. The red and green dots represent up-regulated and down-regulated DEGs in Af0 vs Af0.1 **(A)**, Af0.1 vs Af7.1 **(B)**, and Af7.1 vs Af145.3 **(C)**, respectively. **(D)**: Venn diagram of DEGs distribution among three comparison groups, with the numbers in overlapping regions representing the shared DEGs. **(E)**: Dendrogram and gene modules of a total of 15,874 DEGs obtained through hierarchical clustering analysis. The turquoise module contains the largest number of DEGs.

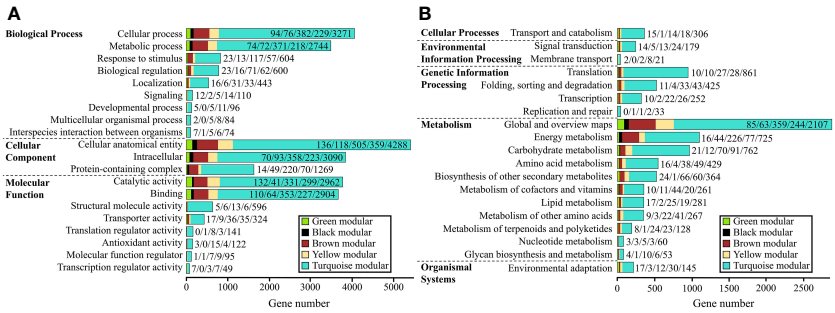



FIGURE 4
GO and KEGG classifications of the five gene modules. **(A, B)**: GO **(A)** and KEGG **(B)** classifications of genes in the five modules. The numbers on each bar represent the count of genes for the green, black, brown, yellow, and turquoise modules, respectively.

*A. flabellulatum*'s gametophytes. Additionally, the two terms with the smallest *Q*-value in the yellow module were "flavonoid biosynthetic process" (*Q*=3.61E-09) and "flavonoid metabolic process" (*Q*=3.87E-09), and they both had the most genes (12) (Supplementary Table 10, Supplementary Figure 4D). The terms enriched in the turquoise module were mainly related to cell damage, such as "cell killing" and "killing of cells of other organism," and they had the most genes (36) (Supplementary Table 10, Supplementary Figure 4E). This indicates that the DEGs in the yellow and turquoise modules mainly participate in resisting light stress caused by strong light.

## 3.5 KEGG analysis of module genes

We performed KEGG annotation on 10,995 DEGs, of which 5,405 genes were successfully annotated with an annotation rate of 49.16%. These genes were classified into five major categories, including "Cellular processes", "Environmental information processing", "Genetic Information Processing", "Metabolism", and "Organismal Systems". Among these categories, the "Metabolism" category had the most DEGs, with a total of 3,363 genes. The five gene modules were annotated to the most genes in the "Transport and catabolism" (354 genes) of "Cellular processes", "Signal transduction" (235 genes) of "Environmental information processing", "Global and overview maps" (2,858 genes) of "Metabolism", and "Environmental adaptation" (207 genes) of "Organismal Systems". In the "Genetic Information Processing" category, the black and turquoise modules had the most genes annotated in "Translation", while the green, brown, and yellow modules were involved in "Folding, sorting and degradation". Importantly, in the "Metabolism" category, all five gene modules had a relatively high number of genes annotated in "Energy metabolism" (1,088 genes) and "Carbohydrate metabolism" (956 genes), indicating that the regulation of these two metabolisms by light is also crucial (Supplementary Tables 11, 12; Figure 4B).

We presented the top 10 KEGG pathways with the smallest *Q*-values in enrichment analysis (Supplementary Figure 5; Figure 5). Firstly, the green, black, and yellow gene modules were all enriched in the "Circadian rhythm-plant" pathway, with the green module

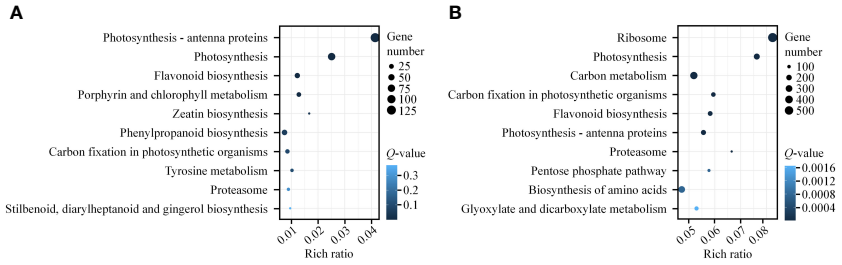



FIGURE 5
KEGG pathway enrichment analysis of brown and turquoise modules. **(A, B)**: KEGG pathway enrichment analysis of genes in the brown **(A)** and turquoise **(B)** modules. The top 10 enriched pathways with the smallest *Q*-values are presented. The size of the circle corresponds to the number of genes in the pathway, and the color intensity reflects the level of significance (darker colors indicate smaller *Q*-values).

having the smallest *Q*-value and the largest number of genes (14) in this pathway (Supplementary Table 13, Supplementary Figure 5A). Secondly, the black, yellow, brown, and turquoise gene modules were all enriched in multiple pathways related to photosynthesis, such as "Photosynthesis-antenna proteins" and "Photosynthesis" (Supplementary Table 13). The black module had the smallest *Q*-value (1.63E-20), highest enrichment rate (0.009240924), and the largest number of genes (28) in the "Photosynthesis" pathway (Supplementary Table 13, Supplementary Figure 5B). In the brown module, the "Photosynthesis-antenna proteins" pathway had the smallest *Q*-value (1.08E-77), highest enrichment rate (0.041363336), and the largest number of genes (125) (Supplementary Table 13; Figure 5A). Additionally, the brown, yellow, and turquoise gene modules were enriched in the "Flavonoid biosynthesis" pathway, which is involved in photoprotection (Supplementary Table 13, Supplementary Figure 5C; Figure 5). In summary, our KEGG analysis results also indicate that light plays an essential regulatory role in light signal transduction, photosynthesis, and photoprotection during gametophyte development in *A. flabellulatum*.

## 3.6 Analysis of transcription factor families for module genes

We predicted transcription factor families for the 10,995 DEGs and identified 34 families, with mTERF, MYB, and bHLH families having the highest number of DEGs at 13 each. The turquoise module had the most predicted transcription factor families (30) and individual transcription factors (95), followed by the yellow module with 13 families and 17 individual factors (Supplementary Table 14, Supplementary Figure 6).

## 3.7 Expression analysis of photosynthesis-related genes

As the "Photosynthesis-antenna proteins" and "Photosynthesis" pathways are significantly enriched in multiple gene modules, we first analyzed the expression of DEGs in these two pathways. The "Photosynthesis-antenna proteins" pathway includes Light-harvesting complex I (LHC I) and Light-harvesting complex II (LHC II), with LHC I containing five subunits (Lhca1-Lhca5) and LHC II containing seven subunits (Lhcb1-Lhcb7). Results showed that among the five gene modules, the "Photosynthesis-antenna proteins" pathway had a total of 327 DEGs, including the 12 subunits mentioned above, mainly belonging to the brown (125) and turquoise (168) modules. These genes were generally expressed at the highest levels under a PPFD of 7.1 μmol $m^{-2}$ $s^{-1}$ (brown module) or 145.3 μmol $m^{-2}$ $s^{-1}$ (turquoise module) (Supplementary Table 15; Figure 6). In addition, the "Photosynthesis" pathway includes the Photosystem II complex (PSII complex), cytochrome $b_6/f$ complex, Photosystem I complex (PSI complex), photosynthetic electron transfer chain, and Adenosine triphosphate synthase (ATP synthase). Results showed that among the five gene modules, the "Photosynthesis" pathway had a total of 370 DEGs, including 30 subunits, such as 11 subunits of the PSII complex (PsbD, PsbO, PsbP, PsbQ, PsbR, PsbS, PsbW, PsbY, PsbZ, Psb27, Psb28), one subunit of the cytochrome $b_6/f$ complex (PetC), 10 subunits of the PSI complex (PsaB, PsaD, PsaE, PsaF, PsaG, PsaH, PsaK, PsaL, PsaN, PsaO), 4 subunits of the photosynthetic electron transfer chain (PetE, PetF, PetH, PetJ), and 4 subunits of ATP synthase (gamma, delta, a, b). These DEGs were also mainly distributed in the brown (76) and turquoise (231) modules (Supplementary Table 16, Figure 7).

We also analyzed the DEGs involved in Calvin cycle (141), chlorophyll (61), and carotenoid (28) biosynthesis, and found that these DEGs were mainly classified into the turquoise module (Supplementary Tables 17-19, Supplementary Figures 7-9).

Moreover, mitochondrial transcription termination factors (mTERFs) located in plastids play important regulatory roles in the development and function of chloroplasts. *AfmTERFs* were detected in the brown, yellow, and turquoise modules, with 1, 1, and 11 members, respectively. Therefore, we constructed a phylogenetic tree of mTERFs from *A. flabellulatum* and *Arabidopsis*. The results showed that four AfmTERFs (isoform_43473, isoform_47623, isoform_50012, and isoform_240354) in *A. flabellulatum* were homologous to AtmTERF6, one (isoform_20996) was homologous to AtmTERF9, and two (isoform_83065 and isoform_91611) were homologous to AtmTERF12 (Supplementary Table 20, Supplementary Figure 10). Interestingly, all seven *AfmTERFs* belonged to the turquoise module.

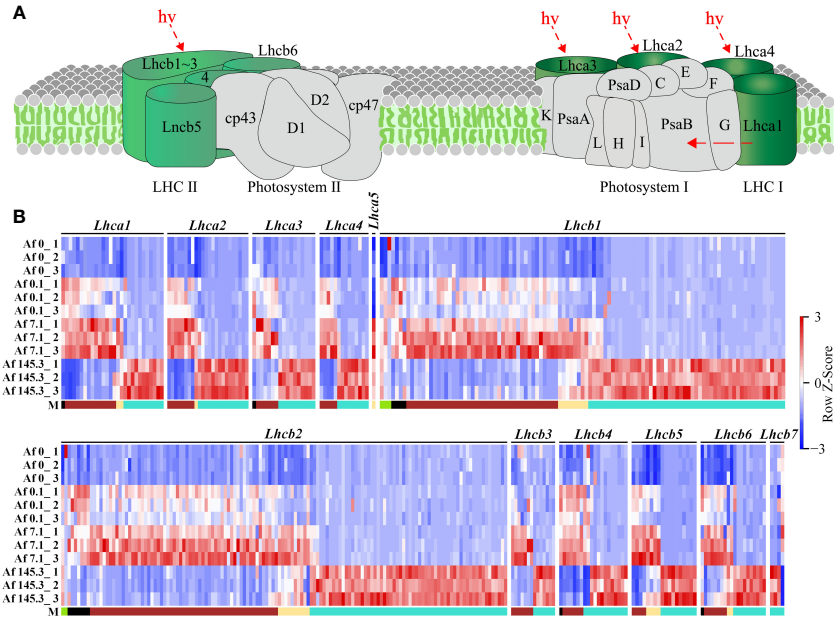



FIGURE 6
Expression of DEGs in the "Photosynthesis-antenna proteins" pathway across 12 samples. **(A)**: Diagram of the "Photosynthesis-antenna proteins" pathway (source: https://www.kegg.jp/pathway/map00196, accessed on April 14, 2023); **(B)**: Expression profiles of DEGs encoding 12 subunits of LHCI and LHCII across 12 samples. "M" represents the gene module. It can be observed that there are more DEGs in the brown and turquoise modules for this pathway.

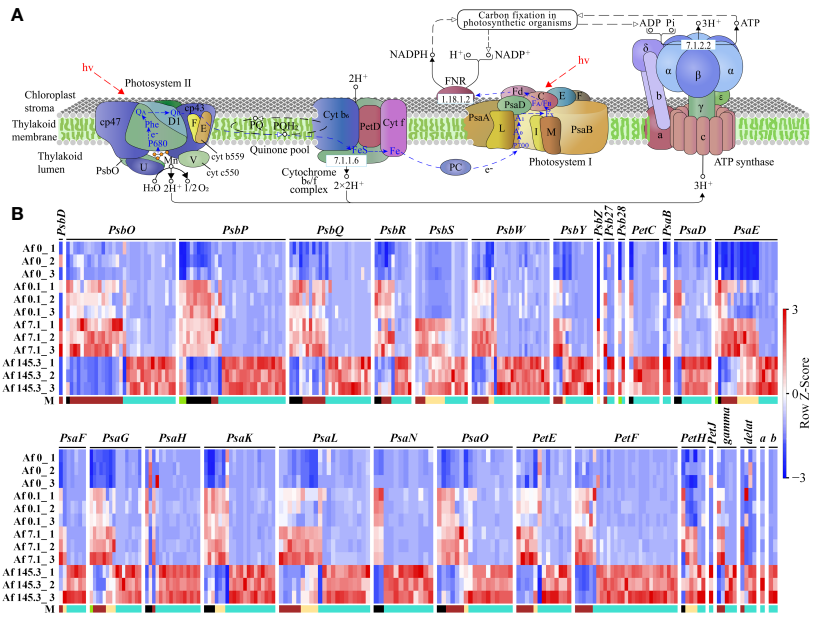



FIGURE 7
Expression of DEGs in the "Photosynthesis" pathway across 12 samples. **(A)**: Diagram of the "Photosynthesis" pathway (source: https://www.kegg.jp/entry/map00195, accessed on April 14, 2023); **(B)**: Expression profiles of DEGs encoding 30 subunits of the PSII complex (11 subunits: PsbD, PsbO, PsbP, PsbQ, PsbR, PsbS, PsbW, PsbY, PsbZ, Psb27, Psb28), cytochrome $b_6/f$ complex (1 subunit: PetC), PSI complex (10 subunits: PsaB, PsaD, PsaE, PsaF, PsaG, PsaH, PsaK, PsaL, PsaN, PsaO), photosynthetic electron transfer chain (4 subunits: PetE, PetF, PetH, PetJ), and ATP synthase (4 subunits: gamma, delta, a, b) in 12 samples. "M" represents gene modules. The brown and turquoise modules contain a larger number of genes.

## 3.8 Expression of flavonoid biosynthesis genes

Based on the analysis in the previous sections, we found that the flavonoid biosynthesis pathway is significantly enriched in multiple gene modules, particularly in the brown, yellow, and turquoise modules. Therefore, we analyzed the expression of DEGs in this pathway. Our results show that there are 210 DEGs encoding 13 enzymes, including CHS, CHI, F3H, and FLS, which are key enzymes in the flavonoid biosynthesis pathway. These DEGs mainly belong to the turquoise module (Supplementary Table 21; Figure 8). Thus, we suggest that light, especially strong light, strongly upregulates the expression of genes encoding flavonoid biosynthesis enzymes in the gametophyte of *A. flabellulatum*.

The MYB and bHLH families of transcription factors are important in plant responses to abiotic stress. These two types of transcription factors were detected in four modules, with the highest number found in the turquoise module. Therefore, we performed co-expression analysis of flavonoid biosynthesis genes and MYB/bHLH transcription factor genes in *A. flabellulatum*. Our results showed that the encoding genes of flavonoid biosynthesis enzymes, CHS, DFR, HCT, and CHI, have a wide range of co-expression relationships with *MYBs* and *bHLHs*, with 39, 20, 20, and 19 genes in the turquoise module, respectively (Supplementary Table 22; Figure 9). Furthermore, we also found co-expression relationships between *MYBs* and *bHLHs* in the turquoise module.

## 3.9 Quantitative real-time polymerase chain reaction

To validate the accuracy of the transcriptomic data, we selected 15 DEGs for qRT-PCR analysis. These included two genes from the "Photosynthesis-Antenna Proteins" pathway (*Lhca3* and *Lhcb1*/*Lhcb2*), three genes from the "Photosynthesis" pathway (*PsaG*, *PsaL*, and *PetE*), nine genes from the "Flavonoid Biosynthesis" pathway (*CHSs*, *CHIs*, *F3H*, *ANS*, *DFR*, and *HCTs*), and *Cryptochrome* (*CRY*). Our qRT-PCR results demonstrated that the expression patterns of these 15 genes were consistent with the transcriptomic data, confirming the reliability of our transcriptomic analysis (Supplementary Table 23; Figure 10).

# 4 Discussion

The gametophytes of ferns are small in size and need to constantly sense and utilize the scattered sunlight that passes

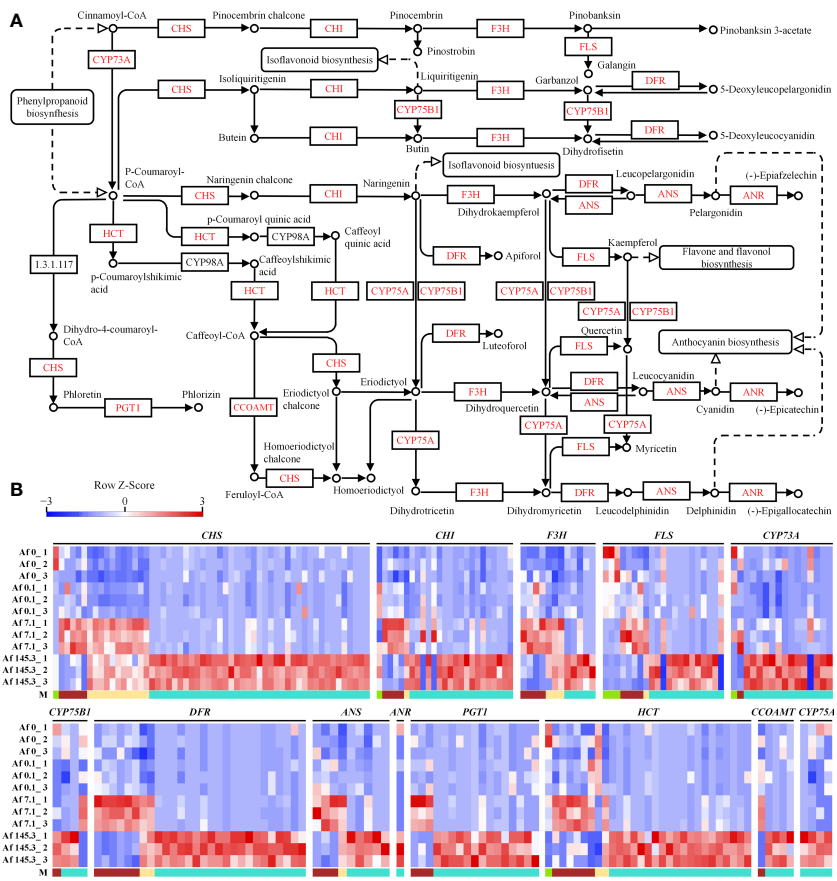



FIGURE 8
Expression of DEGs in the "Flavonoid Biosynthesis" pathway across 12 samples. **(A)**: "Flavonoid Biosynthesis" pathway (Source: https://www.kegg.jp/pathway/map00941, accessed on April 14, 2023); **(B)**: Expression profiles of DEGs encoding CHS, CHI, F3H, FLS, CYP73A, CYP75B1, DFR, ANS, ANR, PGT1, HCT, CCOAMT, and CYP75A in 12 samples. "M" represents gene modules, with a higher number of DEGs observed in the turquoise module.

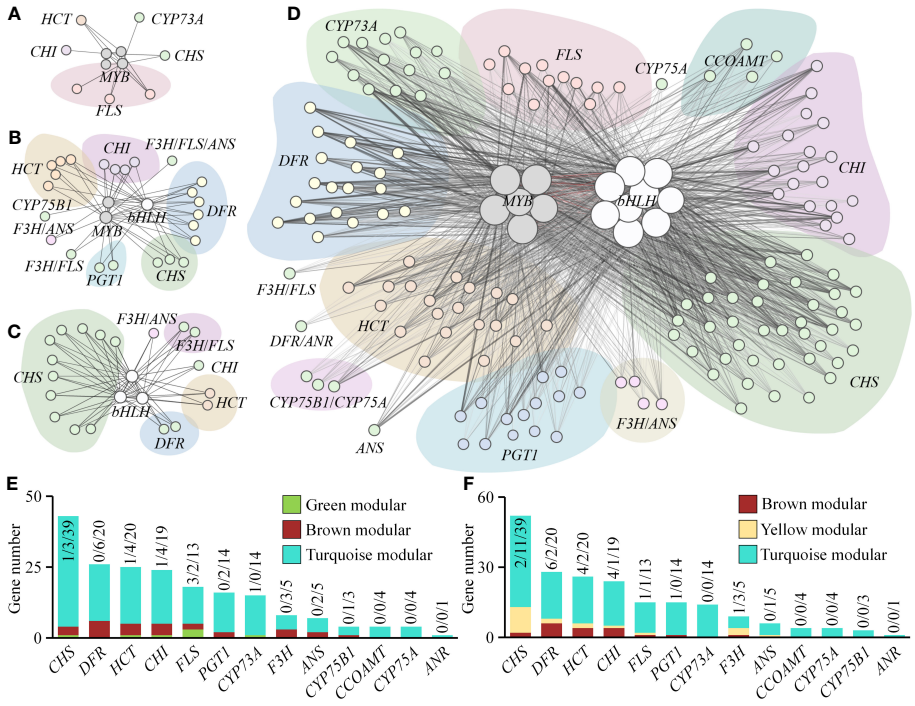



FIGURE 9
Co-expression network of flavonoid biosynthesis genes and MYB/bHLH transcription factor genes. **(A–D)**: Co-expression networks of flavonoid biosynthesis genes and MYB/bHLH transcription factor genes in green, brown, yellow, and turquoise modules, respectively. Nodes represent genes, with larger nodes indicating higher degree values. Lines represent co-expression weights between genes, with thicker and darker lines indicating higher weight values. **(E, F)**: Number of co-expressed flavonoid biosynthetic enzyme-coding genes with MYB and bHLH transcription factor genes, respectively. Additionally, we found a co-expression relationship between *AfMYBs* and *AfbHLHs* in the turquoise module, indicated by red lines.

through the forest canopy to survive. In our study, GO enrichment analysis of the green module identified 8 out of 21 enriched terms related to light signaling. Notably, "photoreceptor activity" had the smallest *Q*-value and the highest number of genes (Supplementary Figure 4A). Additionally, KEGG enrichment analysis showed significant enrichment of the "Circadian rhythm-plant" pathway in the green module, with the smallest *Q*-value and a relatively large number of genes (Supplementary Figure 5A). Phytochromes (PHYs) and cryptochromes (CRYs) are the primary photoreceptors that mediate light-induced signaling and regulate circadian rhythms in plants (Guo et al., 1998; Oakenfull and Davis, 2017). In the green module, we identified three *PHYB* genes and five *CRY1* genes that were significantly downregulated at a PPFD of only 0.1 μmol $m^{-2}$ $s^{-1}$ (Supplementary Table 24, Supplementary Figure 11). We also observed a significant increase in gametophyte perimeter under a low PPFD of 0.1 μmol $m^{-2}$ $s^{-1}$. As *A. flabellulatum* predominantly thrives in humid understory environments (Yang, 2015), we conclude that its gametophytes exhibit high sensitivity to low light, a trait likely linked to their adaptation to such conditions.

Mitochondrial transcription termination factors (mTERFs) are a class of transcriptional regulators encoded by nuclear genes. Originally discovered in animal mitochondria, mTERFs have been found to play a role in transcriptional regulation not only in animal mitochondria but also in plant chloroplasts and mitochondria (Roberti et al., 2009; Babiychuk et al., 2011). In *Arabidopsis*, AtmTERF6 is localized to both chloroplasts and mitochondria, while AtmTERF9 and AtmTERF12 are localized to chloroplasts. Among them, AtmTERF6 and AtmTERF9 have been shown to play important roles in photosynthesis and chloroplast development (Robles et al., 2015; Romani et al., 2015; Xu et al., 2017; Zhang et al., 2018), while the function of AtmTERF12 remains relatively limited (Xu et al., 2017). Single mutants of *mterf6* and *mterf9* in *Arabidopsis* exhibited reduced levels of 16S and 23S rRNA of the small and large plastid ribosomal subunits in chloroplasts, decreased number of chloroplast ribosomes, and affected protein translation in chloroplasts, leading to impaired chloroplast development, reduced chlorophyll a and b content, decreased number of mesophyll cells, and leaf whitening (Robles et al., 2015; Romani et al., 2015; Robles et al., 2018; Zhang et al., 2018; Meteignier et al., 2021). In comparison, the double mutant *mterf6-5 mterf9* in *Arabidopsis* seedlings exhibited a more pronounced pale/albino phenotype, indicating a synergistic effect between the genes (Robles et al., 2018). In this study, we discovered that there are five AfmTERFs (isoform_47623, isoform_43473, isoform_20996, isoform_83065, and isoform_91611) in the gametophytes of *A. flabellulatum*, which show homology to AtmTERF6, AtmTERF9, or AtmTERF12, and their encoding genes were up-regulated under strong light (145.3 μmol $m^{-2}$ $s^{-1}$). Given the localization and functional significance of mTERFs in other plant systems, we speculate that these AfmTERFs may play a crucial role in balancing the effects of light damage in fern gametophytes under strong light stimulation, ensuring the safe and efficient photosynthetic process. Further studies should aim to elucidate

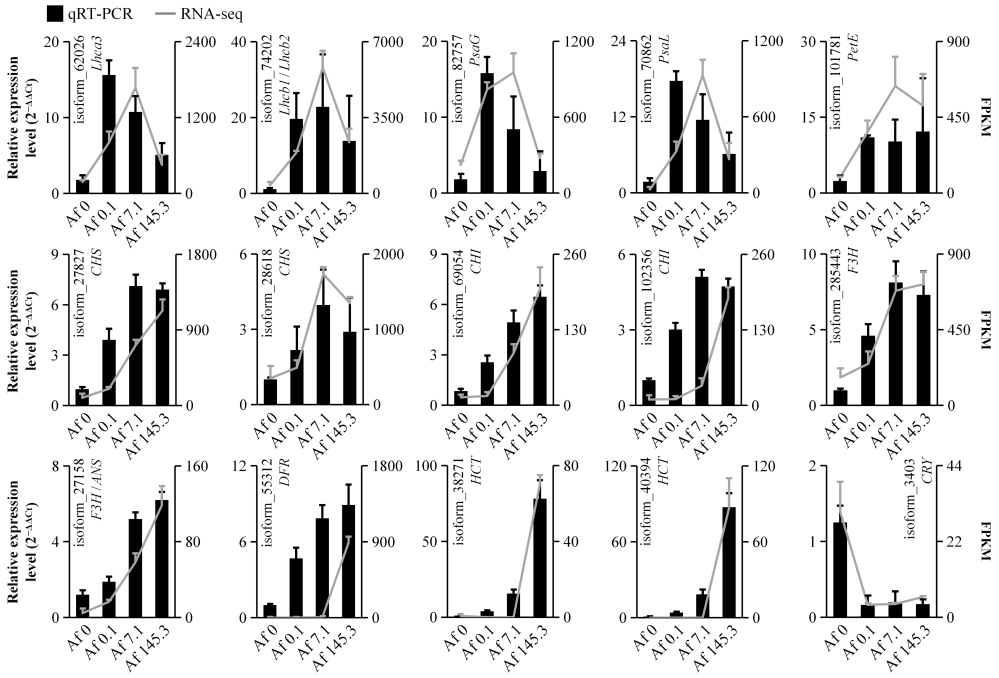



FIGURE 10
Real-time quantitative polymerase chain reaction. The expression trends of 15 DEGs were validated by qRT-PCR, with the bar chart and line chart representing the results of qRT-PCR and transcriptome, respectively. The consistency between qRT-PCR and transcriptome results suggests the reliability of the transcriptome data.

the precise molecular mechanisms underlying AfmTERF function and their impact on fern gametophyte development and survival.

Flavonoids play a vital role in photoprotection in plants by scavenging free radicals and suppressing the generation of reactive oxygen species (Agati et al., 2012). Strong light exposure promotes the expression of genes encoding flavonoid biosynthesis enzymes, such as CHS, CHI, and F3H, in various angiosperms, including *Arabidopsis thaliana*, *Petunia*, and *Medicago sativa* seedlings, enhancing flavonoid biosynthesis and providing photodamage defense. In this study, we found that light exposure, particularly strong light, significantly increased the expression of genes encoding 13 flavonoid biosynthesis enzymes in *A. flabellulatum* gametophytes, including key enzymes such as CHS, CHI, and F3H. This is the first report of such induction in fern gametophytes, although it has been reported in various angiosperms (Albert et al., 2009; Gao et al., 2019; Zheng et al., 2021). Meanwhile, we detected stable total flavonoid content in *A. flabellulatum* gametophytes under different PPFD levels (Supplementary Figure 12). Therefore, we hypothesize that the upregulation of relevant enzyme gene expression maintains the stable total flavonoid content during the process of responding to photodamage in *A. flabellulatum* gametophytes.

Transcription factors, particularly MYB and bHLH, regulate the expression of flavonoid biosynthesis genes. Previous studies have demonstrated that many MYB transcription factors can regulate the expression of flavonoid biosynthesis genes such as *CHS*, *CHI*, *F3H*, and *DFR*, promoting flavonoid biosynthesis (Dubos et al., 2010). For instance, AtMYB11, AtMYB12, and AtMYB111 in *Arabidopsis* can directly bind to the promoter of flavonoid biosynthesis genes like *AtCHS*, *AtCHI*, and *AtF3H*, directly regulating the expression of flavonoid biosynthesis genes (Stracke et al., 2007). Additionally, MYB can form binary or ternary complexes with bHLH and WD40 to co-regulate the expression of flavonoid biosynthesis genes (Xu et al., 2015), as seen in the MYB-bHLH-WD40 ternary complex in *Arabidopsis* and the MYB10-bHLH3 binary protein complex in red-fleshed apples (*Malus sieversii* f. *niedzwetzkyana*), which directly regulate the expression of *DFR*, promoting anthocyanin biosynthesis (Gonzalez et al., 2016; Wang et al., 2017). Light closely regulates *MYB* expression, as seen with *MdMYB1* in apples (*Malus domestica*) and *PpMYB10* in red Chinese sand pears (*Pyrus pyrifolia*), which are induced by light and regulate the expression of *DFR*, *CHS*, and *ANS* genes, promoting flavonoid biosynthesis (Takos et al., 2006; Bai et al., 2017). In this study, we found that several MYB and bHLH transcription factors in *A. flabellulatum* have a strong co-expression relationship with flavonoid biosynthesis genes encoding CHS, CHI, DFR, and F3H. Therefore, we hypothesize that under strong light irradiation, MYB and bHLH transcription factors in *A. flabellulatum* gametophytes are activated to regulate the expression of flavonoid biosynthesis genes, promoting flavonoid biosynthesis and playing a photoprotective role.

In summary, this study not only provides valuable insights into the protection and cultivation of *A. flabellulatum* but also serves as a reference for further research on gametophytes of ferns' responses to light. The comprehensive analysis of the light response in *A. flabellulatum* gametophytes sheds light on their sensitivity to different light conditions, which is likely an adaptation to their natural understory habitats. Furthermore, the identification of

*AfmTERFs* and their potential role in photoprotection enhances our understanding of fern gametophyte survival strategies under varying light intensities. The upregulation of flavonoid biosynthesis genes under strong light exposure highlights the significance of these secondary metabolites in photoprotection, offering potential avenues for future research on fern photobiology and stress responses.

## 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/genbank/, PRJNA868202.

## Author contributions

ZC and JL conceived and designed the study. ZC, XW and ZX performed experiments. ZC, XW, ZX and ZW analyzed the data. ZC, XW, ZX, ZW, XS and JL drafted the manuscript. ZC, XW, ZX, XY, SX and JL finalized the manuscript. JL, ZC and XY provided the funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Major Science and Technology plan of Hainan Province (ZDKJ2021018), the Natural Science Foundation of Hainan Province (No.322RC772; No.321QN316; No.319MS017; No.320RC508), the Program of Hainan Association for Science and Technology Plans to Youth Science and Techonology Talent Innovation (QCQTXM202201), the National Natural Science Foundation of China (No.31660229), the Hainan University Scientific Research Startup Fund Project (No.kyqd1620).

## Acknowledgments

We sincerely thank reviewers and editors of the *Frontiers in Plant Science* for providing valuable suggestions on this paper.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1222414/full#supplementary-material

## References

Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as antioxidants in plants: Location and functional significance. *Plant Sci.* 196, 67–76. doi: 10.1016/j.plantsci.2012.07.014

Albert, N. W., Lewis, D. H., Zhang, H., Irving, L. J., Jameson, P. E., and Davies, K. M. (2009). Light-induced vegetative anthocyanin pigmentation in *Petunia*. *J. Exp. Bot.* 60 (7), 2191–2202. doi: 10.1093/jxb/erp097

Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408 (6814), 796–815. doi: 10.1038/35048692

Babiychuk, E., Vandepoele, K., Wissing, J., Garcia-Diaz, M., De Rycke, R., Akbari, H., et al. (2011). Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. *Proc. Natl. Acad. Sci. U.S.A.* 108 (16), 6674–6679. doi: 10.1073/pnas.1103442108

Bai, S. L., Sun, Y. W., Qian, M. J., Yang, F. X., Ni, J. B., Tao, R. Y., et al. (2017). Transcriptome analysis of bagging-treated red Chinese sand pear peels reveals light-responsive pathway functions in anthocyanin accumulation. *Sci. Rep.* 7 (1), 63. doi: 10.1038/s41598-017-00069-z

Cai, Z. P., Huang, Z., Wang, Z. X., Tao, Y., Wu, F. H., Yu, X. D., et al. (2021). Identification of the related genes on the asymmetric root growth of *Oryza sativa* induced by ethylene through transcriptome sequencing, GO and KEGG analysis. *Acta Physiol Plantarum* 43 (7), 99. doi: 10.1007/s11738-021-03271-9

Cai, Z. P., Xie, Z. Y., Huang, L. Y., Wang, Z. X., Pan, M., Yu, X. D., et al. (2022a). Full-length transcriptome analysis of *Adiantum flabellulatum* gametophyte. *PeerJ* 10, e13079. doi: 10.7717/peerj.13079

Cai, Z. P., Xie, Z. Y., Wang, X. C., Zhang, S. X., Wu, Q., Yu, X. D., et al. (2022b). Excavation of genes responsive to brassinosteroids by transcriptome sequencing in *Adiantum flabellulatum* gametophytes. *Genes (Basel)* 13, (6). doi: 10.3390/genes13061061

Costarelli, A., Cannavo, S., Cerri, M., Pellegrino, R. M., Reale, L., Paolocci, F., et al. (2021). Light and temperature shape the phenylpropanoid profile of *Azolla filiculoides* fronds. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.727667

de Wit, M., Galvao, V. C., and Fankhauser, C. (2016). Light-mediated hormonal regulation of plant growth and development. *Annu. Rev. Plant Biol.* 67, 513–537. doi: 10.1146/annurev-arplant-043015-112252

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15 (10), 573–581. doi: 10.1016/j.tplants.2010.06.005

Ferreyra, M. L. F., Serra, P., and Casati, P. (2021). Recent advances on the roles of flavonoids as plant protective molecules after UV and high light exposure. *Physiol. Plant* 173 (3), 736–749. doi: 10.1111/ppl.13543

Friend, D. J. C. (1974). *Shade adaptation of the Hawaiian tree-fern (Cibotium glaucum* Vol. 41 (Honolulu (HI: Island Ecosystems IRP, U.S. International Biological Program. International Biological Program Technical Report), 39.

Gao, L. M., Liu, Y., Wang, X. F., Li, Y. F., and Han, R. (2019). Lower levels of UV-B light trigger the adaptive responses by inducing plant antioxidant metabolism and flavonoid biosynthesis in *Medicago sativa* seedlings. *Funct. Plant Biol.* 46 (10), 896–906. doi: 10.1071/FP19007

Gonzalez, A., Brown, M., Hatlestad, G., Akhavan, N., Smith, T., Hembd, A., et al. (2016). TTG2 controls the developmental regulation of seed coat tannins in *Arabidopsis* by regulating vacuolar transport steps in the proanthocyanidin pathway. *Dev. Biol.* 419 (1), 54–63. doi: 10.1016/j.ydbio.2016.03.031

Guo, H. W., Yang, H. Y., Mockler, T. C., and Lin, C. T. (1998). Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* 279 (5355), 1360–1363. doi: 10.1126/science.279.5355.1360

Jin, J. P., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J. C., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45 (D1), D1040–D1045. doi: 10.1093/nar/gkw982

Johnson, G. N., Rumsey, F. J., Headley, A. D., and Sheffield, E. (2000). Adaptations to extreme low light in the fern *TrichOmanes speciosum*. *New Phytol.* 148 (3), 423–431. doi: 10.1046/j.1469-8137.2000.00772.x

Krieg, C. P., and Chambers, S. M. (2022). The ecology and physiology of fern gametophytes: A methodological synthesis. *Appl. Plant Sci.* 10, (2). doi: 10.1002/aps3.11464

Li, F. W., Brouwer, P., Carretero-Paulet, L., Cheng, S. F., de Vries, J., Delaux, P. M., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* 4 (7), 460–472. doi: 10.1038/s41477-018-0188-8

Liao, J. X., Zhang, H., Mo, L., Huang, Y. Q., Sun, Y. J., and Li, Y. Q. (2017). Differences in growth and biomass allocation of *Adiantum flabellulatum* and *A. capillus-veneris* as a result of light and water availability. *Bot. Lett.* 164 (4), 393–400. doi: 10.1080/23818107.2017.1396496

Liu, L., Shu, J. P., Wei, H. P., Zhang, R., Shen, H., and Yan, Y. H. (2016). *De novo* transcriptome analysis of the rare fern *Monachosorum maximowiczii* (Dennstaedtiaceae) endemic to East Asia. *Biodiversity Sci.* 24, 1325–1334. doi: 10.17520/biods.2016231

Meteignier, L. V., Ghandour, R., Zimmerman, A., Kuhn, L., Meurer, J., Zoschke, R., et al. (2021). *Arabidopsis* mTERF9 protein promotes chloroplast ribosomal assembly and translation by establishing ribonucleoprotein interactions in *vivo*. *Nucleic Acids Res.* 49 (2), 1114–1132. doi: 10.1093/nar/gkaa1244

Miller, J. H., and Miller, P. M. (1961). The effect of different light conditions and sucrose on the growth and development of the gametophyte of the fern. Onoclea sensibilis. *Botanical Soc. America* 48 (2), 154–159. doi: 10.1002/j.1537-2197.1961.tb11619.x

Nitta, J. H., Watkins, J. E., Holbrook, N. M., Wang, T. W., and Davis, C. C. (2021). Ecophysiological differentiation between life stages in filmy ferns (Hymenophyllaceae). *J. Plant Res.* 134 (5), 971–988. doi: 10.1007/s10265-021-01318-z

Oakenfull, R. J., and Davis, S. J. (2017). Shining a light on the *Arabidopsis* circadian clock. *Plant Cell Environ.* 40 (11), 2571–2585. doi: 10.1111/pce.13033

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29 (9), e45. doi: 10.1093/nar/29.9.e45

Poorter, H., Niinemets, U., Ntagkas, N., Siebenkas, A., Maenpaa, M., Matsubara, S., et al. (2019). A meta-analysis of plant responses to light intensity for 70 traits ranging from molecules to whole plant performance. *New Phytol.* 223 (3), 1073–1105. doi: 10.1111/nph.15754

Roberti, M., Polosa, P. L., Bruni, F., Manzari, C., Deceglie, S., Gadaleta, M. N., et al. (2009). The MTERF family proteins: mitochondrial transcription regulators and beyond. *Biochim. Biophys. Acta* 1787 (5), 303–311. doi: 10.1016/j.bbabio.2009.01.013

Robles, P., Micol, J. L., and Quesada, V. (2015). Mutations in the plant-conserved MTERF9 alter chloroplast gene expression, development and tolerance to abiotic stress in *Arabidopsis thaliana*. *Physiol. Plant* 154 (2), 297–313. doi: 10.1111/ppl.12307

Robles, P., Nunez-Delegido, E., Ferrandez-Ayela, A., Sarmiento-Manus, R., Micol, J. L., and Quesada, V. (2018). *Arabidopsis* mTERF6 is required for leaf patterning. *Plant Sci.* 266, 117–129. doi: 10.1016/j.plantsci.2017.11.003

Romani, I., Manavski, N., Morosetti, A., Tadini, L., Maier, S., Kuhn, K., et al. (2015). A member of the *Arabidopsis* Mitochondrial Transcription Termination Factor family is required for maturation of chloroplast transfer RNAIle (GAU). *Plant Physiol.* 169 (1), 627–646. doi: 10.1104/pp.15.00964

Santos, E. L., Maia, B.H.L.N.S., Ferriani, A. P., and Teixeira, S. D. (2017). *Flavonoids: Classification, biosynthesis and chemical ecology*. In flavonoids : from biosynthesis to human health. (London, UK: IntechOpen) 13, 78–94. doi: 10.5772/67861

Schneider, H., Schuettpelz, E., Pryer, K. M., Cranfill, R., Magallon, S., and Lupia, R. (2004). Ferns diversified in the shadow of angiosperms. *Nature* 428 (6982), 553–557. doi: 10.1038/nature02361

Sessa, E. B., and Der, J. P. (2016). Evolutionary genomics of ferns and lycophytes. *Adv. Bot. Res.* 78, 215–254. doi: 10.1016/bs.abr.2016.02.001

Shahriar, M., and Kabir, S. (2011). Analgesic activity of *Adiantum flabellulatum*. *Dhaka Univ. J. Biol. Sci.* 20, 91–93. doi: 10.3329/dujbs.v20i1.8877

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi: 10.1101/gr.1239303

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., et al. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50 (4), 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

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

Sun, L. L., Li, Y., Li, S. S., Wu, X. J., Hu, B. Z., and Chang, Y. (2014). Identification and characterisation of *DfCHS*, a chalcone synthase gene regulated by temperature and ultraviolet in *Dryopteris fragrans*. *Cell Mol. Biol. (Noisy-le-grand)* 60 (6), 1–7. doi: 10.14715/cmb/2014.60.6.1

Takahashi, S., and Badger, M. R. (2011). Photoprotection in plants: a new light on photosystem II damage. *Trends Plant Sci.* 16 (1), 53–60. doi: 10.1016/j.tplants.2010.10.001

Takos, A. M., Jaffe, F. W., Jacob, S. R., Bogs, J., Robinson, S. P., and Walker, A. R. (2006). Light-induced expression of a *MYB* gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol.* 142 (3), 1216–1232. doi: 10.1104/pp.106.088104

Tian, F., Yang, D. C., Meng, Y. Q., Jin, J., and Gao, G. (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* 48 (D1), D1104–D1113. doi: 10.1093/nar/gkz1020

Tian, T., You, Q., Zhang, L., Yi, X., Yan, H. Y., Xu, W. Y., et al. (2016). SorghumFDB: sorghum functional genomics database with multidimensional network analysis. *Database* 2016, baw099. doi: 10.1093/database/baw099

Wang, N., Xu, H. F., Jiang, S. H., Zhang, Z. Y., Lu, N. L., Qiu, H. R., et al. (2017). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant J.* 90 (2), 276–292. doi: 10.1111/tpj.13487

Xu, W. J., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* 20 (3), 176–185. doi: 10.1016/j.tplants.2014.12.001

Xu, D. R., Leister, D., and Kleine, T. (2017). *Arabidopsis thaliana* mTERF10 and mTERF11, but not mTERF12, are involved in the response to salt stress. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01213

Yang, X. B. (2015). *"Pteridaceae" in Flora of Hainan: Volume I*. Ed. D. M. Yang, et al (Beijing: Science Press), 156.

Yang, M. Q., You, W. J., Wu, S. W., Fan, Z., Xu, B. F., Zhu, M. L., et al. (2017). Global transcriptome analysis of *Huperzia serrata* and identification of critical genes involved in the biosynthesis of huperzine A. *BMC Genomics* 18 (1), 245. doi: 10.1186/s12864-017-3615-8

Zhang, Y., Cui, Y. L., Zhang, X. L., Yu, Q. B., Wang, X., Yuan, X. B., et al. (2018). A nuclear-encoded protein, mTERF6, mediates transcription termination of rpoA polycistron for plastid-encoded RNA polymerase-dependent chloroplast gene expression and chloroplast development. *Sci. Rep.* 8 (1), 11929. doi: 10.1038/s41598-018-30166-6

Zheng, X. T., Yu, Z. C., Tang, J. W., Cai, M. L., Chen, Y. L., Yang, C. W., et al. (2021). The major photoprotective role of anthocyanins in leaves of *Arabidopsis thaliana* under long-term high light treatment: antioxidant or light attenuator? *Photosynth Res.* 149 (1-2), 25–40. doi: 10.1007/s11120-020-00761-8

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 October 2023
DOI 10.3389/fpls.2023.1245433

Check for updates

OPEN ACCESS

EDITED BY
Eleni Tani,
Agricultural University of Athens, Greece

REVIEWED BY
Panagiotis Madesis,
University of Thessaly, Greece
Alma Balestrazzi,
University of Pavia, Italy

*CORRESPONDENCE
Ester Stajič
ester.stajic@bf.uni-lj.si

RECEIVED 23 June 2023
ACCEPTED 13 September 2023
PUBLISHED 02 October 2023

CITATION
Stajič E and Kunej U (2023) Optimization of cabbage (*Brassica oleracea* var. *capitata* L.) protoplast transformation for genome editing using CRISPR/Cas9.
*Front. Plant Sci.* 14:1245433.
doi: 10.3389/fpls.2023.1245433

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

# Optimization of cabbage (*Brassica oleracea* var. *capitata* L.) protoplast transformation for genome editing using CRISPR/Cas9

Ester Stajič* and Urban Kunej

Department of Agronomy, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

Genome editing techniques, such as Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated systems (CRISPR/Cas9) are undoubtedly becoming an indispensable tool for improving food crops and tackling agricultural challenges. In the present study, key factors affecting transformation efficiency, such as PEG4000 concentration, incubation time, and plasmid amount were evaluated to achieve efficient delivery of CRISPR/Cas9 vector into cabbage protoplasts. Using amplicon sequencing, we confirmed a significant effect of PEG4000 concentration and incubation time on the induced target mutations. By optimizing the transformation protocol, editing efficiency of 26.4% was achieved with 40 µg of plasmid and 15 minutes incubation with 50% PEG4000. While these factors strongly affected the mutation rate, the viability of the transformed protoplasts remained high. Our findings would be useful for successful genome editing in cabbage and other brassicas, as well as in research areas such as gene function analysis and subcellular localization that rely on transient transformation methods in protoplasts.

KEYWORDS

**CRISPR/Cas9, genome editing, protoplasts, PEG-mediated transformation, B. oleracea**

## 1 Introduction

*Brassica oleracea* includes many important vegetable crops with high nutritional value such as kale, cauliflower, cabbage, broccoli and kohlrabi. Cabbage (*Brassica oleracea* var. *capitata* L.) is an important source of vitamins, folic acid, flavonoids and calcium. In addition, it contains phytochemicals (glucosinolates) that are of great benefit to human health as they have anti-inflammatory and anti-cancer effects (Šamec et al., 2017). Cabbage is a biennial plant with an obligate requirement for vernalization and is self-incompatible (Ma et al., 2019). Therefore, breeding by conventional methods is a lengthy process, and modern biotechnology methods are needed to overcome these limitations.

Doubled haploid (DH) technology by microspore isolation is used in cabbage breeding to shorten the breeding process of elite F1 hybrids. However, regeneration of DH plants is highly dependent on genotype (Rudolf et al., 1999). Haploids could be obtained using haploid inducer lines with mutant alleles that trigger chromosome elimination after pollination, an approach commonly used in maize and other cereals (Ren et al., 2017). One of the genes involved in haploid induction is *CENH3* (centromere-specific histone H3). The role of CENH3 protein in kinetochore formation and the effects of modifications in this protein on chromosome segregation and production of spontaneous haploids after pollination have been described in many publications (Ravi and Chan, 2010; Karimi-Ashtiyani et al., 2015; Kuppu et al., 2015; Kuppu et al., 2020).

Genome editing techniques such as CRISPR/Cas9 can be used to make precise changes in specific regions of the genome. Applications of these methods range from basic biological studies of gene function to agricultural breeding and the development of new plant varieties with improved genomic traits or suppressed undesirable functions (Zhu et al., 2020). *Agrobacterium*-mediated transformations are currently predominant to introduce CRISPR/Cas9 reagents into plant cells (Laforest and Nadakuduti, 2022). However, stable transformation protocols have many limitations as they are time-consuming and labor-intensive. Moreover, to develop an efficient CRISPR/Cas9 protocol, various factors affecting the mutation rate need to be evaluated before practical use. The protoplast-based transient transformation system provides a rapid and efficient platform to validate multiple mutagenesis parameters in a short time. Moreover, protoplasts can be further regenerated into plants (Lin et al., 2018; Nadakuduti et al., 2019). Protoplast isolation and transformation protocols have been reported for numerous agriculturally important crops (He et al., 2016; Xiong et al., 2019; Wang et al., 2021). Some studies have also been published for the genus *Brassica* (Murovec et al., 2018; Stajič et al., 2019; Li et al., 2021; Sivanandhan et al., 2021); however, the effects of PEG-transformation parameters on editing efficiency have not been investigated in detail.

In the present study, we aimed to optimize the conditions for PEG-mediated protoplast transformation to enhance editing efficiency using CRISPR/Cas9 in a non-model horticultural plant by targeting the cabbage *CENH3* gene associated with haploid induction. Amplicon sequencing was used to determine the mutation rate in different transformation experiments. In addition, high viability of transformed protoplasts was confirmed by FDA staining.

# 2 Materials and methods

## 2.1 Plant material

Two cabbage cultivars, namely 'Rebecca F1' and 'Reball F1' (Syngenta) were used in this study. Seeds were sterilized in a 1.66% (w/v) solution of dichloroisocyanuric acid (Sigma) for 15 min and then rinsed three times with sterile distilled water. Seeds were placed for germination on MS (Murashige and Skoog, 1962) culture medium (MS basal salts plus vitamins, 3% sucrose, 0.8% agar, pH 5.8) in a growth chamber with a 16-hour light/8-hour night photoperiod at 21°C. After one week, seedlings were transferred to ECO2 plastic containers with the same medium and maintained at 21°C until plantlets with developed leaves were obtained.

## 2.2 Plasmid preparation

Single-guide RNAs (sgRNAs) targeting the cabbage *CENH3* gene were designed using CRISPR RGEN tools (Park et al., 2015). To clone sgRNAs, primers (Supplementary Table 1) were annealed and ligated into pHSN401 (Addgene #50588) using T4 DNA ligase (NEB) and *Bsa*I-HF (NEB) in a Golden Gate reaction. To confirm correct vector construction, we used Sanger sequencing. All vector construction procedures were performed according to Xing et al. (2014).

## 2.3 Protoplast isolation

Protoplasts were isolated according to the protocol of Kiełkowska and Adamus (2012) with minor modifications. Briefly, 1.5 g of young leaves were cut with a sharp razor in a 9-cm Petri dish containing 15 ml of plasmolysis solution (Table 1) and incubated for 1 h at 25°C in the dark. After pretreatment, the solution was discarded and 8 ml of enzyme solution was added to the cut leaves. The material was digested overnight at 25°C while gently shaking. The next day, the digested material was filtered

TABLE 1 Solutions and their composition used for cabbage protoplast isolation and transformation.

| Solution name | Composition |
|---|---|
| Plasmolysis solution | 0.5 M mannitol (pH 5.7) |
| Enzyme solution | 0.5% cellulase Onozuka RS (Yakult), 0.1% macerozyme R-10 (Duchefa), 0.4 M mannitol, 3 mM $CaCl_2$, 2 mM MES (pH 5.6) |
| SAH/MES solution | 0.5 M sucrose, 1 mM MES (pH 5.7) |
| W5 solution | 154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES (pH 5.7) |
| MMG solution | 0.4 M mannitol, 15 mM $MgCl_2$, 4 mM MES (pH 5.7) |
| PEG solution | 0.2 M mannitol, 100 mM $CaCl_2$, 10-50% PEG4000 |

through a 40-μm cell strainer and centrifuged at 900 RPM for 5 min. The supernatant was discarded and the protoplast pellet was resuspended in 8 ml of SAH/MES solution and overlaid with 2 ml of W5 solution. After centrifugation at 1,100 RPM for 10 min, the protoplasts from interphase were transferred to a fresh centrifuge tube and rinsed with 10 ml of W5 solution. The protoplast pellet was resuspended in MMG solution at a density of $2.5 \times 10^6$ protoplasts/ml. The viability of isolated protoplasts was evaluated by staining with 0.01% fluorescein diacetate (FDA). The percentage of viable protoplasts was calculated based on the number of protoplasts with green fluorescence divided by the total number of cells. Viability assessment was repeated three times.

## 2.4 PEG-mediated transformation of protoplasts

For protoplast transformation, plasmid (20-80 μg) was added to 200 μl of protoplasts, followed by addition of equal volume of PEG solution. The tubes were mixed gently and incubated in the dark at room temperature for 5 s to 25 min, depending on the protocol tested. After incubation, the reaction was immediately stopped by adding an equal volume of W5 solution followed by two volumes of W5 solution. Samples were then centrifuged at 700 RPM for 5 min, and the supernatant was removed. The transformed protoplasts were washed a final time with 10 ml of W5 solution and resuspended in 1 ml of the same solution. Protoplasts were transferred to 24-well plates (Sigma) and incubated in the dark at room temperature for 48 h. After incubation, protoplast viability was determined as described above, and protoplasts were collected by centrifugation at 12,000 RPM before DNA isolation was performed.

## 2.5 Genomic DNA isolation and Ampliseq

Genomic DNA was extracted from the pellet of protoplasts using E.Z.N.A.® Plant DNA DS Kit (Omega BIO-TEK). Target sites were amplified with Q5® High-Fidelity DNA Polymerase (NEB) using the primers listed in Supplementary Table 1. The same target-specific primers followed by unique barcode were used for each sample. PCR was performed with initial denaturation for 30 s at 98°C, 35 cycles of denaturation for 5 s at 98°C, annealing for 10 s at 65°C, elongation for 10 s at 72°C and final elongation for 2 min at 72°C. PCR products were purified using GFX™ PCR DNA and Gel Band Purification Kit (Cytiva) according to the manufacturer protocol. Samples were then pooled and sequenced on the Illumina HiSeq at Eurofins Genomics Europe Sequencing GmbH (Konstanz, Germany). The obtained next-generation sequencing data were analyzed using Cas-analyzer software (Park et al., 2017) to determine mutation efficiency. Three biological replicates were performed for each experiment and results are presented as the mean ± standard error (SE). Detailed information is provided at the end of the corresponding tables and figures.

# 3 Results

## 3.1 Effect of PEG4000 incubation time on protoplast viability and mutation rate

High viability and number of protoplasts are one of the main requirements for successful protoplast transformation. Mesophyll protoplasts with high viability (93.6% in cultivar 'Reball F1' and 86.7% in cultivar 'Rebecca F1') were isolated from young leaves of cabbage plants grown *in vitro* after overnight digestion in 0.5% cellulase Onozuka RS and 0.1% macerozyme R-10. A sucrose density gradient was used to obtain debris-free protoplasts. The average yield of isolated protoplasts was $5.3 \times 10^6$ p/g (protoplasts per g of leaves) for the cultivar 'Reball F1' and $4.0 \times 10^6$ p/g for the cultivar 'Rebecca F1'.To optimize the PEG-mediated transformation protocol in cabbage protoplasts, the effect of incubation time with PEG4000 on protoplast viability and indel frequencies were investigated. No indel mutations could be detected in control samples transformed without vectors. Editing efficiency in 'Reball F1' protoplasts increased from 1.57% to 7.13% within 5 s to 15 min of incubation, but then decreased to 5.13% when protoplasts were incubated with PEG4000 for 25 min. In 'Rebecca F1', the detected mutation frequencies were low (0.5-1.8%) regardless of the incubation time (Figure 1). Incubation time with PEG4000 did not affect viability, and it remained high (over 83.3% on average) in both genotypes even when the exposure time was extended to 25 min (Figure 2).

Two sgRNAs targeting the same gene were also tested in 'Reball F1' protoplasts. The results showed that higher indel frequencies were obtained with sgRNA-CENH3-A than with sgRNA-CENH3-B (Figure 3; Table 2). The editing efficiencies ranged from 8.63% to 9.37% when protoplasts were transformed with sgRNA-CENH3-A and the incubation time was 5-25 min. The lowest mutation rate (5.07%) was observed after 5 s incubation with PEG4000.

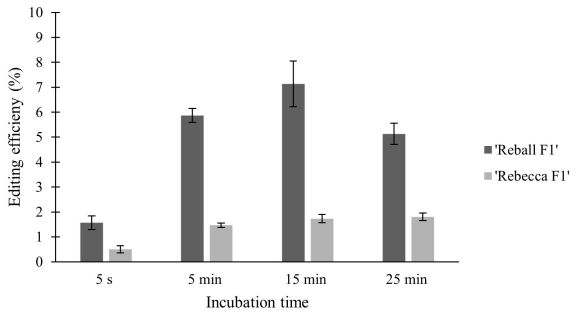



FIGURE 1
Effect of PEG4000 incubation time on editing efficiency in two cabbage cultivars. Values represent means ± SE of three replicates.

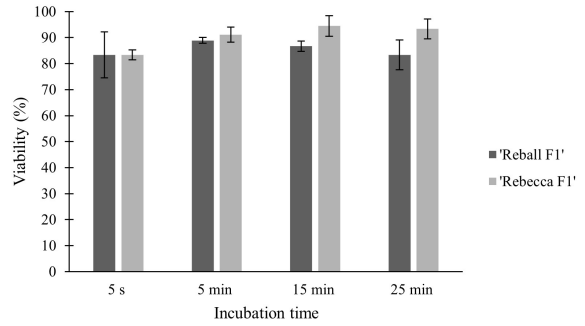



FIGURE 2
Effect of PEG4000 incubation time on viability of cabbage protoplasts. Values represent means ± SE of three replicates.

## 3.2 Effect of PEG4000 concentration on protoplast viability and mutation rate

In the second part of the study, protoplasts of both cultivars were transformed with different concentrations (10-50%) of PEG4000. The PEG4000 concentration substantially affected the efficiency of protoplast transformation. At 10%, the editing efficiency was 0.73% ('Reball F1') and 1.23% ('Rebecca F1'). With increasing PEG4000 concentration, the average mutation frequencies in 'Reball F1' and 'Rebecca F1' protoplasts increased to 13.43% and 25.23%, respectively (Figure 4; Table 3). No effect of PEG4000 concentration on the viability of transformed protoplasts of both genotypes was detected (Figure 5). Based on these results, 50% PEG4000 was chosen as the optimal PEG4000 concentration.

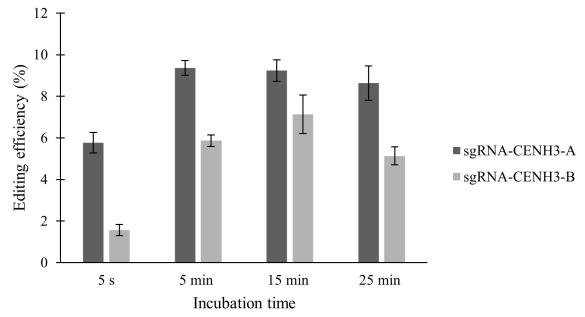



FIGURE 3
Effect of PEG4000 incubation time on editing efficiency in 'Reball F1' protoplasts with two different sgRNAs. Values represent means ± SE of three replicates.

## 3.3 Effect of plasmid amount on mutation rate and viability

After optimization of PEG4000 concentration and incubation time, different amounts (20 to 80 μg) of the plasmid were transformed into cabbage protoplasts. The results showed that for both cultivars, mutation rates increased when 40 μg of plasmid was added instead of 20 μg, but then decreased when the amount increased to 60 μg and 80 μg (Figure 6; Table 4), suggesting that 40 μg of plasmid might be the threshold and was chosen as the optimal plasmid amount for cabbage protoplast transformation. However, viability remained high (91.1% for 'Reball F1' and 83.3% for 'Rebecca F1') even when 80 μg of plasmid was used for transformation (Figures 7, 8). Figure 9 shows the most frequent alleles found around the cleavage site in the protoplast sample of 'Reball F1' transformed with 40 μg plasmid, 50% PEG4000, and 15 min incubation. The most frequent mutations were 1 bp insertions, followed by 3 bp deletions.

TABLE 2 Indel frequencies detected in the *CENH3* gene by amplicon sequencing after protoplast transformation with different incubation time.

| Genotype | Incubation time | sgRNA | No. of reads analyzed[a] | No. of deletions[a] | No. of insertions[a] | Indel frequency ± SE (%)[a] |
|---|---|---|---|---|---|---|
| 'Reball F1' | 5 s | CENH3-A | 254,061 | 8,186 | 3,656 | 5.07 ± 0.49 |
| | | CENH3-B | 35,669 | 185 | 311 | 1.57 ± 0.27 |
| | 5 min | CENH3-A | 283,542 | 19,810 | 6,511 | 9.37 ± 0.35 |
| | | CENH3-B | 70,485 | 1,393 | 2,723 | 5.87 ± 0.28 |
| | 15 min | CENH3-A | 180,306 | 12,474 | 4,391 | 9.23 ± 0.52 |
| | | CENH3-B | 66,817 | 1,582 | 3,009 | 7.13 ± 0.92 |
| | 25 min | CENH3-A | 191,875 | 12,086 | 4,482 | 8.63 ± 0.83 |
| | | CENH3-B | 23,493 | 487 | 720 | 5.13 ± 0.43 |

[a]Based on three biological replicates.

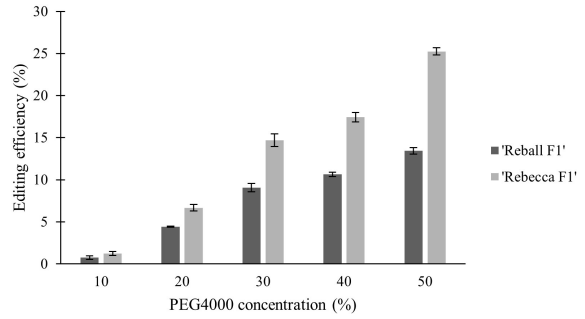



FIGURE 4
Effect of PEG4000 concentration on editing efficiency in two cabbage cultivars. Values represent means ± SE of three replicates.

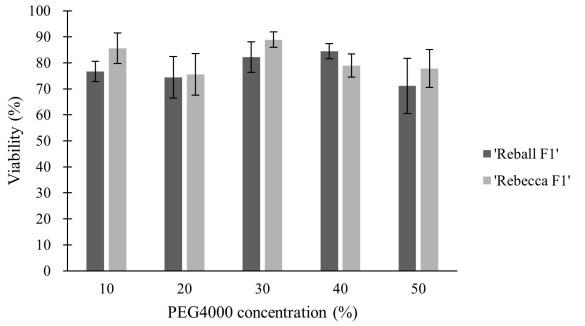



FIGURE 5
Effect of PEG4000 concentration on viability of cabbage protoplasts. Values represent means ± SE of three replicates.

## 4 Discussion

The rapidly growing world population and unpredictable environmental changes pose challenges to agricultural production that need to be addressed. Modern genetic tools such as CRISPR/Cas9 provide the opportunity to introduce precise genetic changes and develop new varieties with higher yields, better quality, and improved resistance (Zhu et al., 2020). However, a successful genome editing protocol depends on a combination of factors, and efficient delivery of genome editing reagents into plant cells remains the major bottleneck in plant genome editing (Laforest and Nadakuduti, 2022). Currently, protoplast transformation is gaining interest because it can be used in various areas of plant science research and protoplasts can also be regenerated into modified plants without the incorporation of foreign DNA (Nadakuduti et al., 2019; Park et al., 2019). For protoplast transformation and further regeneration into plants, yield, viability, and transformation efficiency are important factors.

In our experiments, protoplasts from two cabbage cultivars were isolated with high yield and viability, which indicates that our protoplast isolation protocol is suitable to obtain high-quality protoplasts appropriate for downstream applications. To optimize the PEG-mediated delivery of CRISPR/Cas9 plasmids, we targeted the *CENH3* gene of high breeding importance. Editing this gene could lead to the development of a haploid inducer line and the generation of haploids in recalcitrant genotypes (Kuppu et al., 2020). The efficiency of CRISPR/Cas9-induced mutations in protoplasts varies greatly, even among *Brassica* species. Murovec et al. (2018) detected only 2.25% of indel mutations in cabbage protoplasts when RNPs were introduced, whereas mutation rates were higher in *Brassica rapa* (up to 24.51%). In our previous report, editing efficiencies of 1.27-11.95% were obtained in cabbage protoplasts after plasmid transformation (Stajič et al., 2019). We therefore decided to test different transformation parameters (PEG4000 concentration, incubation time, and plasmid amount) to achieve higher editing efficiencies.

Our data suggest that prolonged incubation of 15 to 25 min with PEG4000 leads to a decrease in editing efficiency, which was confirmed with two different sgRNAs. In contrast, Sivanandhan

TABLE 3 Indel frequencies detected by amplicon sequencing after protoplast transformation with different PEG4000 concentrations.

| Genotype | PEG4000 concentration (%) | No. of reads analyzed[a] | No. of deletions[a] | No. of insertions[a] | Indel frequency ± SE (%)[a] |
|---|---|---|---|---|---|
| 'Reball F1' | 10 | 105,037 | 613 | 240 | 0.73 ± 0.20 |
| | 20 | 119,472 | 3,895 | 1,339 | 4.40 ± 0.06 |
| | 30 | 184,038 | 12,302 | 4,473 | 9.07 ± 0.49 |
| | 40 | 216,983 | 17,050 | 6,120 | 10.63 ± 0.27 |
| | 50 | 78,456 | 7,691 | 2,887 | 13.43 ± 0.38 |
| 'Rebecca F1' | 10 | 165,561 | 1,598 | 523 | 1.23 ± 0.24 |
| | 20 | 181,203 | 8,851 | 3,112 | 6.67 ± 0.37 |
| | 30 | 187,111 | 19,201 | 8,165 | 14.70 ± 0.76 |
| | 40 | 183,655 | 22,100 | 9,106 | 17.43 ± 0.57 |
| | 50 | 67,115 | 12,003 | 5,033 | 25.23 ± 0.43 |

[a]Based on three biological replicates.

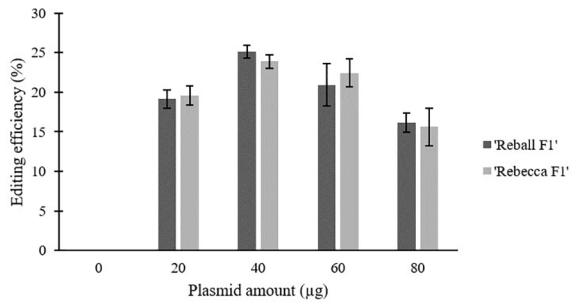



FIGURE 6
Effect of plasmid amount on editing efficiency in two cabbage cultivars. Values represent means ± SE of three replicates.

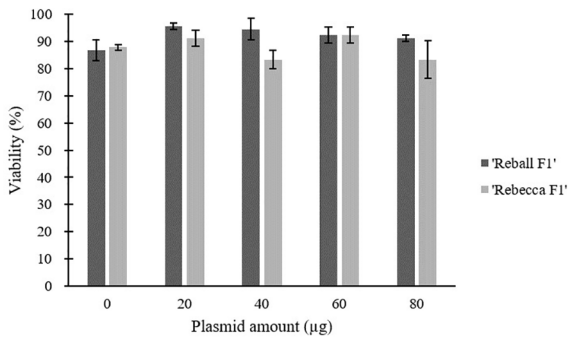



FIGURE 7
Effect of plasmid amount on viability of cabbage protoplasts. Values represent means ± SE of three replicates.

et al. (2021) found that incubation for 30 min instead of 20 min positively affected transformation efficiency in *Brassica rapa* protoplasts. In most protocols, 15 min of exposure time and 40% PEG4000 are taken as standard. PEG4000 concentration substantially influenced editing efficiency in our experiments and increased significantly when PEG4000 concentration was increased to 50%. For broccoli protoplasts, 20% PEG4000 was determined as the optimal concentration, as higher concentrations led to protoplast rupture (Yang et al., 2022). Similarly, a concentration of 25% PEG4000 was used for the transformation of *Brassica napus* protoplasts (Li et al., 2021). In our experiments, the viability of transformed protoplasts remained high even when 50% PEG4000 was used. Another important factor influencing transformation efficiency in protoplasts is amount of plasmid. Optimal dosage of plasmid differs greatly, in broccoli protoplasts, no differences in transformation efficiency were found when 5-15 µg of plasmid was used (Yang et al., 2022), while 40 µg of the plasmid was reported as optimal amount in *Brassica rapa* protoplasts (Sivanandhan et al., 2021). Transformation of cabbage protoplasts with 40 instead 20 µg of plasmid improved the editing efficiency in our experiments. However, higher amounts (60 and 80 µg) had a negative effect.

## 5 Conclusion

The results obtained in this study show that higher editing efficiency can be achieved by refining the transformation protocol. The data collected in our experiments show not only high viability of isolated protoplasts but also higher success than other current protocols in *Brassica oleracea*, with editing efficiency reaching 26.4% and 25.4% in 'Reball F1' and 'Rebecca F1' protoplasts, respectively, by subjecting protoplasts to optimized conditions (15 min incubation with 50% PEG4000 and 40 µg plasmid), while

TABLE 4 Indel frequencies detected in the *CENH3* gene by amplicon sequencing after protoplast transformation with different amounts of plasmid.

| Genotype | Plasmid amount (µg) | No. of reads analyzed[a] | No. of deletions[a] | No. of insertions[a] | Indel frequency ± SE (%)[a] |
|---|---|---|---|---|---|
| 'Reball F1' | 0 | 179,802 | 0 | 5 | 0.00 ± 0.00 |
| | 20 | 140,051 | 19,935 | 7,425 | 19.13 ± 1.19 |
| | 40 | 188,124 | 33,954 | 13,742 | 25.13 ± 0.78 |
| | 60 | 113,868 | 16,442 | 7,131 | 20.93 ± 2.70 |
| | 80 | 145,988 | 17,217 | 6,570 | 16.20 ± 1.21 |
| 'Rebecca F1' | 0 | 166,441 | 0 | 3 | 0.00 ± 0.00 |
| | 20 | 115,687 | 15,992 | 6,781 | 19.57 ± 1.19 |
| | 40 | 264,442 | 42,655 | 20,897 | 23.87 ± 0.87 |
| | 60 | 94,288 | 15,067 | 6,472 | 22.43 ± 1.76 |
| | 80 | 75,695 | 8,233 | 3,584 | 15.63 ± 2.38 |

[a]Based on three biological replicates.

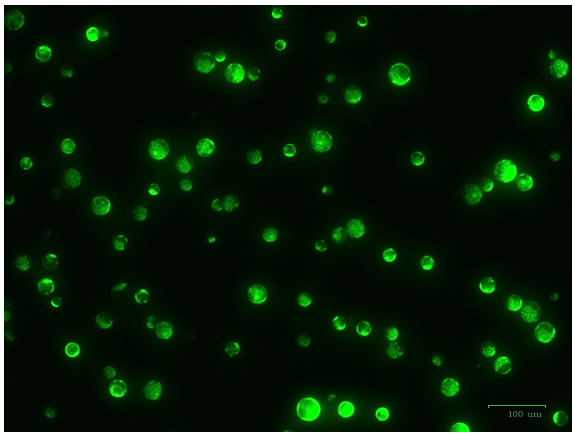



FIGURE 8
High viability of 'Reball F1' protoplasts 48 h after transformation with 80 µg plasmid and 15 min incubation with 50% PEG4000. Scale bar represents 100 µm.

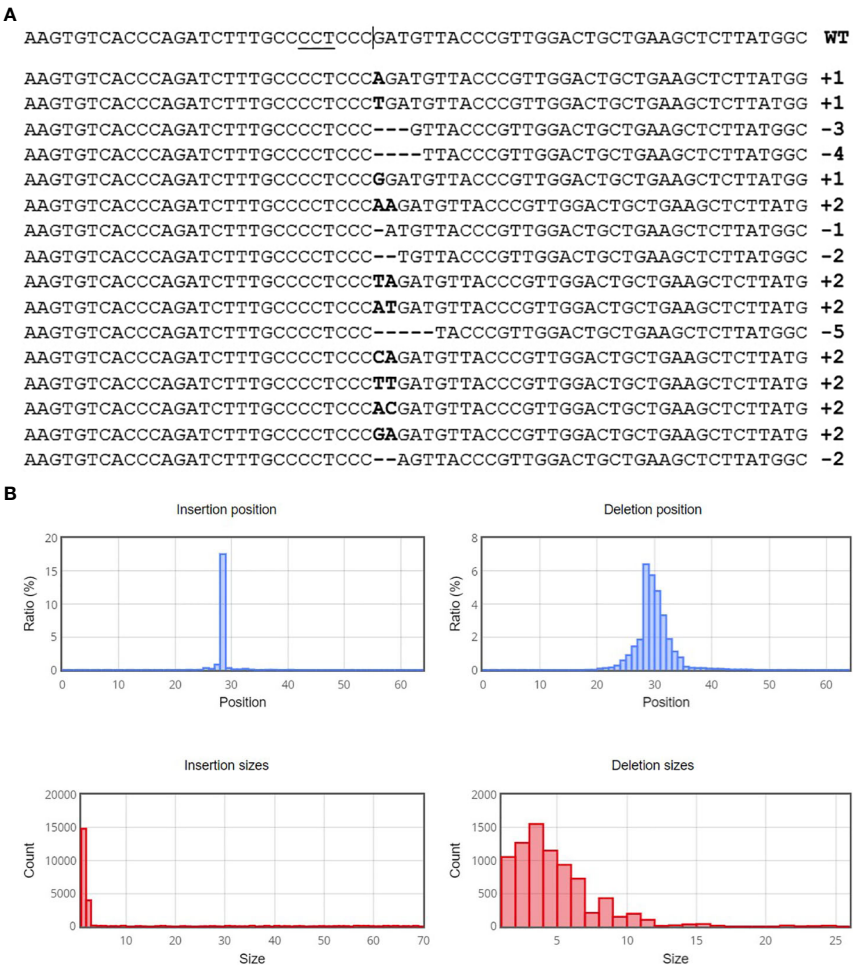



FIGURE 9
Sequence analysis with Cas-analyzer in 'Reball F1'protoplasts transformed with 40 µg plasmid, 50% PEG4000 and 15 min incubation. **(A)** Identified alleles around the predicted cleavage site (vertical line) in descending order by read count. Inserted nucleotides are shown in bold font and horizontal dashed lines indicate deletions. **(B)** Size of induced mutations and their distribution.

remaining high viability. All these data indicate that this isolation and transformation protocol is highly efficient and could be further used for plant regeneration from protoplast to obtain modified plants. Our optimized protocol could also be used for further optimization of editing efficiency (e.g. enhancing expression of CRISPR/Cas9 reagents) and would contribute to successful genome editing in other brassicas as well.

## 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

ES conceived the project. UK contributed to data analysis. ES wrote the first draft of the manuscript. Both authors performed the experiments and contributed to the manuscript revision. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Slovenian Research Agency (postdoctoral research project Z4-3215 and research programme P4-0077).

## Acknowledgments

We thank Gabriela Fretes Ocampos for assistance during protoplast transformation experiments.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1245433/full#supplementary-material

## References

He, F., Chen, S., Ning, Y., and Wang, G.-L. (2016). Rice (Oryza sativa) protoplast isolation and its application for transient expression analysis. *Curr. Protoc. Plant Biol.* 1, 373–383. doi: 10.1002/cppb.20026

Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., et al. (2015). Point mutation impairs centromeric CENH3 loading and induces haploid plants. *Proc. Natl. Acad. Sci. U. S. A.* 112, 11211–11216. doi: 10.1073/pnas.1504333112

Kiełkowska, A., and Adamus, A. (2012). An alginate-layer technique for culture of Brassica oleracea L. protoplasts. *Vitr. Cell. Dev. Biol. - Plant* 48, 265–273. doi: 10.1007/s11627-012-9431-6

Kuppu, S., Ron, M., Marimuthu, M. P. A., Li, G., Huddleson, A., Siddeek, M. H., et al. (2020). A variety of changes, including CRISPR/Cas9-mediated deletions, in CENH3 lead to haploid induction on outcrossing. *Plant Biotechnol. J.* 18, 2068–2080. doi: 10.1111/pbi.13365

Kuppu, S., Tan, E. H., Nguyen, H., Rodgers, A., Comai, L., Chan, S. W. L., et al. (2015). Point mutations in centromeric histone induce post-zygotic incompatibility and uniparental inheritance. *PloS Genet.* 11, 1–18. doi: 10.1371/journal.pgen.1005494

Laforest, L. C., and Nadakuduti, S. S. (2022). Advances in delivery mechanisms of CRISPR gene-editing reagents in plants. *Front. Genome Ed.* 4. doi: 10.3389/fgeed.2022.830178

Li, X., Sandgrind, S., Moss, O., Guan, R., Ivarson, E., Wang, E. S., et al. (2021). Efficient protoplast regeneration protocol and CRISPR/cas9-mediated editing of glucosinolate transporter (GTR) genes in rapeseed (Brassica napus L.). *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.680859

Lin, C. S., Hsu, C. T., Yang, L. H., Lee, L. Y., Fu, J. Y., Cheng, Q. W., et al. (2018). Application of protoplast technology to CRISPR/Cas9 mutagenesis: from single-cell mutation detection to mutant plant regeneration. *Plant Biotechnol. J.* 16, 1295–1310. doi: 10.1111/pbi.12870

Ma, C., Zhu, C., Zheng, M., Liu, M., Zhang, D., Liu, B., et al. (2019). CRISPR / Cas9-mediated multiple gene editing in Brassica oleracea var. capitata using the endogenous tRNA-processing system. *Hortic. Res.* 6, 1–15. doi: 10.1038/s41438-018-0107-1

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Murovec, J., Guček, K., Bohanec, B., Avbelj, M., and Jerala, R. (2018). DNA-Free Genome Editing of Brassica oleracea and B. rapa Protoplasts Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01594

Nadakuduti, S. S., Starker, C. G., Ko, D. K., Jayakody, T. B., Buell, C. R., Voytas, D. F., et al. (2019). Evaluation of methods to assess in *vivo* activity of engineered genome-editing nucleases in protoplasts. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00110

Park, J., Bae, S., and Kim, J.-S. (2015). Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* 31, 4014–4016. doi: 10.1093/bioinformatics/btv537

Park, J., Lim, K., Kim, J.-S., and Bae, S. (2017). Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* 33, 286–288. doi: 10.1093/bioinformatics/btw561

Park, S. C., Park, S., Jeong, Y. J., Lee, S. B., Pyun, J. W., Kim, S., et al. (2019). DNA-free mutagenesis of GIGANTEA in Brassica oleracea var. capitata using CRISPR/Cas9 ribonucleoprotein complexes. *Plant Biotechnol. Rep.* 13, 483–489. doi: 10.1007/s11816-019-00585-6

Ravi, M., and Chan, S. W. L. (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature* 464, 615–618. doi: 10.1038/nature08842

Ren, J., Wu, P., Trampe, B., Tian, X., Lübberstedt, T., and Chen, S. (2017). Novel technologies in doubled haploid line development. *Plant Biotechnol. J.* 15, 1361–1370. doi: 10.1111/pbi.12805

Rudolf, K., Bohanec, B., and Hansen, M. (1999). Microspore culture of white cabbage, Brassica oleracea var. capitata L.: Genetic improvement of non-responsive cultivars and effect of genome doubling agents. *Plant Breed.* 118, 237–241. doi: 10.1046/j.1439-0523.1999.118003237.x

Šamec, D., Pavlović, I., and Salopek-Sondi, B. (2017). White cabbage (Brassica oleracea var. capitata f. alba): botanical, phytochemical and pharmacological overview. *Phytochem. Rev.* 16, 117–135. doi: 10.1007/s11101-016-9454-4

Sivanandhan, G., Bae, S., Sung, C., Choi, S.-R., Lee, G.-J., and Lim, Y.-P. (2021). Optimization of Protoplast Isolation from Leaf Mesophylls of Chinese Cabbage (Brassica rapa ssp. pekinensis) and Subsequent Transfection with a Binary Vector. *Plants (Basel Switzerland)* 10. doi: 10.3390/plants10122636

Stajič, E., Kiełkowska, A., Murovec, J., and Bohanec, B. (2019). Deep sequencing analysis of CRISPR/Cas9 induced mutations by two delivery methods in target model genes and the CENH3 region of red cabbage (Brassica oleracea var. capitata f. rubra). *Plant Cell Tissue Organ Cult.* 139, 227–235. doi: 10.1007/s11240-019-01665-9

Wang, Q., Yu, G., Chen, Z., Han, J., Hu, Y., and Wang, K. (2021). Optimization of protoplast isolation, transformation and its application in sugarcane (Saccharum spontaneum L). *Crop J.* 9, 133–142. doi: 10.1016/j.cj.2020.05.006

Xing, H. L., Dong, L., Wang, Z. P., Zhang, H. Y., Han, C. Y., Liu, B., et al. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol.* 14, 327. doi: 10.1186/s12870-014-0327-y

Xiong, L., Li, C., Li, H., Lyu, X., Zhao, T., Liu, J., et al. (2019). A transient expression system in soybean mesophyll protoplasts reveals the formation of cytoplasmic GmCRY1 photobody-like structures. *Sci. China Life Sci.* 62, 1070–1077. doi: 10.1007/s11427-018-9496-5

Yang, D., Zhao, Y., Liu, Y., Han, F., and Li, Z. (2022). A high-efficiency PEG-Ca2+-mediated transient transformation system for broccoli protoplasts. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1081321

Zhu, H., Li, C., and Gao, C. (2020). Applications of CRISPR-Cas in agriculture and plant biotechnology. *Nat. Rev. Mol. Cell Biol.* 21, 661–677. doi: 10.1038/s41580-020-00288-9

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 October 2023
DOI 10.3389/fpls.2023.1268537

Check for updates

OPEN ACCESS

EDITED BY
Christos Bazakos,
Max Planck Institute for Plant Breeding Research, Germany

REVIEWED BY
Michail Michailidis,
Aristotle University of Thessaloniki, Greece
Yuan Linxi,
Xi'an Jiaotong-Liverpool University, China

*CORRESPONDENCE
Lu Wang
wanglu317@tricaas.com
Jianming Zeng
zengjm@tricaas.com

†These authors contributed equally to this work

RECEIVED 28 July 2023
ACCEPTED 08 September 2023
PUBLISHED 02 October 2023

CITATION
Zheng Q, Guo L, Huang J, Hao X, Li X, Li N, Wang Y, Zhang K, Wang X, Wang L and Zeng J (2023) Comparative transcriptomics provides novel insights into the mechanisms of selenium accumulation and transportation in tea cultivars (*Camellia sinensis* (L.) O. Kuntze).
*Front. Plant Sci.* 14:1268537.
doi: 10.3389/fpls.2023.1268537

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

# Comparative transcriptomics provides novel insights into the mechanisms of selenium accumulation and transportation in tea cultivars (*Camellia sinensis* (L.) O. Kuntze)

Qinghua Zheng†, Lina Guo†, Jianyan Huang, Xinyuan Hao, Xiaoman Li, Nana Li, Yueqi Wang, Kexin Zhang, Xinchao Wang, Lu Wang* and Jianming Zeng*

Key Laboratory of Biology, Genetics and Breeding of Special Economic Animals and Plants, Ministry of Agriculture and Rural Affairs, National Center for Tea Plant Improvement, Tea Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou, China

Tea plants (*Camellia sinensis*) show discrepancies in selenium accumulation and transportation, the molecular mechanisms of which are not well understood. Hence, we aimed to conduct a systematic investigation of selenium accumulation and transportation mechanisms in different tea cultivars via transcriptome analysis. The $Na_2SeO_3$ and $Na_2SeO_4$ treatments improved selenium contents in the roots and leaves of three tea cultivars. The high selenium-enrichment ability (HSe) tea cultivars accumulated higher selenium contents in the leaves than did the low selenium-enrichment ability (LSe) tea cultivars. Transcriptome analysis revealed that differentially expressed genes (DEGs) under the $Na_2SeO_3$ and $Na_2SeO_4$ treatments were enriched in flavonoid biosynthesis in leaves. DEGs under the $Na_2SeO_3$ treatment were enriched in glutathione metabolism in the HSe tea cultivar roots compared to those of the LSe tea cultivar. More transporters and transcription factors involved in improving selenium accumulation and transportation were identified in the HSe tea cultivars under the $Na_2SeO_3$ treatment than in the $Na_2SeO_4$ treatment. In the HSe tea cultivar roots, the expression of sulfate transporter 1;2 (*SULTR1;2*) and *SULTR3;4* increased in response to $Na_2SeO_4$ exposure. In contrast, ATP-binding cassette transporter genes (*ABC*s), glutathione *S*-transferase genes (*GST*s), phosphate transporter 1;3 (*PHT1;3*), nitrate transporter 1 (*NRT1*), and 34 transcription factors were upregulated in the presence of $Na_2SeO_3$. In the HSe tea cultivar leaves, ATP-binding cassette subfamily B member 11 (*ABCB11*) and 14 transcription factors were upregulated under the $Na_2SeO_3$ treatment. Among them, *WRKY75* was explored as a potential transcription factor that regulated the

accumulation of $Na_2SeO_3$ in the roots of HSe tea cultivars. This study preliminary clarified the mechanism of selenium accumulation and transportation in tea cultivars, and the findings have important theoretical significance for the breeding and cultivation of selenium-enriched tea cultivars.

KEYWORDS

tea cultivars, high selenium-enrichment ability, transcriptome analyses, $Na_2SeO_3$, $Na_2SeO_4$

## 1 Introduction

Selenium (Se) is an essential trace element in the human body that plays an important role in immune regulation and disease prevention (Xiang et al., 2022). Organic selenium has higher bioavailability and fewer toxic side effects in humans than those caused by inorganic selenium such as $Na_2SeO_4$ and is an important means for Se uptake (Wang et al., 2022b). Plants are important organic selenium sources for the human body; they uptake inorganic selenium from the soil through their roots and convert it into absorbable organic selenium (Ren et al., 2022). People mainly consume Se-rich grains and horticultural crops as a source of Se (Ye et al., 2015; Chao et al., 2022). Nevertheless, the global distribution of Se resources in the soil is extremely uneven. More than 15% of the world population suffers from Kashin–Beck disease and Keshan disease, and approximately 72% of the soil in China is Se-deficient (Chen et al., 2022). Therefore, in low-Se areas, Se biofortification techniques such as soil fertilization and foliar spraying are utilized to enhance Se uptake in plants (Zhang et al., 2013; Xiong et al., 2019; Zhao et al., 2019; Huang et al., 2020). Meanwhile, the Se contents of plants were also increased by cultivating the Se-enrichment cultivars (Cabannes et al., 2011).

The main forms of Se in the soil are selenate and selenite, and their ratio in the soil is controlled mainly by soil redox state and pH (Elrashidi et al., 1987; Wang et al., 2022a). The accumulation and transportation of these two forms of Se differ among plants; then, both are eventually metabolized into Se compounds (Raina et al., 2021). In higher plants, the uptake of $Na_2SeO_4$ occurs mainly through sulfate transport into the plant, which is assimilated by the sulfur assimilation pathway (El Mehdawi et al., 2018; Yu et al., 2019). Studies on the model organism *Arabidopsis thaliana* have found that sulfate transporters (SULTR1;1 and SULTR1;2) participate in $Na_2SeO_4$ transport, and SULTR1;1 increases resistance to $Na_2SeO_4$ (Yoshimoto et al., 2002; Barberon et al., 2008). On the contrary, the accumulation and transportation mechanisms of $Na_2SeO_3$ in plants are relatively complex and currently unclear. A previous study found that $Na_2SeO_3$ entered plants primarily via passive diffusion (Shrift and Ulrich, 1969). However, Broyer et al. found that an increase in phosphate concentration can inhibit plant root accumulation of $Na_2SeO_3$ (Hopper and Parker, 1999). The accumulation of $Na_2SeO_3$ is an active process in plant roots, with a similar accumulation mechanism to that of phosphorus, and both phosphate transporters (OsPHT1;2 and OsPHT1;8) are involved in $Na_2SeO_3$ accumulation and transportation in rice (Li et al., 2008; Zhang et al., 2014). There is also evidence that the aquaporin NIP2;1 aids in the accumulation of $Na_2SeO_3$ in rice (Zhao et al., 2010). Recent studies have shown that nitrate transporter (NRT1.1B) promotes the transport of selenomethionine (SeMet) in rice (Zhang et al., 2019). However, the molecular mechanisms of Se accumulation and transportation in various tea cultivars with high and low Se-enrichment abilities have not been vastly investigated. As tea can be an important Se source, it is necessary to select and breed tea cultivars with high Se-enrichment ability and explore the discrepant mechanisms between high and low Se accumulation.

As an important cash crop, the tea plant (*Camellia sinensis* (L.) O. Kuntze) has a high ability to enrich Se (Ren et al., 2022). Research on Se in tea plants has mainly focused on the effects of exogenous Se on the quality and yield of tea, the effects of soil factors on the accumulation of Se, and the mechanism of Se tolerance (Tang et al., 2012; Zhou et al., 2015). Expression analysis of genes related to Se accumulation and transportation in tea has mainly focused on sulfate and phosphate transporters. CsSULTR1;1, CsSULTR1;2, and CsSULTR1;3 have been suggested to play important roles in Se accumulation (Zhang et al., 2022). Cao et al. found that CsPHTs might play vital roles in $Na_2SeO_3$ accumulation, transportation, and homeostasis (Cao et al., 2018; Cao et al., 2021). During transcriptome analysis, *PHT3;1a*, *PHT1;3b*, *PHT1;8*, and *NIP2;1* were found to be upregulated under the $Na_2SeO_3$ treatment (Ren et al., 2022). However, there are few reports on the regulatory network and function of Se accumulation and transportation in tea plants. Hence, in this study, we aimed to gain more insights into the roles of key genes by systematically analyzing the accumulation and transportation mechanisms of Se in various tea cultivars.

In this study, three tea cultivars with contrasting Se accumulation capacities were used as experimental materials to explore by transcriptome analysis the key genes related to Se accumulation and transportation. The research provided a theoretical basis for the selection and breeding of high Se-enrichment ability tea cultivars.

# 2 Materials and methods

## 2.1 Plant material and treatments

Various cultivars were grown in Hangzhou and Enshi and bred continuously for 3 years from 2019 to 2022 as part of a pioneer field experiment. The field experiment results showed that the varieties 'Zhongcha xicha 1 hao (XC 1)' and 'Zhongcha xicha 4 hao (XC 4)' had high selenium-enrichment ability (HSe) and that the variety 'CT2009-0025' had low selenium-enrichment ability (LSe). Therefore, three cultivars were selected as the experimental materials for this study. One-year-old tea plant cuttings were cultured in 1/4 nutrient solution (macronutrients mmol/L: N 2.0, P 0.07, K 0.6, Mg 0.67, Ca 0.53, and Al 0.07; microelements μmol/L: Fe 4.2, Mn 1, Zn 0.67, Cu 0.13, B 7, and Mo 0.33). The plants were grown in a greenhouse with 12-h illumination, at 25°C and 70% relative humidity. The nutrient solution was replaced once weekly until new roots emerged. The cuttings were then transferred to treatment solutions without Se or supplemented with 5 μmol/L of $Na_2SeO_3$ or $Na_2SeO_4$ for 15 days. Each treatment comprised four biological replicates. Mature leaves and roots were removed for transcriptome analyses, immediately frozen in liquid nitrogen, and stored at −80°C.

## 2.2 Detection of Se contents in roots and leaves

Roots were washed with Milli-Q water containing 2 mmol/L of MES and 1 mmol/L of $CaSO_4$, and the surface water was dried with absorbent paper. The roots and leaves were freeze-dried for 48 h in a lyophilizer (TF-FD-1; Zhejiang Nade Scientific Instrument Co., Ltd., Hangzhou, China). The total Se contents were determined as follows: 0.2-g samples (accurate to 0.0001 g) were weighed in a digestive tube, 4 mL of nitric acid and 2 mL of hydrogen peroxide were added, and the tube was sealed in a Mars 6 microwave digestion instrument (CEM Corp., Matthews, NC, USA). The digestion was conducted as follows: heated to 130°C for 10 min, kept for 5 min, brought to 200°C in 10 min, and held for 30 min. After the samples were digested and cooled to room temperature, they were diluted with ultrapure water to 50 mL and shaken. The total Se content of the sample was determined by NexIon 300 ICP-MS (PerkinElmer Inc., Waltham, MA, USA) with a radiofrequency (RF) power of 1,100 W; the flow rate of plasma gas (Ar) was 14 L/min, and the flow rate of reaction ($CH_4$) was 0.90 mL/min. The Rpq value was 0.8, and the atomization gas flow rate was 0.98 L/min.

## 2.3 RNA extraction, library construction, and sequencing

RNA was extracted from the roots and leaves of tea plants using an RNAprep Pure Polysaccharide Plant Total RNA Extraction Kit (Tiangen Biotech, Beijing, China). RNA integrity was verified by RNase-free agarose gel electrophoresis, and RNA purity was quantified using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, US). First, a sequencing library was generated using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, MA, USA), and sequencing was performed on the HiSeq platform (Illumina, San Diego, CA, USA) to obtain transcriptome data. Second, the raw data obtained by sequencing were filtered and checked for sequencing error rates and guanine–cytosine (GC) content distribution to obtain clean reads to ensure good-quality, reliable data analysis. FASTP software was used to filter the raw sequence. The filtered data were sufficiently clean for subsequent analysis. Sequencing was performed by Novogene Technology Co., Ltd. (Beijing, China). In this study, the whole genome of Longjing 43 was selected as the reference sequence, and the reference genome and annotation files were downloaded from the National Genomics Data Center (https://ngdc.cncb.ac.cn/search/?dbId=gwh&q=GWHACFB00000000). Finally, the HISAT2 software was used to compare clean reads with the reference genome quickly and accurately to obtain location information of the reads on the reference genome (Pertea et al., 2015).

## 2.4 Transcriptome data analysis

The DESeq2 package of R software was used to analyze differentially expressed genes (DEGs) between the $Na_2SeO_3$ vs. CK and $Na_2SeO_4$ vs. CK (Love et al., 2014). A negative binomial distribution was used to calculate the hypothesis test probability ($p$-value), and the obtained $p$-value was corrected using Benjamini and Hochberg's method of controlling the false discovery rate. In pairwise comparison, DEGs were clustered with $padj < 0.05$ and $|\log_2\text{FoldChange}| > 0.5$. The Novogene platform (https://magic.novogene.com/customer/main#/homeNew) was used to construct a Venn diagram. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed for all DEGs as previously described (Wang et al., 2019b).

## 2.5 Validation of RNA-Seq results by qRT-PCR

To verify the accuracy of the transcriptome data, 10 DEGs in the roots and leaves were selected for qRT-PCR, and gene-specific primers were designed using Oligo7. The primer sequences are listed in Supplementary Table 1. RNA was extracted from tea roots and leaves using the RNAprep Pure Polysaccharide Plant Total RNA Extraction Kit (Tiangen), followed by the PrimeScript RT Reagent Kit (Takara Bio, Kosatsu, Shiga, Japan) to reverse transcription of 1 μg of RNA into cDNA. A 10-fold dilution of cDNA was used for qRT-PCR of the target genes using a SYBR Green I Master kit (Roche, Basel, Switzerland). *CsPTB* was used as the reference gene (Hao et al., 2014).

## 2.6 Statistical analysis

All data were analyzed by SPSS Statistics v. 26 (IBM Corp., Armonk, NY, USA), and one-way ANOVA followed the least significant difference (LSD) test at $p \leq 0.05$. Column plots were

10.3389/fpls.2023.1268537

constructed using the GraphPad Prism 8 software (GraphPad Software, La Jolla, CA, USA).

# 3 Results

## 3.1 Effects of $Na_2SeO_3$ and $Na_2SeO_4$ on Se contents in tea cultivars

Total Se was measured in the roots and leaves of various HSe and LSe tea varieties treated with 5 μmol/L of $Na_2SeO_3$ or $Na_2SeO_4$ for 15 days (Figure 1). The total Se contents of tea plants in the roots and leaves were significantly increased by either the $Na_2SeO_3$ or $Na_2SeO_4$ treatment compared to those of the control plant treated without Se. Compared with the $Na_2SeO_4$ treatment, the $Na_2SeO_3$ treatment significantly increased the total Se in the roots and significantly decreased the total Se in the leaves. When treated with $Na_2SeO_3$ or $Na_2SeO_4$, the total Se in the roots of 'XC 4' was significantly higher than that in the roots of 'LSe', and the total Se in the leaves of either 'XC 4' or 'XC 1' was significantly higher than that in the leaves of 'LSe'. Among them, the total Se in the leaves of 'XC 1' was the highest when treated with $Na_2SeO_4$, reaching 17.53 mg/kg. When treated with $Na_2SeO_3$, the total Se in the leaves of 'XC 4' was the highest, reaching 1.56 mg/kg.

## 3.2 Identification and KEGG enrichment analysis of DEGs in response to $Na_2SeO_3$ and $Na_2SeO_4$

To elucidate the discrepancies in the molecular mechanisms of Se accumulation and transportation between the HSe and LSe tea cultivars, relative genes in untreated hydroponic 'LSe', 'XC 4', and 'XC 1' were compared to those subjected to 5 μmol/L of $Na_2SeO_3$ or $Na_2SeO_4$ for 15 days; the criteria for screening were $|\log_2 FoldChange| > 0.5$ and $padj < 0.05$ (Figure 2A). The number of genes responding to the $Na_2SeO_3$ treatment in the roots and leaves of HSe and LSe tea cultivars was higher than that responding to the $Na_2SeO_4$ treatment. When the plants were treated with $Na_2SeO_3$, there were 683, 1,964, and 1,506 upregulated genes in the roots and 1,258, 954, and 1,164 upregulated genes in the leaves in 'LSe', 'XC 4', and 'XC 1' cultivars, respectively. The number of downregulated genes in the roots was 1,168, 2,912, and 1,521 and in the leaves was 1,456, 714, and 970 in 'LSe', 'XC 4', and 'XC 1', respectively. When the plants were treated with $Na_2SeO_4$, the number of upregulated genes in 'LSe', 'XC 4', and 'XC 1' roots was 392, 391, and 427, whereas that in the leaves was 298, 424, and 599, respectively; the number of downregulated genes in the roots was 499, 720, and 430 and in the leaves was 862, 234, and 715 in 'LSe', 'XC 4', and 'XC 1', respectively. Notably, the number of DEGs after treatment with $Na_2SeO_4$ was lower than that after the $Na_2SeO_3$ treatment in all three cultivars, indicating higher sensitivities of tea plants to $Na_2SeO_3$ than to $Na_2SeO_4$.

A Venn diagram was used to analyze the DEGs of HSe and LSe tea cultivars under Se treatment. Under the $Na_2SeO_3$ treatment, the number of DEGs in the roots and leaves of the HSe tea cultivars was 942 and 385, respectively (Figures 2B, C), indicating that these genes responded to $Na_2SeO_3$ in the HSe tea cultivars. Under the $Na_2SeO_4$ treatment, the number of DEGs in the roots and leaves of the HSe tea cultivars was 100 and 118, respectively (Figures 2D, E), indicating that these genes responded to $Na_2SeO_4$ in the HSe tea cultivars. By comparing different tissues under the same treatment, the related tissue-specific DEGs responding to Se were determined. Analysis of the UpSet map showed that 748 and 288 genes were tissue-specific (Supplementary Figure 1A) and responded separately to $Na_2SeO_3$ in the roots and leaves of the HSe tea cultivars, respectively. Similarly, 83 and 105 genes (Supplementary Figure 1B) responded separately to $Na_2SeO_4$ in the roots and leaves of HSe tea cultivars, respectively, indicating tissue specificity.

KEGG enrichment analysis was performed on the DEGs in response to Se and the tissue-specific DEGs in the roots and leaves of HSe tea cultivars. Under the $Na_2SeO_3$ treatment, KEGG enrichment analysis was performed on 942 DEGs and 748 tissue-specific DEGs in the roots of the HSe tea cultivars. The DEGs were significantly enriched with the glutathione metabolism, zeatin biosynthesis, and phenylpropanoid biosynthesis pathways (Figure 3A). KEGG enrichment analysis was performed on 385 DEGs and 288 tissue-specific DEGs in the leaves of HSe tea

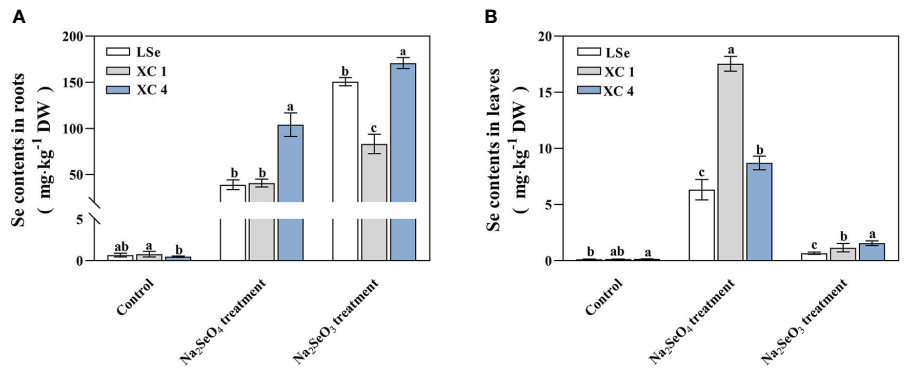



FIGURE 1
Total selenium contents in the roots and leaves of tea plants treated with $Na_2SeO_3$ and $Na_2SeO_4$. **(A)** Selenium contents in roots. **(B)** Selenium contents in leaves. Different lowercase letters above the bar indicate significant differences at the $p < 0.05$ level in the same treatment.

10.3389/fpls.2023.1268537

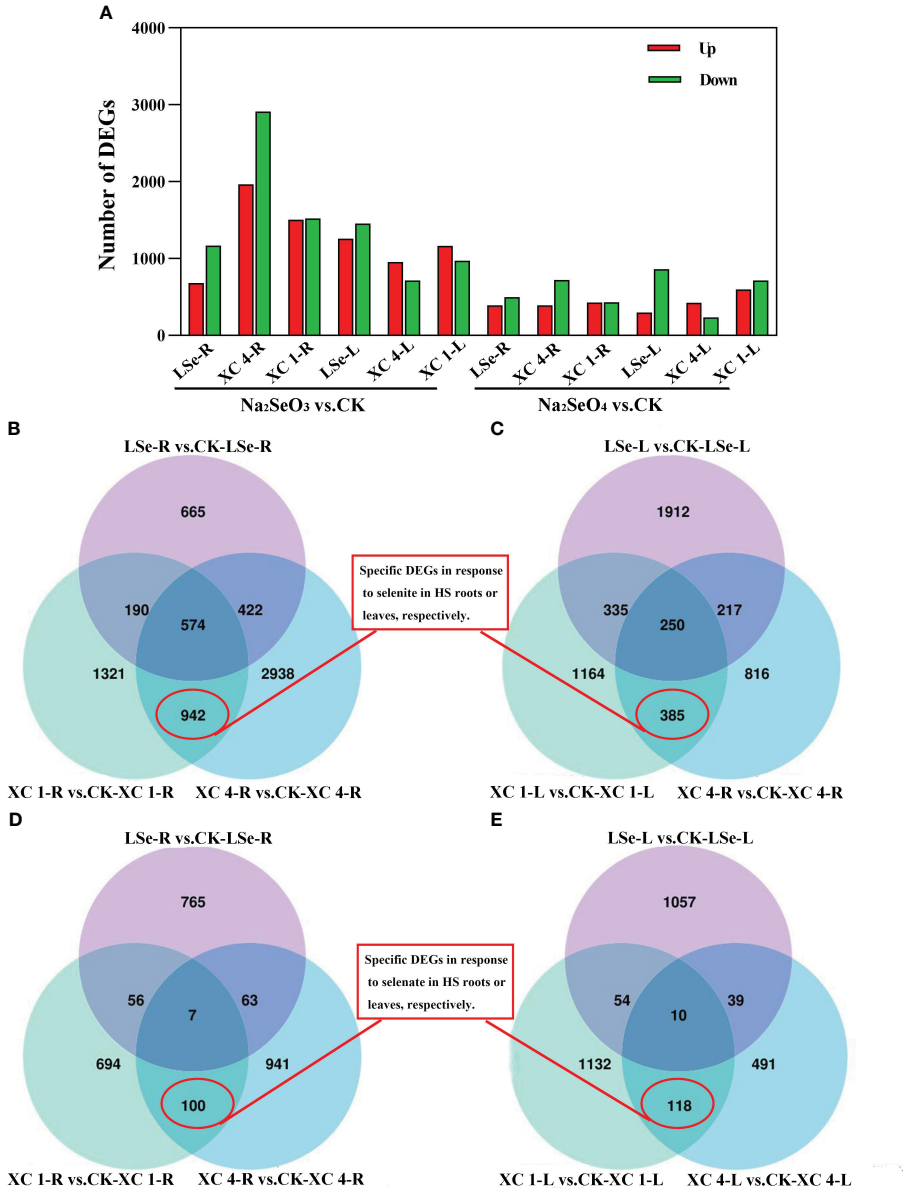



FIGURE 2
The numbers of DEGs and Venn diagram analyses between $Na_2SeO_3$ vs. control and $Na_2SeO_4$ vs. control in the three cultivars. **(A)** The numbers of upregulated and downregulated DEGs. **(B, C)** Venn diagram showing DEGs of $Na_2SeO_3$ treatment in the three cultivars roots and leaves. **(D, E)** Venn diagram showing DEGs of $Na_2SeO_4$ treatment in the three cultivars roots and leaves. 'XC 1-R' and 'XC 4-R' represent the root tissues of 'Zhongcha xicha 1 hao' and 'Zhongcha xicha 4 hao', respectively. 'XC 1-L' and 'XC 4-L' represent the leaf tissues of 'Zhongcha xicha 1 hao' and 'Zhongcha xicha 4 hao', respectively. 'LSe-R' and 'LSe-L' represent the root and leaf tissues of 'CT2009-0025', respectively. CK represents no selenium treatment. DEGs, differentially expressed genes.

cultivars, and the DEGs were significantly enriched with the flavonoid biosynthesis and ribosome pathways (Figure 3A). Similarly, KEGG enrichment analysis was performed of 100 DEGs and 83 tissue-specific DEGs, which were associated with $Na_2SeO_4$ in the roots of HSe tea cultivars, and it was found that they were significantly enriched with α-linolenic acid metabolism p (Figure 3B). The 118 DEGs and 105 tissue-specific DEGs in the leaves of the HSe tea cultivars were mainly enriched with flavonoid biosynthesis pathways (Figure 3B). Notably, the enriched pathways of DEGs in the roots of HSe tea cultivars in response to the $Na_2SeO_4$ and $Na_2SeO_3$ treatments were different, whereas the DEGs in the leaves of HSe tea cultivars were mainly enriched with the flavonoid biosynthesis pathways under both treatments.

## 3.3 DEGs involved in Se accumulation and transportation of the HSe tea cultivars

The transcriptome analysis identified 33 and 63 transporters in the leaves and roots, respectively (Supplementary Tables 2, 3). Among them, 22 and 11 transporters in the leaves, and 57 and six transporters in the roots responded to the $Na_2SeO_3$ and $Na_2SeO_4$

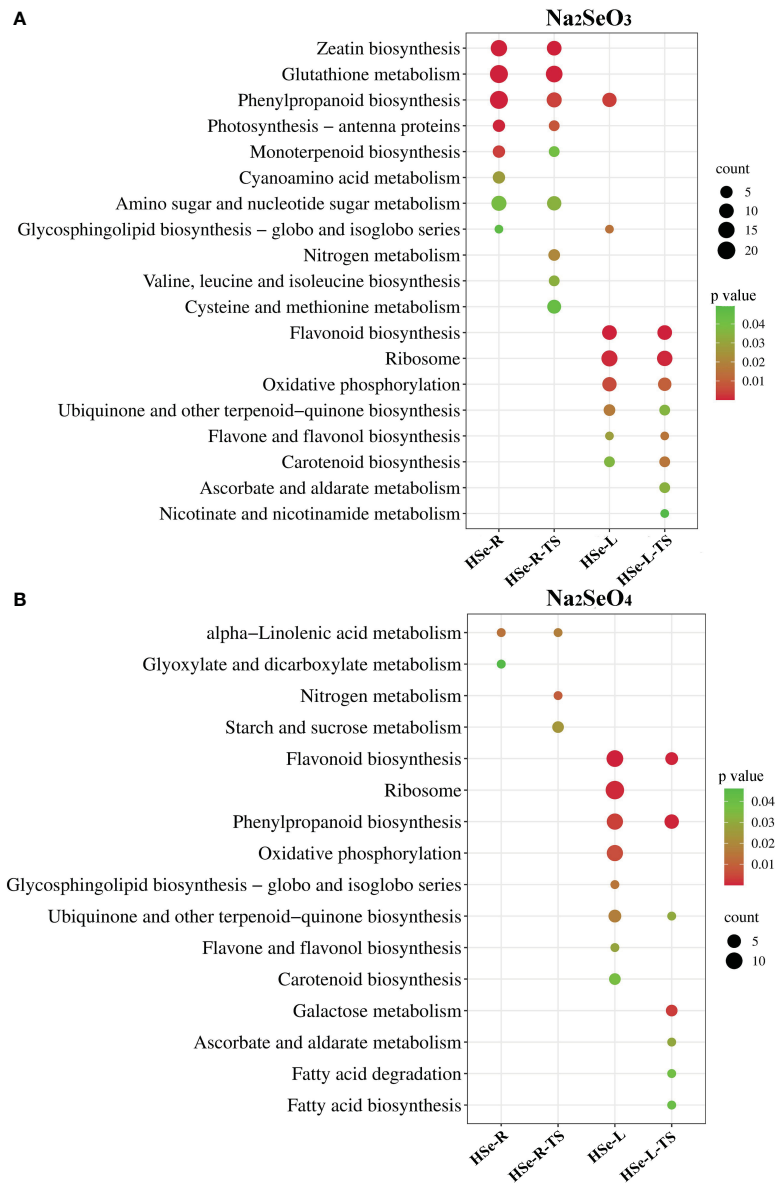



FIGURE 3
KEGG pathway analysis of DEGs in response to $Na_2SeO_3$ and $Na_2SeO_4$ treatment. **(A)** Treatment with $Na_2SeO_3$. **(B)** Treatment with $Na_2SeO_4$; HSe-R and HSe-L represent KEGG enrichment analyses of DEGs in the roots and leaves of HSe tea cultivars, respectively. HSe-R-TS and HSe-L-TS represent KEGG enrichment of tissue-specific DEGs of the HSe tea cultivars in roots and leaves, respectively. KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes.

treatments, respectively. Under the $Na_2SeO_4$ treatment, *ChaUn24016.1* (*PHT1;4*) in the leaves and *Cha09g005000* (*SULTR1;2*) and *Cha03g006400* (*SULTR3;4*) in the roots were identified in HSe tea cultivars. The roots of the HSe tea cultivars had more DEGs in response to the $Na_2SeO_3$ treatment than in response to the $Na_2SeO_4$ treatment. The transporters responding to $Na_2SeO_3$ mainly included ABC transporters, magnesium transporters, sugar transporters, phosphate transporters, oligopeptide transporters, and lysine histidine transporters (Figures 4, 5), which may be beneficial for exploring the key genes involved in $Na_2SeO_3$ accumulation and transportation in HSe tea cultivars. Seventeen ABC transporters were identified in the roots of HSe tea cultivars, of which nine were tissue-specific. There were 12 upregulated genes and five downregulated genes, of which *Cha01g013100* (*ABCG11*) was upregulated by 2.01- and 2.85-fold in 'XC 4' and 'XC 1', respectively. Among the two phosphate transporters, *PHT3;1* was upregulated by 2.92- and 1.99-fold in 'XC 4' and 'XC 1', respectively. *Cha12g006480* (*NRT1*) was upregulated by 1.18- and 1.85-fold in 'XC 4' and 'XC 1', respectively.

Meanwhile, the upregulated genes related to transferase, cytochrome P450, dehydrogenase, reductase, and ABC transporter genes, which were significantly differentially expressed in the HSe tea cultivars compared with their expression in the LSe tea cultivar, were filtered twofold in roots in response to $Na_2SeO_3$ (Table 1). Notably, nine glutathione transferases were involved in response to $Na_2SeO_3$, of which the expression of *ChaUn6696.4*

10.3389/fpls.2023.1268537

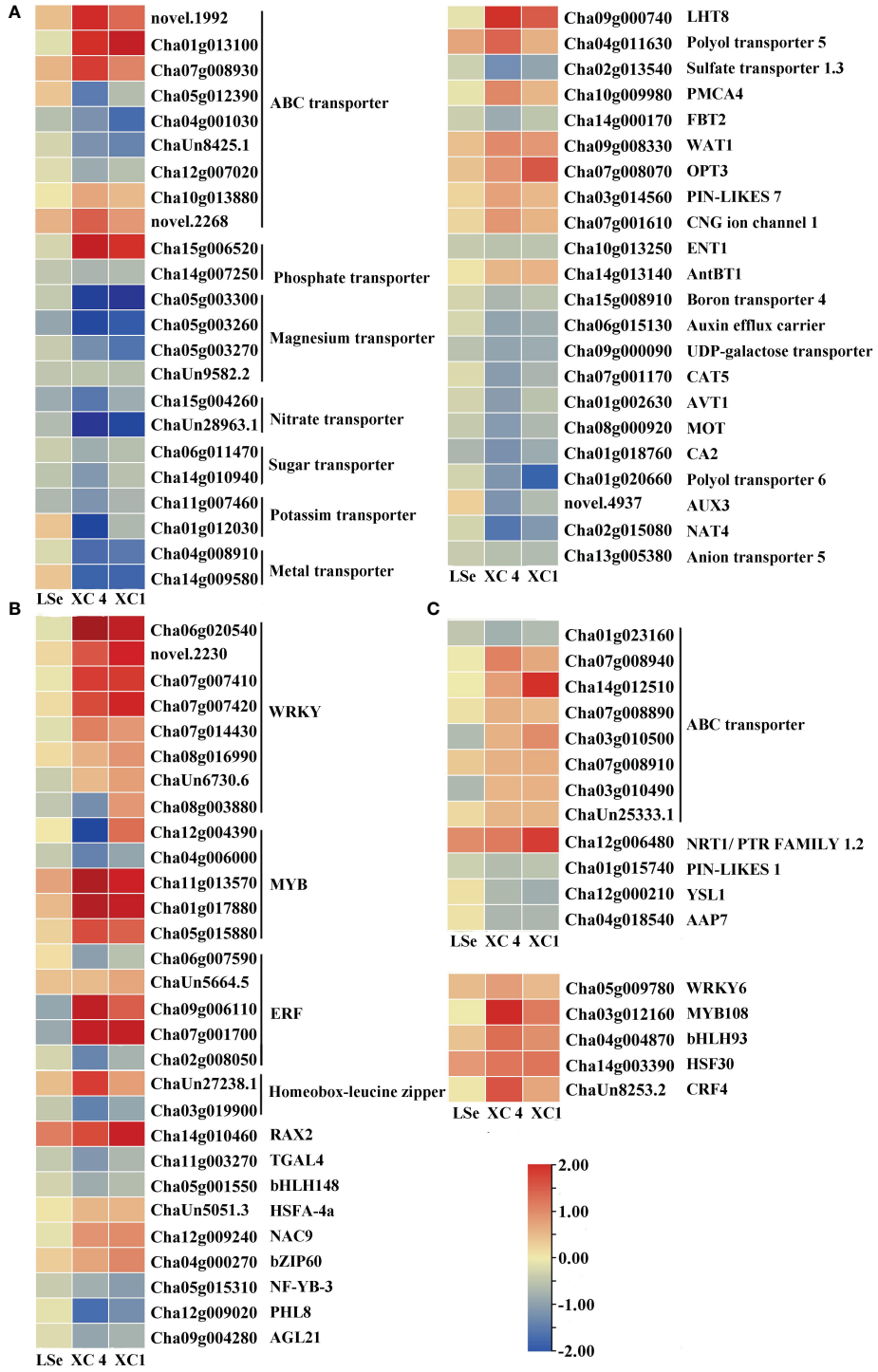



FIGURE 4
Heatmap of putative Se transporters and transcription factors identified as DEGs that responded to $Na_2SeO_3$ in the roots of the HSe tea cultivars. **(A, B)** The heatmap of tissue-specific DEGs related to transporters and transcription factors in the roots of HSe tea cultivars. **(C, D)** The heatmap of DEGs related to transporters and transcription factors in the roots of HSe tea cultivars. DEGs, differentially expressed genes; HSe, high selenium-enrichment ability.

(*GSTU17*) was 4.40- and 2.73-fold higher in 'XC 4' and 'XC 1', respectively. Moreover, nine genes were related to glucosyltransferases, of which *ChaUn5691.2* (crocetin-glucosyltransferase) expression was 4.49- and 2.69-fold higher in 'XC 4' and 'XC 1', respectively, than in LSe. Additionally, *Cha01g009120* (*ABCG8*) and *Cha13g008780* (*ABCC4*) were also identified. We also found that in two enzymes in the ethylene synthesis pathway, *Cha06g014280* (1-aminocyclopropane-1-carboxylate oxidase 1) expression was 2.83- and 3.06-fold higher in 'XC 4' and 'XC 1', respectively, than in LSe, and *Cha12g001220*

10.3389/fpls.2023.1268537

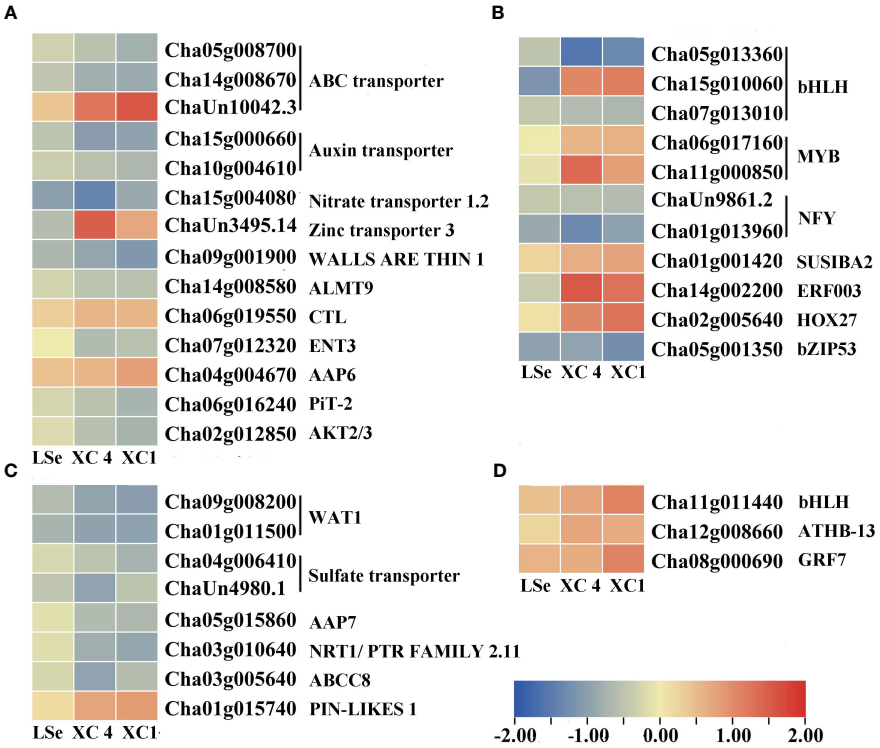



FIGURE 5
Heatmap of putative Se transporters and transcription factors identified as DEGs that responded to $Na_2SeO_3$ in the leaves of the HSe tea cultivars. **(A, B)** The heatmap of tissue-specific DEGs related to transporters and transcription factors in the leaves of HSe tea cultivars. **(C, D)** The heatmap of DEGs related to transporters and transcription factors in the leaves of HSe tea cultivars. DEGs, differentially expressed genes; HSe, high selenium-enrichment ability.

TABLE 1 The co-upregulated DEGs in the roots of HSe and LSe cultivars under $Na_2SeO_3$ treatment.

| Gene ID | Description | LSe $log_2FC$ (+$Se^{4+}$ vs -$Se^{4+}$) | XC 4 $log_2FC$ (+$Se^{4+}$ vs -$Se^{4+}$) | XC 1 $log_2FC$ (+Se4+ vs -Se4+) | XC 4/LSe $log_2FC$/ $log_2FC$ | XC 1/LSe $log_2FC$/ $log_2FC$ |
|---|---|---|---|---|---|---|
| Transferase | | | | | | |
| Cha05g011690 | Probable glutathione *S*-transferase | 0.84 | 1.89 | 1.73 | 2.25 | 2.06 |
| ChaUn6696.4 | Glutathione *S*-transferase U17 | 1.99 | 8.74 | 5.43 | 4.40 | 2.73 |
| Cha08g013620 | Probable glutathione *S*-transferase | 0.73 | 2.13 | 1.76 | 2.92 | 2.41 |
| Cha08g013640 | Probable glutathione *S*-transferase | 0.87 | 2.51 | 2.04 | 2.90 | 2.35 |
| Cha01g021070 | Glutathione *S*-transferase U10 | 1.74 | 4.28 | 3.46 | 2.46 | 1.99 |
| ChaUn15528.2 | Glutathione *S*-transferase U8 | 1.07 | 2.60 | 2.95 | 2.43 | 2.75 |
| Cha08g013610 | Probable glutathione *S*-transferase | 1.23 | 2.85 | 2.60 | 2.31 | 2.11 |
| Cha05g011780 | Probable glutathione *S*-transferase | 0.95 | 2.17 | 1.91 | 2.28 | 2.00 |
| Cha03g014180 | Microsomal glutathione *S*-transferase 3 | 0.53 | 1.35 | 0.89 | 2.55 | 1.69 |
| ChaUn5691.2 | Crocetin glucosyltransferase, chloroplastic | 0.54 | 2.43 | 1.46 | 4.49 | 2.69 |
| ChaUn5691.1 | Crocetin glucosyltransferase, chloroplastic | 1.24 | 5.31 | 3.90 | 4.28 | 3.15 |
| ChaUn9614.2 | Beta-D-glucosyl crocetin beta-1,6-glucosyltransferase | 1.33 | 2.93 | 2.48 | 2.21 | 1.87 |

*(Continued)*

10.3389/fpls.2023.1268537

TABLE 1 Continued

| Gene ID | Description | LSe $log_2$FC (+$Se^{4+}$ vs -$Se^{4+}$) | XC 4 $log_2$FC (+$Se^{4+}$ vs -$Se^{4+}$) | XC 1 $log_2$FC (+Se4+ vs -Se4+) | XC 4/LSe $log_2$FC/ $log_2$FC | XC 1/LSe $log_2$FC/ $log_2$FC |
|---|---|---|---|---|---|---|
| Cha01g002120 | Beta-D-glucosyl crocetin beta-1,6-glucosyltransferase | 0.93 | 2.36 | 2.44 | 2.52 | 2.62 |
| Cha13g010050 | Raucaffricine-*O*-beta-D-glucosidase | 0.55 | 1.25 | 0.75 | 2.26 | 1.36 |
| Cha01g022200 | Phenolic glucoside malonyltransferase 2 | 0.75 | 1.61 | 1.14 | 2.16 | 1.53 |
| ChaUn16290.1 | UDP-glycosyltransferase 73C6 | 0.72 | 2.41 | 2.17 | 3.33 | 3.00 |
| Cha06g010160 | UDP-glycosyltransferase 71K1 | 1.38 | 3.74 | 3.01 | 2.71 | 2.18 |
| Cha06g013540 | UDP-glycosyltransferase 73C3 | 0.65 | 1.54 | 0.99 | 2.35 | 1.51 |
| ChaUn17166.1 | BAHD acyltransferase | 1.21 | 3.18 | 2.27 | 2.63 | 1.88 |
| ChaUn12774.5 | Probable lipid transfer | 1.46 | 3.19 | 2.30 | 2.18 | 1.57 |
| Cha08g010320 | Tyrosine aminotransferase | 1.84 | 3.93 | 3.99 | 2.14 | 2.17 |
| Cha02g015780 | Scopoletin glucosyltransferase | 1.46 | 3.08 | 2.59 | 2.11 | 1.78 |
| Cytochrome P450 | | | | | | |
| Cha14g003030 | Cytochrome P450 CYP72A219 | 1.21 | 3.18 | 1.89 | 2.63 | 1.56 |
| ChaUn5945.1 | Cytochrome P450 94A2 | 0.74 | 2.79 | 2.96 | 3.75 | 3.98 |
| ChaUn13615.4 | Cytochrome P450 CYP72A219 | 0.74 | 3.10 | 0.97 | 4.22 | 1.31 |
| Cha04g001400 | Cytochrome P450 83B1 | 0.92 | 1.95 | 1.33 | 2.12 | 1.44 |
| Cha14g011380 | Cytochrome P450 89A9 | 0.95 | 2.26 | 1.54 | 2.38 | 1.62 |
| Dehydrogenase and reductase | | | | | | |
| ChaUn22024.2 | Short-chain type dehydrogenase/ reductase | 0.54 | 2.20 | 1.09 | 4.05 | 2.01 |
| Cha13g001990 | Probable mannitol dehydrogenase | 2.63 | 6.71 | 5.58 | 2.55 | 2.12 |
| Cha06g007800 | (+)-Neomenthol dehydrogenase | 1.09 | 2.77 | 2.54 | 2.53 | 2.32 |
| Cha14g007840 | (+)-Neomenthol dehydrogenase | 1.18 | 2.79 | 2.62 | 2.35 | 2.21 |
| Cha14g007830 | Short-chain dehydrogenase/reductase 2b | 2.44 | 6.52 | 5.50 | 2.68 | 2.26 |
| ChaUn20934.2 | Salutaridine reductase | 2.09 | 4.33 | 3.75 | 2.08 | 1.80 |
| ABC transporter | | | | | | |
| Cha01g009120 | ABC transporter G family member 8 | 1.43 | 2.87 | 2.04 | 2.00 | 1.42 |
| Cha13g008780 | ABC transporter C family member 4 | 1.37 | 2.94 | 1.97 | 2.14 | 1.44 |
| ACC oxidative synthase | | | | | | |
| Cha06g014280 | 1-Aminocyclopropane-1-carboxylate oxidase 1 | 1.05 | 2.96 | 3.19 | 2.83 | 3.06 |
| Cha12g001220 | 1-Aminocyclopropane-1-carboxylate synthase | 1.51 | 3.15 | 1.71 | 2.08 | 1.13 |
| Transcription factor | | | | | | |
| Cha11g001130 | Ethylene-responsive transcription factor ERF071 | 1.03 | 2.49 | 3.55 | 2.40 | 3.43 |

DEGs, differentially expressed genes; HSe, high selenium-enrichment ability; LSe, low selenium-enrichment ability.

(1-aminocyclopropane-1-carboxylate synthase) expression was 2.08- and 1.13-fold higher in 'XC 4' and 'XC 1', respectively, than in LSe. However, the co-upregulated genes in response to $Na_2SeO_3$ in the leaves were mainly heat shock proteins, auxin-responsive protein IAA7, and glycine-rich protein 2 (Table 2). In particular, heat shock proteins were obviously upregulated more in the HSe cultivars than in LSe by 2.18- and 1.94-fold for 'XC 4' and 'XC 1', respectively.

## 3.4 DEGs involved in Se regulatory network of the HSe tea cultivars

Five and four transcription factors responded to $Na_2SeO_4$ in the roots and leaves of the HSe tea cultivars, respectively (Supplementary Table 4). However, 34 and 14 transcription factors were significantly expressed in the roots and leaves of the HSe tea cultivars, respectively, in response to $Na_2SeO_3$ (Supplementary Table 5). Among them, *Cha07g001700* (ethylene-responsive transcription factor (ERF) 1B) was upregulated by 2.98- and 2.85-fold in the roots of 'XC 4' and 'XC 1', respectively; *Cha09g006110* (*ERF110*) was upregulated by 3.36- and 1.49-fold in the roots of 'XC 4' and 'XC 1', respectively; *Cha14g002200* (*ERF003*) was significantly upregulated in leaves. With the $Na_2SeO_3$ treatment, the *Cha11g013570* (*MYB44*), *Cha01g017880* (*MYB75*), and *Cha05g015880* (*MYB80*) transcription factors in roots were significantly upregulated, while the *Cha11g00085* (*MYB12*) and *Cha06g017160* (*MYB1R1*) transcription factors in the leaves were significantly upregulated. In addition, many *WRKY* transcription factors were identified in response to the $Na_2SeO_3$ treatment. For example, *Cha07g007420* (*WRKY6*), *Cha07g007410* (*WRKY42*), *novel.2230* (*WRKY51*), and *Cha06g020540* (*WRKY75*) were significantly upregulated in the roots of HSe tea cultivars. In particular, *WRKY75* was upregulated by 3.42- and 5.10-fold in 'XC 1' and 'XC 4', respectively. Only one *WRKY* transcription factor responded to $Na_2SeO_3$ in the leaves.

We also identified upregulated genes in three tea cultivars. Among them, *Cha11g001130* (*ERF071*) showed a higher induction degree in response to $Na_2SeO_3$ in the 'XC 4' and 'XC 1' than that in LSe by 2.40- and 3.43-fold, respectively.

## 3.5 DEGs involved in glutathione metabolism of the HSe tea cultivars

KEGG enrichment analysis of root tissue-specific DEGs after the $Na_2SeO_4$ treatment showed enrichment primarily of the glutathione metabolic pathway. Therefore, we analyzed DEG patterns related to glutathione metabolism (Figure 6A). Sixteen root tissue-specific DEGs were identified in the glutathione metabolic pathway of the HSe tea cultivars (Figure 6B). There were 13 DEGs in the glutathione *R*-transferase (2.5.1.18) pathway, and all of them were upregulated. Among them, *Cha09g013850* (glutathione *S*-transferase (*GST*) *U8*), *ChaUn26564.1* (*GSTU18*), and *Cha04g014010* (glutathione *S*-transferase part A) were upregulated more than twofold in the HSe tea cultivars (Supplementary Table 6). Hence, we speculated that *GST* genes enhanced the accumulation of Se in the roots of Se-enrichment tea plants and participated in the detoxification of Se.

## 3.6 Validation of DEGs by qRT-PCR

To verify the accuracy and reliability of the RNA-Seq data, we screened 10 genes related to transporters and transcription factors for qRT-PCR analysis in root and leaf tissues. The expression patterns demonstrated by RNA-Seq and qRT-PCR were consistent for all the tested 10 genes in the roots and leaves, and the results are shown in Figure 7. Therefore, the RNA-Seq data can be considered reliable.

# 4 Discussion

Based on their Se accumulation ability, plants can be divided into non-accumulator, Se-indicator, and Se-accumulator species (White, 2016). Tea plants have strong Se enrichment abilities, and various tea cultivars show significant differences in their ability to accumulate Se. Tea plants are more likely to accumulate $Na_2SeO_3$ than $Na_2SeO_4$ in root tissues and rarely transport it to leaf tissues. Based on previous reports, the Se contents in tea leaves reached Se-enriched levels after being treated with 5 μmol/L of Se concentrations for more than 14 days (Fang and Shen, 1992; Liu

TABLE 2 The co-upregulated DEGs in the leaves of HSe and LSe cultivars under $Na_2SeO_3$ treatment.

| Gene ID | Description | LSe $log_2FC$ ($+Se^{4+}$ vs $-Se^{4+}$) | XC 4 $log_2FC$ ($+Se^{4+}$ vs $-Se^{4+}$) | XC 1 $log_2FC$ ($+Se^{4+}$ vs $-Se^{4+}$) | XC 4/LSe $log_2FC$/ $log_2FC$ | XC 1/LSe $log_2FC$/ $log_2FC$ |
|---|---|---|---|---|---|---|
| Cha03g013630 | 15.7 kDa heat shock protein | 0.62 | 1.35 | 1.20 | 2.18 | 1.94 |
| ChaUn9222.6 | Thaumatin-like protein 1 | 0.74 | 1.37 | 1.25 | 1.84 | 1.68 |
| Cha12g002960 | Auxin-responsive protein IAA7 | 0.74 | 1.26 | 0.91 | 1.72 | 1.24 |
| Cha06g001000 | Glycine-rich protein 2 | 0.56 | 0.95 | 0.77 | 1.70 | 1.37 |
| novel.5594 | Wound-induced basic protein | 0.58 | 0.93 | 1.12 | 1.61 | 1.94 |
| ChaUn7167.6 | E3 ubiquitin protein ligase RIE1 | 0.51 | 1.04 | 0.80 | 2.02 | 1.57 |
| Cha06g006700 | Protein transport protein Sec61 subunit beta | 0.75 | 1.35 | 0.81 | 1.79 | 1.07 |
| Cha13g008460 | Phospholipase D zeta 2 | 0.57 | 0.99 | 0.84 | 1.75 | 1.48 |
| Cha01g017260 | Alpha/beta hydrolase family | 0.64 | 1.07 | 0.94 | 1.68 | 1.47 |

DEGs, differentially expressed genes; HSe, high selenium-enrichment ability; LSe, low selenium-enrichment ability.

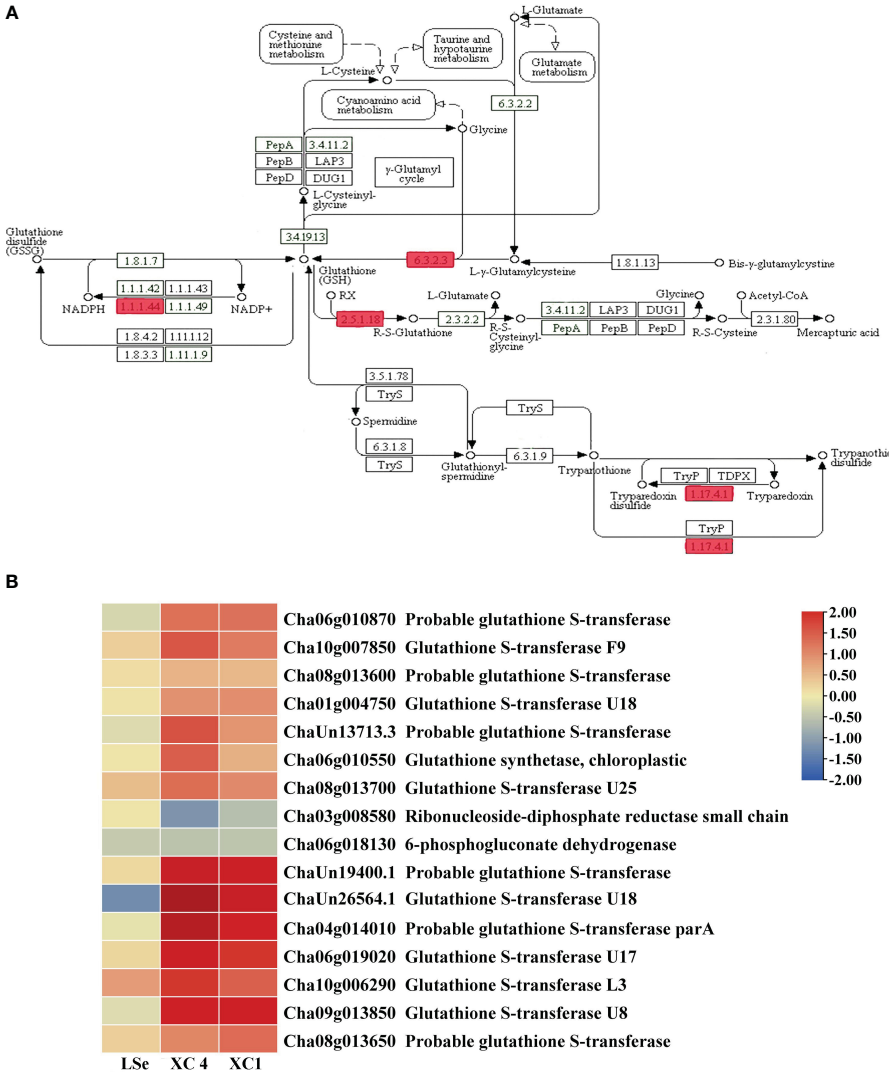



FIGURE 6
The DEGs in the roots of tissue-specific HSe tea cultivars are involved in the glutathione metabolic pathway. **(A)** Glutathione metabolic pathway. **(B)** Heatmap of differential genes. Red rectangles represent DEGs. 'XC 4' and 'XC 1' represent tissue-specific DEGs identified during analysis of $Na_2SeO_3$-treated vs. control samples in the roots of 'XC 4' and 'XC 1', respectively. Different colors indicate different levels of gene expression based on $log_2$FoldChange. DEGs, differentially expressed genes; HSe, high selenium-enrichment ability.

et al., 2021). In our study, under the $Na_2SeO_4$ or $Na_2SeO_3$ treatment for 15 days, 'XC 1' had the highest Se transportation ability from roots to leaves. The roots of 'XC 4' were readily able to accumulate Se but translocated only small amounts to the leaves. To explore the networks that regulate Se accumulation, transcriptome analyses were performed on various Se-enriched cultivars.

## 4.1 Transporters involved in Se accumulation and transportation in HSe tea cultivars

The RNA-Seq method has been used to analyze DEGs and has become a standard in multiple research fields (Costa-Silva et al., 2023). RNA-Seq technology can be used to identify genes related to Se transportation, accumulation, and assimilation (Hu et al., 2022). The transport mechanism of $Na_2SeO_3$ is more complex than that of $Na_2SeO_4$. $Na_2SeO_3$ has different ionization forms under various pH conditions. At pH 5, 5 μmol/L of $Na_2SeO_3$ mainly exists in the form of $HSeO_3^-$ (97.2%), $SeO_3^{2-}$ (2.4%), and $H_2SeO_3$ (0.4%) (Zhang et al., 2010). Plants accumulate $HSeO_3^-$ via phosphate transporters and $H_2SeO_3$ via silicon transporters (Zhao et al., 2010; Wang et al., 2019a). The accumulation of $SeO_3^{2-}$ is partly through ion channels, and most of the mechanisms are unclear. In the study, transcriptome analysis was performed to elucidate the accumulation and transportation mechanisms of $Na_2SeO_3$ and $Na_2SeO_4$ in various Se-enriched tea cultivars.

ABC transporters share a conserved ATPase domain and catalyze ATP to supply the energy required for the transmembrane transport of substrates, thus participating in important physiological processes, such as plant secondary metabolite accumulation and biotic and abiotic stress response

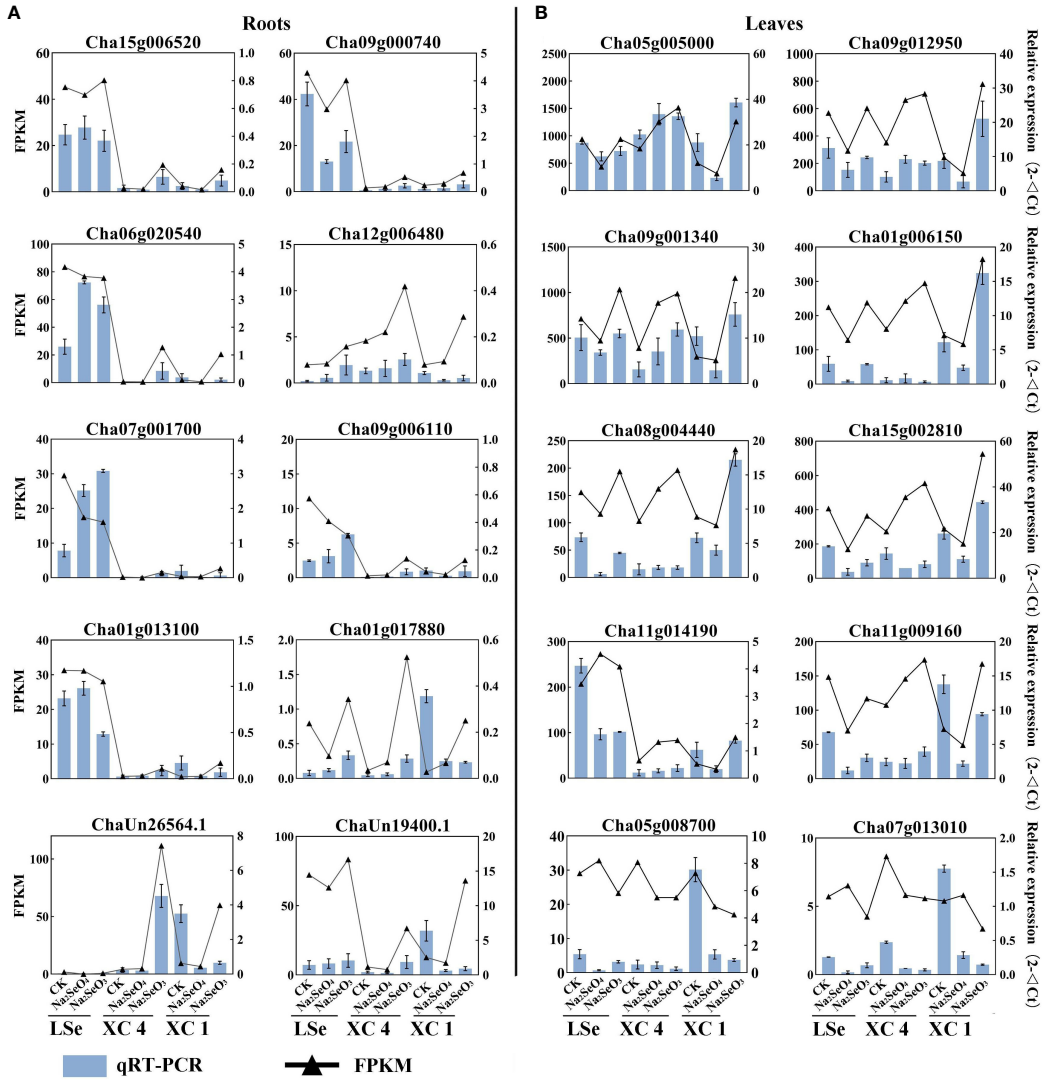



FIGURE 7
The expression of DEGs in roots **(A)** and leaves **(B)**. The black line represents expression levels as detected by RNA-Seq. The blue bars represent expression as detected by qRT-PCR. The error bar represents the standard deviation. DEGs, differentially expressed genes.

(Byrne et al., 2010). These transporters can be divided into eight subfamilies, from ABCA to ABCG, and ABCI (Verrier et al., 2008). In rice, ABC transporters were detected under the $Na_2SeO_3$ treatment, indicating that they might participate in Se accumulation and transportation (Kong et al., 2021). Meanwhile, the ABCB, ABCC, and ABCG transporters have been suggested to be involved in the accumulation and transport of Se in cowpeas (Li et al., 2023). Fourteen upregulated *ABC* genes were also found in our study (Figure 4), including *ABCA*, *ABCB*, *ABCC*, and *ABCG*, indicating that they may play vital roles in the accumulation and transportation of Se in tea roots.

Se mainly exists in organic forms, including selenoproteins, Se-polysaccharides, and Se-nucleotides (Zhou et al., 2020). However, there are few reports on the accumulation and transportation of Se-polysaccharides. In this study, sugar metabolites related to the abundance of glucosyltransferase genes in HSe cultivar roots were upregulated in response to $Na_2SeO_3$ (Table 1). Hence, we speculate that glucosyltransferase genes may participate in the formation and accumulation of Se-polysaccharides in tea plants.

In plants, sulfate transporters are involved in the accumulation and transportation of $Na_2SeO_4$, whereas phosphate transporters are involved in the accumulation of $Na_2SeO_3$ (Lazard et al., 2010; Tombuloglu et al., 2017). Cao et al. found that the response number and expression level of *CsPHT* gene in tea roots increased with increasing $Na_2SeO_3$ concentration (Cao et al., 2021). We found that only one SULTR1;2 and PHT3;1 might play important roles in $Na_2SeO_4$ and $Na_2SeO_3$ uptake, respectively, by HSe cultivars. Most $Na_2SeO_3$ accumulates in the root tissues after being converted into organic selenium, making its transport to the leaf tissues difficult (Ren et al., 2022). This finding is similar to that regarding the mode of accumulation of $Na_2SeO_3$ in rice. The NRT1.1B transporter in rice enhances the ability to transfer SeMet from roots to shoots (Zhang et al., 2019). In tea plants, we found that NRT1 may play an important role in the transportation and allocation of Se.

## 4.2 Transcription factors involved in Se accumulation and transportation in HSe tea cultivars

Reactive oxygen species (ROS) production in plants promotes an increase in the jasmonic acid and ethylene stress hormone levels (Overmyer et al., 2003). In *Arabidopsis*, jasmonic and ethylene respond to $Na_2SeO_3$, which induces *ERF* to participate indirectly in the regulation of plant defense (Lorenzo and Solano, 2005). Thirty-four and 14 $Na_2SeO_3$-regulated transcription factors were identified in the roots and leaves of the HSe tea cultivars, respectively. Among these, *ERF118* and *ERF110* were significantly upregulated in the roots of HSe cultivars, and *ERF003* was significantly upregulated in the leaves. Cao et al. found that jasmonic and ethylene could regulate a defensive network by upregulating the expression levels of transcriptional factors, including *ERF* and *MYB* in tea plant roots under the $Na_2SeO_3$ treatment (Cao et al., 2018). Our study suggested that similar regulatory mechanisms of *ERF*s and *MYB*s may exist in tea plants. *WRKY*s are involved in plant nutrition stress response. For example, *WRKY75* was significantly upregulated (Devaiah et al., 2007) and positively regulated phosphate and Se accumulation under P-deficient conditions. In this study, *WRKY75* was also found to be involved in regulating the accumulation of $Na_2SeO_3$ in tea plant roots.

## 4.3 Glutathione metabolism genes involved in Se accumulation and transportation in HSe tea cultivars

Accumulation of excessive Se can lead to oxidative responses in plants, resulting in excessive ROS that induce apoptosis. Glutathione can remove the excessive ROS produced by plants to protect tissues from oxidative damage and improve plant tolerance to Se (Zou et al., 2021). After treatment with high $Na_2SeO_3$ concentrations, 15 genes involved in glutathione metabolism were identified in tea plants. Genes encoding glutathione *S*-transferase, glutathione synthetase, glutathione peroxidase, and glutathione reductase were significantly upregulated, suggesting that they may increase the tolerance and accumulation of Se in tea plants (Cao et al., 2018). In the study, 13 *GST* genes were upregulated in the roots of HSe tea cultivars treated with $Na_2SeO_3$. Furthermore, the nine co-upregulated *GST*s showed significantly higher expression in HSe cultivar roots, as compared to the LSe cultivar. GSTs might be induced to transfer $SeO_4^{2-}$ ions to GSH to form glutathione *S*-conjugates (GS-X) under the $Na_2SeO_3$ treatment, which can explain the accumulation of abundant Se in roots (Zhou et al., 2018). Therefore, we speculate that *GST*s are likely key genes for accumulating Se in tea plant roots.

# 5 Conclusion

In this study, the effects of different Se sources on the Se contents of the roots and leaves of various high and low Se-enrichment ability tea cultivars were analyzed. Under treatment, the Se contents of roots were obviously higher than those of leaves, especially after treatment with $Na_2SeO_3$, and the Se contents of HSe cultivars in leaves were significantly higher than those in LSe leaves. The RNA-Seq analysis showed that the number of DEGs after the $Na_2SeO_3$ treatment was higher than that after the $Na_2SeO_4$ treatment, indicating that the tea plant responded more strongly to $Na_2SeO_3$. Further studies suggested that *GST*s might participate in the accumulation of abundant Se in roots. The accumulation of $Na_2SeO_3$ was increased in line with the upregulated expression levels of *ABC*, glucosyltransferases, *NRT1*, and *PHT3;1* in tea plants (Figure 8). Meanwhile, the expression levels of *SULTR1;2* and *SULTR3;4* genes in HSe tea cultivars were obviously induced by $Na_2SeO_4$, which participates in the accumulation of Se. These data provide a basis for the mechanism of Se accumulation and transportation in various high and low Se-enrichment ability tea cultivars.

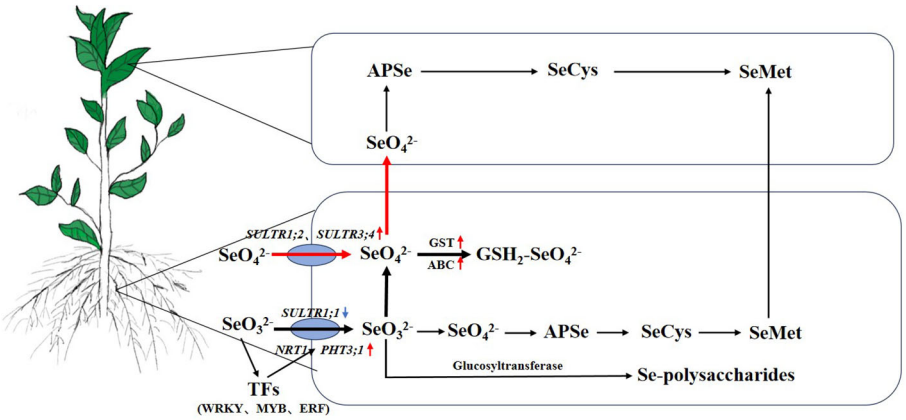



FIGURE 8
The uptake, transport, and metabolism of Se in the HSe tea cultivars. The red arrow represents upregulated genes. The blue arrow represents downregulated genes. The black bold lines represent the main metabolic pathways of $Na_2SeO_3$. The red bold lines represent the main metabolic pathways of $Na_2SeO_4$. HSe, high selenium-enrichment ability.

## 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 authors. The datasets presented in this study can be found in online repositories. The raw RNA - 438 Seq generated are available in the NCBI Sequence Read Archive (SRA) database under 439 accession number SRR25823839-SRR25823909 and SRR25823916-SRR25823917.

## Author contributions

QZ: Writing – original draft. LG: Writing – original draft. JH: Writing – review & editing. XH: Software, Supervision, Writing – review & editing. XL: Methodology, Writing – review & editing. NL: Writing – review & editing. YW: Methodology, Software, Writing – review & editing. KZ: Writing – review & editing. XW: Project administration, Writing – review & editing. LW: Methodology, Supervision, Writing – review & editing. JZ: Writing – review & editing, Funding acquisition, Project administration, Supervision.

## Funding

This work was supported by theAgricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (No. CAAS-XTCX20190025-7), and the Agricultural Sciences and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (ASTIP).

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1268537/full#supplementary-material

## References

Barberon, M., Berthomieu, P., Clairotte, M., Shibagaki, N., Davidian, J. C., and Gosti, F. (2008). Unequal functional redundancy between the two *Arabidopsis thaliana* high-affinity sulphate transporters SULTR1;1 and SULTR1;2. *New Phytol.* 180, 608–619. doi: 10.1111/j.1469-8137.2008.02604.x

Byrne, S. L., Durandeau, K., Nagy, I., and Barth, S. (2010). Identification of ABC transporters from *Lolium perenne* L. that are regulated by toxic levels of selenium. *Planta* 231, 901–911. doi: 10.1007/s00425-009-1096-y

Cao, D., Liu, Y., Ma, L., Jin, X., Guo, G., Tan, R., et al. (2018). Transcriptome analysis of differentially expressed genes involved in selenium accumulation in tea plant (*Camellia sinensis*). *PLoS One* 13, e0197506. doi: 10.1371/journal.pone.0197506

Cao, D., Liu, Y., Ma, L., Liu, Z. H., Li, J., Wen, B., et al. (2021). Genome-wide identification and characterization of phosphate transporter gene family members in tea plants (*Camellia sinensis* (L.). O. kuntze) under different selenite levels. *Plant Physiol. Biochem.* 166, 668–676. doi: 10.1016/j.plaphy.2021.06.038

Cabannes, E., Buchner, P., Broadley, M. R., Hawkesford, M. J., et al. (2011) A comparison of sulfate and selenium accumulation in relation to the expression of sulfate transporter genes in *Astragalusspecies. Plant Physiol.* 157, 2227–2239. doi: 10.1104/pp.111.183897

Chao, W., Rao, S., Chen, Q., Zhang, W., Liao, Y., Ye, J., et al. (2022). Advances in research on the involvement of selenium in regulating plant ecosystems. *Plants-Basel* 11, 2712. doi: 10.3390/plants11202712

Chen, Y., Deng, Y., Wu, X., Zhang, D., Wang, F., Liu, K., et al. (2022). The levels of selenium in tea from China and associated human exposure. *J. Food Compos. Anal.* 110, 104567. doi: 10.1016/j.jfca.2022.104567

Costa-Silva, J., Hungria, M., Domingues, D. S., Menotti, D., and Lopes, F. M. (2023). Temporal progress of gene expression analysis with RNA-Seq data: A review on the relationship between computational methods. *Comput. Struct. Biotec.* 21, 86–98. doi: 10.1016/j.csbj.2022.11.0512001-0370

Devaiah, B. N., Karthikeyan, A. S., and Raghothama, K. G. (2007). *WRKY75* transcription factor is a modulator of phosphate acquisition and root development in arabidopsis. *Plant Physiol.* 143, 1789–1801. doi: 10.1104/pp.106.093971

El Mehdawi, A. F., Jiang, Y., Guignardi, Z. S., Esmat, A., Pilon, M., Pilon-Smits, E. A. H., et al. (2018). Influence of sulfate supply on selenium uptake dynamics and expression of sulfate/selenate transporters in selenium hyperaccumulator and nonhyperaccumulator *Brassicaceae. New Phytol.* 217, 194–205. doi: 10.1111/nph.14838

Elrashidi, M. A., Adriano, D. C., Workman, M., and Lindsay, W. L. (1987). Chemical equilibria of selenium in soils: a theoretical development. *Soil Sci.* 144, 141–152. doi: 10.1097/00010694-198708000-00008

Fang, X., and Shen, X. (1992). Effects of selenium on growth and substance metabolism of the tea plant. *China Tea.* 02, 28–30.

Hao, X., Horvath, D. P., Chao, W., Yang, Y., Wang, X., and Xiao, B. (2014). Identification and evaluation of reliable reference genes for quantitative real-time PCR analysis in tea plant (*Camellia sinensis* (L.) O. Kuntze). *Int. J. Mol. Sci.* 15, 22155–22172. doi: 10.3390/ijms151222155

Hopper, J. L., and Parker, D. R. (1999). Plant availability of selenite and selenate as influenced by the competing ions phosphate and sulfate. *Plant Soil.* 210, 199–207. doi: 10.1023/a:1004639906245

Hu, H., Hu, J., Wang, Q., Xiang, M., and Zhang, Y. (2022). Transcriptome analysis revealed accumulation-assimilation of selenium and physio-biochemical changes in alfalfa (*Medicago sativa* L.) leaves. *J. Sci. Food Agr.* 102, 4577–4588. doi: 10.1002/jsfa.11816

Huang, C., Liu, Y., Pan, L., Liao, Q., Shen, P., Huang, C., et al. (2020). Research and prospect analysis of microorganisms in selenium-rich soil. *Plant Dis.* 11, 36-38, 46.

Kong, Q., Li, F., Qin, L., and Chen, E. (2021). Screening and analysis of Se responsive genes in leaves of foxtail millet. *Mol. Plant Breed.* 19, 2798–2810. doi: 10.13271/j.mpb.019.002798

Lazard, M., Blanquet, S., Fisicaro, P., Labarraque, G., and Plateau, P. (2010). Uptake of selenite by *Saccharomyces Cerevisiae* involves the high and low affinity orthophosphate transporters. *J. Biol. Chem.* 285, 32029–32037. doi: 10.1074/jbc.M110.139865

Li, H., McGrath, S. P., and Zhao, F. (2008). Selenium uptake, translocation and speciation in wheat supplied with selenate or selenite. *New Phytol.* 178, 92–102. doi: 10.1111/j.1469-8137.2007.02343.x

Li, L., Xiong, Y., Wang, Y., Wu, S., Xiao, C., Wang, S., et al. (2023). Effect of nano-selenium on nutritional quality of cowpea and response of ABCC transporter family. *Molecules* 28, 1398. doi: 10.3390/molecules28031398

Liu, K., Hu, X., Ding, S., Wan, B., Zhou, X., Ji, X., et al. (2021). Research on the selenium accumulation ability of different tea varieties. *J. Tea Commun.* 48, 638–643.

Lorenzo, O., and Solano, R. (2005). Molecular players regulating the jasmonate signalling network. *Curr. Opin. Plant Biol.* 8, 532–540. doi: 10.1016/j.pbi.2005.07.003

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome. Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Overmyer, K., Brosche, M., and Kangasjarvi, J. (2003). Reactive oxygen species and hormonal control of cell death. *Trends Plant Sci.* 8, 335–342. doi: 10.1016/s1360-1385(03)00135-3

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, 290. doi: 10.1038/nbt.3122

Raina, M., Sharma, A., Nazir, M., Kumari, P., Rustagi, A., Hami, A., et al. (2021). Exploring the new dimensions of selenium research to understand the underlying mechanism of its uptake, translocation, and accumulation. *Physiol. Plant* 171, 882–895. doi: 10.1111/ppl.13275

Ren, H., Li, X., Guo, L., Wang, L., Hao, X., and Zeng, J. (2022). Integrative transcriptome and proteome analysis reveals the absorption and metabolism of selenium in tea plants (*Camellia sinensis* (L.) O. Kuntze). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.848349

Shrift, A., and Ulrich, J. M. (1969). Transport of selenate and selenite into *Astragalus* roots. *Plant Physiol.* 44, 893–896. doi: 10.1104/pp.44.6.893

Tang, H., Tang, J., Li, J., and Zhao, C. (2012). Effect of selenium-rich,yields and quality by applying selenium fertilizer on Yinghong 9 tea trees. *Guangdong Agr. Sci.* 39, 52–54.

Tombuloglu, H., Filiz, E., Aydin, M., and Koc, I. (2017). Genome-wide identification and expression analysis of sulphate transporter (*SULTR*) genes under sulfur deficiency in *Brachypodium* distachyon. *J. Plant Biochem. Biotechnol.* 26, 263–273. doi: 10.1007/s13562-016-0388-0

Verrier, P. J., Bird, D., Buria, B., Dassa, E., Forestier, C., Geisler, M., et al. (2008). Plant ABC proteins - a unified nomenclature and updated inventory. *Trends Plant Sci.* 13, 151–159. doi: 10.1016/j.tplants.2008.02.001

Wang, Z., Huang, W., and Pang, F. (2022a). Selenium in soil-plant-microbe: a review. *B. Environ. Contam. Tox.* 108, 167–181. doi: 10.1007/s00128-021-03386-2

Wang, M., Yang, W., Zhou, F., Du, Z., Xue, M., Chen, T., et al. (2019a). Effect of phosphate and silicate on selenite uptake and phloem-mediated transport in tomato (*Solanum lycopersicum* L.). *Environ. Sci. pollut. Res.* 26, 20475–20484. doi: 10.1007/s11356-019-04717-x

Wang, L., Yao, L., Hao, X., Li, N., Wang, Y., Ding, C., et al. (2019b). Transcriptional and physiological analyses reveal the association of ROS metabolism with cold tolerance in tea plant. *Environ. Exp. Bot.* 160, 45–58.

Wang, Z., Yuan, L., Li, J., and Hu, Z. (2022b). Research progress of selenium-rich plant protein. *J. Chin. Cereals Oils Assoc.* 37, 196–202.

White, P. J. (2016). Selenium accumulation by plants. *Ann. Bot.* 117, 217–235. doi: 10.1093/aob/mcv180

Xiang, J., Rao, S., Chen, Q., Zhang, W., Cheng, S., Cong, X., et al. (2022). Research progress on the effects of selenium on the growth and quality of tea plants. *Plants-Basel* 11, 2491. doi: 10.3390/plants11192491

Xiong, J., Li, W., Pan, W., Wei, M., Tang, X., Yan, H., et al. (2019). Effects of exogenous selenium application on selenium content and yield of purple sweet potato under different soil conditions. *J. South. Agr.* 50, 1211–1218.

Ye, F., Gong, Z., Gao, S., Zhang, Q., Cui, Q., and Liang, J. (2015). Investigation of the selenium element in tea plantation of Enshi district, Hubei province. *J. Sichuan Agr. Univ.* 33, 275–278.

Yoshimoto, N., Takahashi, H., Smith, F. W., Yamaya, T., and Saito, K. (2002). Two distinct high-affinity sulfate transporters with different inducibilities mediate uptake of sulfate in *Arabidopsis* roots. *Plant J.* 29, 465–473. doi: 10.1046/j.0960-7412.2001.01231.x

Yu, Y., Liu, Z., Luo, L. Y., Fu, P., Wang, Q., and Li, H. (2019). Selenium uptake and biotransformation in *Brassica rapa* supplied with selenite and selenate: a hydroponic work with HPLC speciation and RNA-sequencing. *J. Agr. Food Chem.* 67, 12408–12418. doi: 10.1021/acs.jafc.9b05359

Zhang, H., Hao, X., Zhang, J., Wang, L., Wang, Y., Li, N., et al. (2022). Genome-wide identification of *SULTR* genes in tea plant and analysis of their expression in response to sulfur and selenium. *Protoplasma* 259, 127–140. doi: 10.1007/s00709-021-01643-z

Zhang, L., Hu, B., Deng, K., Gao, X., Sun, G., Zhang, Z., et al. (2019). *NRT1.1B* improves selenium concentrations in rice grains by facilitating selenomethinone translocation. *Plant Biotechnol. J.* 17, 1058–1068. doi: 10.1111/pbi.13037

Zhang, L., Hu, B., Li, W., Che, R., Deng, K., Li, H., et al. (2014). OsPT2, a phosphate transporter, is involved in the active uptake of selenite in rice. *New Phytol.* 201, 1183–1191. doi: 10.1111/nph.12596

Zhang, Y., Xia, W., and Ren, A. (2013). A Research on synthesis of microorganism template nano-selenium material. *Comput. Aided Des.* 443, 639 doi: 10.4028/www.scientific.net/AMM.443.639

Zhang, L., Yu, F., Shi, W., Li, Y., and Miao, Y. (2010). Physiological characteristics of selenite uptake by maize roots in response to different pH levels. *J. Plant Nutr. Soil Sci.* 173, 417–422. doi: 10.1002/jpln.200900260

Zhao, X., He, L., Li, Y., Wei, H., Fang, S., Lu, S., et al. (2019). Effect of selenium-containing fertilizer on ability of selenium accumulation in peanut. *Southwest China J. Agr. Sci.* 32, 2350–2354. doi: 10.16213/j.cnki.scjas.2019.10.015

Zhao, X., Mitani, N., Yamaji, N., Shen, R., and Ma, J. (2010). Involvement of silicon influx transporter OsNIP2;1 in selenite uptake in rice. *Plant Physiol.* 153, 1871–1877. doi: 10.1104/pp.110.157867

Zhou, C., Hu, Y., Zeng, J., Yang, J., and Chen, L. (2015). Effects of soil factors on the selenium absorption characteristics of tea plant. *J. Tea Sci.* 35, 429–436.

Zhou, N., Long, H., Wang, C., Yu, L., Zhao, M., and Liu, X. (2020). Research progress on the biological activities of selenium polysaccharides. *Food Funct.* 11, 4834–4852. doi: 10.1039/c9fo02026h

Zhou, Y., Tang, Q., Wu, M., Mou, D., Liu, H., Wang, S., et al. (2018). Comparative transcriptomics provides novel insights into the mechanisms of selenium tolerance in the hyperaccumulator plant *Cardamine hupingshanensis*. *Sci. Rep.* 8, 2789. doi: 10.1038/s41598-018-21268-2

Zou, Y., Han, C., Wang, F., Tan, Y., Yang, S., Huang, C., et al. (2021). Integrated metabolome and transcriptome analysis reveal complex molecular mechanisms underlying selenium response of *Aloe vera* L. *J Plant Biol* 64(17):135–143.. doi: 10.1007/s12374-020-09285-z

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 October 2023
DOI 10.3389/fpls.2023.1277436

Check for updates

OPEN ACCESS

EDITED BY
Eleni Tani,
Agricultural University of Athens, Greece

REVIEWED BY
Xiaomei Zhang,
Chinese Academy of Agricultural Sciences, China
Hulya Ilbi,
Ege University, Türkiye

*CORRESPONDENCE
Hongna Zhang
13692476979@139.com
Yongzan Wei
wyz4626@163.com

RECEIVED 14 August 2023
ACCEPTED 11 October 2023
PUBLISHED 26 October 2023

CITATION
Zhang X, Ouyang Y, Zhao L, Li Z, Zhang H and Wei Y (2023) Genome-wide identification of PEBP gene family in pineapple reveal its potential functions in flowering.
*Front. Plant Sci.* 14:1277436.
doi: 10.3389/fpls.2023.1277436

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

# Genome-wide identification of PEBP gene family in pineapple reveal its potential functions in flowering

Xiaohan Zhang[1], Yanwei Ouyang[1], Lei Zhao[1,2], Ziqiong Li[1], Hongna Zhang[1]* and Yongzan Wei[2]*

[1]School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan University, Sanya, China, [2]Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Hainan Institute for Tropical Agricultural Resources, Haikou, China

Phosphatidylethanolamine binding protein (PEBP) plays an important role in regulating flowering time and morphogenesis of plants. However, the identification and functional analysis of *PEBP* gene in pineapple (*AcPEBP*) have not been systematically studied. The pineapple genome contained 11 PEBP family members, which were subsequently classified into three subfamilies (FT-like, TFL-like and MFT-like) based on phylogenetic relationships. The arrangement of these 11 shows an unequal pattern across the six chromosomes of pineapple the pineapple genome. The anticipated outcomes of the promoter cis-acting elements indicate that the *PEBP* gene is subject to regulation by diverse light signals and endogenous hormones such as ethylene. The findings from transcriptome examination and quantitative real-time polymerase chain reaction (qRT-PCR) indicate that FT-like members *AcFT3* and *AcFT4* display a heightened expression level, specifically within the floral structures. The expression of *AcFT3* and *AcFT4* increases sharply and remains at a high level after 4 days of ethylene induction, while the expression of *AcFT7* and *AcMFT1* decreases gradually during the flowering process. Additionally, *AcFT3, AcFT4* and *AcFT7* show specific expression in different floral organs of pineapple. These outcomes imply that members belonging to the FT-like subfamily may have a significant impact on the process of bud differentiation and flower development. Through transcriptional activation analysis, it was determined that *AcFT4* possesses transcriptional activation capability and is situated in the nucleus and peripheral cytoplasm. Overexpression of *AcFT4* in *Arabidopsis* resulted in the promotion of early flowering by 6-7 days. The protein interaction prediction network identified potential flower regulators, including CO, AP1, LFY and SOC1, that may interact with PEBP proteins. This study explores flower development in pineapple, thereby serving as a valuable reference for future research endeavors in this domain.

KEYWORDS

**pineapple (*Ananas comosus* (L.) Merr.), PEBP, genome-wide, expression profiles, flowering**

## Introduction

Phosphatidylethanolamine-binding proteins (PEBP) have been found to have significant involvement in flower induction, regulation of flowering time, and seed germination in plants (Zhang et al., 2020; Jin et al., 2021). PEBP has been categorized into three subfamilies, namely FLOWERING LOCUS T-like (FT-like), TERMINAL FLOWER 1 (TFL1-like) and MOTHER OF FT and TFL1 (MFT-like). Within the FT-Like subfamily, there are two members, namely Flowering Locus T (FT) and TWIN SISTER OF FT (TSF). FT serves as an integral factor in the flowering pathway and plays a crucial role in the regulation of the flowering process (Adeyemo et al., 2019). Under conditions of extended daylight, the expression of FT is induced, resulting in the formation of the FT-FD complex with FD. This complex then activates the expression of *APETALA1* (*AP1*) and *LEAFY* (*LFY*) genes, thereby facilitating the transition to the flowering stage (Abe et al., 2005; Wigge et al., 2005). In rice, *Hd3a* and *RFT1* play a role in activating the expression of *AP1* homologs, namely *OsMADS14* and *OsMADS15*, thereby promoting flowering (Komiya et al., 2008). Furthermore, *FT* homologs have been extensively identified and studied in various other plant species, such as *Lilium* (Yan et al., 2021), *Cucumis melo L* (Zhang et al., 2020; Zhang and Zhang. 2020), sugarcane (Venail et al., 2022), *Perilla frutescens* (Xu et al., 2022) and *Dendrobium huoshanense* (Song et al., 2021). There is an increasing body of evidence indicating the significant involvement of *FT* genes in the flowering mechanism of plants. In the case of grapes, the expression of *VvFT* serves as an indicator that the developmental transition from juvenile to adult stage has occurred (Carmona et al., 2007). In apples, *MdFT1* and *MdFT2* exhibit distinct expression patterns, yet both possess the capability to initiate early flowering in apple plants (Kotoda et al., 2010). In cassava, the *Arabidopsis FT* gene accelerates flower bud formation and increases lateral branching, and its overexpression also significantly increases flower yield and prolongs flower development life (Adeyemo et al., 2017). Furthermore, the translocation expression of *CsFTL3* in chrysanthemum has been found to stimulate early flowering in both *Arabidopsis* and *chrysanthemum* plants (Oda et al., 2012). These findings suggest that the role of the *FT* gene in flower induction is conserved across various plant species.

MFT, acknowledged as the most ancient constituent of the PEBP family, governs the process of seed germination by means of the abscisic acid (ABA) and gibberellin (GA) signaling pathways in *Arabidopsis* (Xi et al., 2010). The TFL1-like subfamily encompasses TERMINAL FLOWER 1 (TFL1), BROTHER OF FT AND TFL1 (BFT), and CENTRORADIALIS (CEN). Both the TFL1-like and FT-like subfamilies participate in the management of flowering. Nevertheless, in comparison to FT, TFL1 exhibits inhibitory characteristics towards the flowering process (Zhu et al., 2020; Song et al., 2021). In cucumber, the antagonistic interaction between *CsTFL1* and *CsFT*, along with the competitive binding with the CsNOT2a-CsFDP complex, hinders cucumber growth and the development of terminal flowers (Wen et al., 2019). Similarly, in rice, the TFL1-like protein RCN inhibits flower development by competitively binding 14-3-3 and FD proteins to Hd3a/RFT1 (Kaneko-Suzuki et al., 2018). Additionally, the overexpression of *MiTFL1-1* and *MiTFL1-2* genes in mango has been found to significantly impede the flowering in *Arabidopsis* (Gafni et al., 2022).

The pineapple (*Ananas comosus* (L.) Merr.) is a tropical fruit that thrives in tropical and subtropical regions. The timing of its flowering significantly impacts both its yield and the optimal time for harvesting. While the PEBP family of genes is recognized for its crucial role in regulating flowering time and flower development (Jin et al., 2021; Kim et al., 2022), the specific characteristics of the PEBP gene family in pineapple and the specific members involved in the flowering process remain uncertain. This study comprehensively analyzed and identified all members of the PEBP gene family in the pineapple genome, encompassing their phylogenetic relationships, chromosome localization, gene structure, conserved motifs and promoter cis-acting elements. Additionally, considering the pivotal role of the *PEBP* gene in flower development, the expression pattern of AcPEBP and the expression characteristics of FT-like subfamily members were further investigated during flowering. Ultimately, this study provides a theoretical basis for future investigations into the regulatory role of *PEBP* genes in pineapple flowering.

## Materials and methods

### Identification of *AcPEBP* genes in pineapple

Download pineapple gene sequence data from the pineapple genome database (http://pineapple.angiosperms.org/) (Xu et al., 2018). Using the Arabidopsis PEBP protein sequence as a query, the amino acid sequence of Arabidopsis *PEBP* is obtained from the TAIR database (http://www.arabidopsis.org/) (Garcia-Hernandez et al., 2002). The pineapple database was searched with Blastp, and candidate pineapple AcPEBP family gene sequences were identified. After removing redundant sequences, the candidate AcPEBP family genes were submitted to the NCBI database (https://www.ncbi.nlm.nih.gov/cdd) to confirm the presence of the *PEBP* gene domains. The ExPASy website (http://web.expasy.org/protparam/) is to estimate the molecular weight (MW), isoelectric point (PI) and grand average of hydropathicity (GRAVY) for proteins found in pineapples (Wilkins et al., 1999). The subcellular localization prediction was performed using ProtComp v.9.0 (http://linux1.softberry.com/cgi-bin/programs/proloc/protcomppl.pl) (Li et al., 2020).

### Phylogenetic analyses

A multiple sequence alignment was conducted using ClustalX software for pineapple, Arabidopsis, and rice amino acid sequences. Based on 1000 bootstrap analyses performed using MEGA7.0 software with the neighbor-joining (NJ) algorithm (Kumar et al., 2016), the phylogenetic tree was constructed. The tree was visually

enhanced and beautified using the chiplot website (https://www.chiplot.online/).

## Conserved motifs and exon-intron structure analysis

Based on the whole-genome GFF annotation file, the analysis of exons and introns was performed using the Gene Structure Display Server 2.0 (http://gsds.cbi.pku.edu.cn). Based on the conserved domain information obtained from NCBI (https://www.ncbi.nlm.nih.gov/), the conserved motifs of the AcPEBP gene family were determined using the MEME online tool (http://meme-suite.org/tools/meme) (Bailey et al., 2009). The results of gene structure and conserved motif analysis were visualized using the TBtools software (Chen et al., 2020).

## *Cis*-element identification and protein interaction prediction analysis

The promoter sequence of the *AcPEBP* gene, spanning 2000bp upstream of the transcription start site, was submitted to the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) for the speculation of cis-regulatory elements in the promoter region (Lescot et al., 2002). The protein-protein interaction (PPI) of AcPEBP was prophesied using the STRING website (https://cn.string-db.org/) (Szklarczyk et al., 2021).

## Pineapple AcPEBP synteny analysis by chromosome location and duplication event

By utilizing the genome sequence and general feature format (GFF) file, the TBtools software was employed to retrieve the chromosomal positions of the pineapple PEBP family genes, enabling the creation of a gene localization information map. Obtain the genome sequences of maize, banana, rice, tomato, grape and *Arabidopsis* by downloading them from the Ensembl plants database (http://plants.ensembl.org/index.html). Collinearity files between each pair of species were obtained using the MCScanX software (http://chibba.pgml.uga.edu/mcscan2/) (Wang et al., 2012). The TBtools software was used to generate the collinearity analysis plot.

## Expression profiling of *AcPEBP* genes by RNA-Seq

The expression data for the *AcPEBP* genes in various pineapple tissues were obtained from the pineapple transcriptome dataset. Furthermore, the transcriptome data for pineapple were publicly deposited in the National Genomics Data Center database (https://ngdc.cncb.ac.cn/gsa/ (accessed on 6 May 2022)), The assigned accession of the submission is CRA006826.The quantification of transcript abundance for each *AcPEBP* gene was performed using the Fragments Per Kilobase of exon per Million mapped reads (FPKM) metric. The visualization of RNA-seq data was achieved by generating heatmaps using TBtools.

## Plant materials and treatments

Pineapple (*Ananas comosus* cv. Comte de Paris) was sourced from the Zhanjiang Subtropical Crop Research Institute in China. Plant materials approximately 30 cm long were selected when the pineapples were around 15 months old. The plants were induced with 30 mL of 200 mg/L$^{-1}$ ethephon, while an equivalent amount of water was used as a control for the untreated materials. Apical buds of pineapple were collected at different time points, including 0 h, 12 h, 1 d, 2 d, 4 d, 1 w, 2 w, 3 w, 4 w, 5 w, 6 w and 7 w after treatment. Each sample was collected with three separate biological replicates. The specimens were rapidly placed in liquid nitrogen and then kept at -80°C in a freezer for later RNA extraction.

## RNA extraction and real-time quantitative PCR analysis

Total RNA was extracted from pineapple samples using a polysaccharide polyphenol plant RNA extraction kit (China Huatuanyang Biotechnology Co., LTD.), and the quality and concentration of total RNA were determined by NanoDropTM One/OneC (Thermo Fisher Scientific, USA), an ultra-fine spectrophotometer. Total RNA First Strand cDNA was synthesized by RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA). qRT-PCR was performed using the SYBR Green qPCR Master MIX Kit (Thermo Fisher Scientific, USA). Employing the *AcActin* gene as the internal reference gene, qRT-PCR was performed to calculate the relative expression level of the *AcPEBP* gene through the $2^{-\Delta\Delta Ct}$ method (Rao et al., 2013). Each sample was replicated three times.

## Subcellular localization of PEBP proteins

To validate the subcellular localization of the AcPEBP protein in plant cells, the full-length sequence of *AcFT4* was cloned into the pCAMBIA2300-GFP vector to generate the fusion construct (2300 GFP-*AcFT4*). The fusion construct and control vector were transformed into the Agrobacterium tumefaciens strain GV3101 (ANGYUBio, China). Subsequently, they were used for transformation into healthy tobacco leaves. After 2-3 d, the transient expression of *AcFT4*-GFP was observed using a confocal laser scanning microscope (Zeiss, Germany).

## Transcriptional activity of PEBP proteins

The pGBKT7-*AcFT4* fusion vector was constructed and transferred into the yeast strain AH109 for the analysis of AcFT4 protein's transcriptional activation activity (AngYuBio, China). The

transformed colonies were streaked on selective medium SD/−Trp and incubated at 30°C for 48-72 h. Subsequently, the colonies were further selected on SD/−Trp/−His/−Ade medium, and X-α-Gal was added to observe if the AcFT4 protein exhibited transcriptional activation activity.

## Expression vector construction and plant transformation

The agrobacterium containing the recombinant plasmid was introduced into *Arabidopsis* flowers through the floral dip method (Zhang et al., 2006). *Arabidopsis* plants were cultured at room temperature (22°C) until they matured. Subsequently, transgenic positive plants were selected on MS medium supplemented with 50 mg/mL kanamycin. Healthy and green *Arabidopsis* plants were chosen for rapid PCR-based DNA screening to confirm positive transgenic individuals. The selected positive seedlings were further screened for T3 generation, to obtain homozygous transgenic lines, and the phenotypes of the transgenic homozygous lines were observed.

# Results

## Characterization and identification of PEBP proteins in pineapple

The BLASTP search was conducted in the pineapple genome database, utilizing the Arabidopsis PEBP protein as the query. Following the elimination of redundant members, a total of 11 *AcPEBP* genes were identified and subsequently named based on their resemblance to homologous genes in Arabidopsis. These genes encompass *AcFT1-AcFT7*, *AcTFL1a-AcTFL1b* and *AcMFT1-AcMFT2*. As depicted in Table 1, the AcPEBP family consists of 11 members, exhibiting varying protein lengths that range from 110 amino acids (AcMFT2) to 238 amino acids (AcFT6). The molecular weights of the proteins in this family vary from 12.34 kDa (AcMFT2) to 26.93 kDa (AcFT6). The isoelectric points of these proteins range from 6.96 (AcMFT1) to 9.61 (AcFT1). All members of this protein family exhibit a negative GRAVY value, indicating their hydrophobic nature. Subcellular localization prediction analysis revealed that, with the exception of AcMFT1, which is localized in Chloroplasts and Cytoplasm, the other AcPEBP member proteins are predicted to be located in the nucleus. Additionally, AcFT1 and AcMFT2 are also distributed in the Cytoplasm.

## Phylogenetic analysis

In order to examin the evolutionary relationships among AcPEBP members, a phylogenetic tree (Figure 1) was generated using the neighbor-joining technique with MEGA software. This analysis was based on the full-length amino acid sequences obtained from pineapple, Arabidopsis, and rice. The resulting phylogenetic tree classified the 11 AcPEBP proteins, 5 AtPEBP proteins, and 19 OsPEBP proteins into three distinct subfamilies, namely FT-like, TFL-like and MFT-like, as depicted in Figure 1. Among the AcPEBP family members, the FT-like subfamily displayed the greatest abundance of members, with a total of 7 AcPEBP proteins. In contrast, both the TFL-like and MFT-like subfamilies consisted of only 2 members each.

## Gene structures, conserved domains, and motif analysis

To acquire a more thorough comprehension of the AcPEBP gene members, we utilized the MEME website to forecast the conserved motifs present within the AcPEBP family. As illustrated in Figure 2, the number of conserved motifs in each

TABLE 1 PEBP gene family physicochemical characterization in pineapple.

| Gene | GeneBank ID | Protein/AA | MW (kDa) | pI | GRAVY | Subcellular localization |
|---|---|---|---|---|---|---|
| *AcFT1* | Aco004292.1.v3 | 176 | 19.80 | 9.61 | -0.261 | Cytoplasm. Nucleus. |
| *AcFT2* | Aco004692.1.v3 | 222 | 25.26 | 9.03 | -0.627 | Nucleus. |
| *AcFT3* | Aco010683.1.v3 | 215 | 24.27 | 7.83 | -0.101 | Nucleus. |
| *AcFT4* | Aco010684.1.v3 | 177 | 19.84 | 6.96 | -0.214 | Nucleus. |
| *AcFT5* | Aco006026.1.v3 | 180 | 20.18 | 8.74 | -0.280 | Nucleus. |
| *AcFT6* | Aco003470.1.v3 | 238 | 26.93 | 8.70 | -0.206 | Nucleus. |
| *AcFT7* | Aco008070.1.v3 | 231 | 25.92 | 8.91 | -0.223 | Nucleus. |
| *AcTFL1a* | Aco031443.1.v3 | 173 | 19.37 | 9.00 | -0.360 | Nucleus. |
| *AcTFL1b* | Aco016718.1.v3 | 147 | 16.75 | 9.09 | -0.463 | Nucleus. |
| *AcMFT1* | Aco006778.1.v3 | 184 | 20.81 | 6.96 | -0.296 | Chloroplasts. Cytoplasm. |
| *AcMFT2* | Aco012712.1.v3 | 110 | 12.34 | 9.30 | -0.462 | Cytoplasm. Nucleus. |

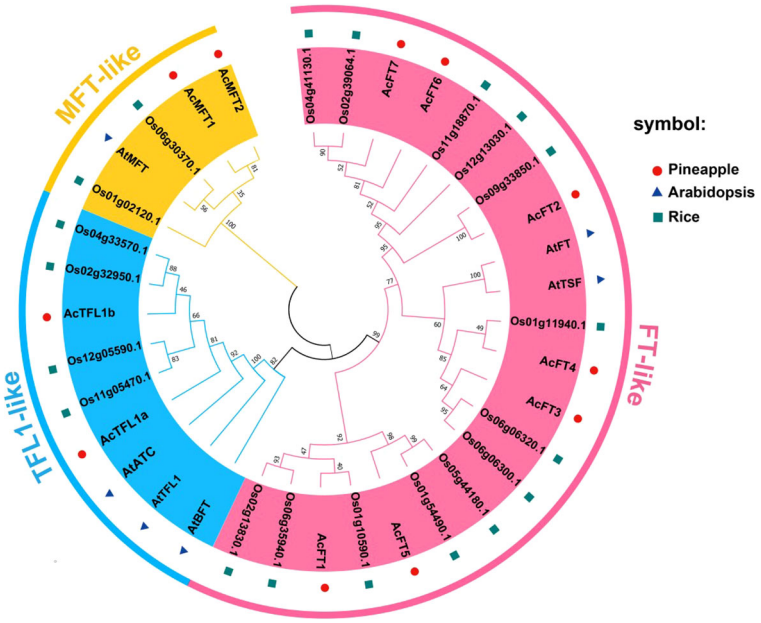



FIGURE 1
The phylogenetic analysis of PEBP proteins among pineapple, *Arabidopsis* and rice. The phylo-genetic tree was constructed using MEGA7.0 program. (The neighbor-joining (NJ) algorithm, bootstrap value = 1000). The phylogenetic connections and structural characteristics of the 11 AcPEBP genes.

AcPEBP protein ranges from Motif 4 to Motif 8. All members include motifs Motif 2, Motif 4, and Motif 5, while a majority of the members exhibit motifs 1 through 5 (Figure 2B). This implies that the AcPEBP protein demonstrates a notable degree of conservation among its constituents. Figure 2C shows that the quantity of introns in the AcPEBP gene family, ranging from 2 to 5. Among them, the MFT subfamily exhibits the highest number of introns in *AcMFT1*, reaching 5, while *AcMFT2* within the same subfamily displays the lowest count, with only 2 introns. Furthermore, other subfamily members share similarities in terms of the amount of introns and exons, suggesting a potential correlation between genetic structure and specific biological functions, regulatory pathways, or evolutionary relationships (Liu et al., 2021).

## *Cis*-element analysis of AcPEBPs

For the purpose of comprehending the potential roles and expression regulatory mechanisms of the *AcPEBP* gene, cis-acting elements located within the gene's promoter region (upstream 2 kb region) were predicted using the PlantCARE website (Figure 3). A total of 27 cis-acting elements were identified, encompassing 7 light-responsive elements, 9 plant hormone-responsive elements, 5 stress-related elements and 6 growth and development-responsive elements. The plant hormone-responsive elements, including encompassing gibberellin-responsive elements (TATC-box, P-box and GARE motif), auxin-responsive elements (TGA-element and AuxRR core), salicylic acid-responsive elements (TCA-element),

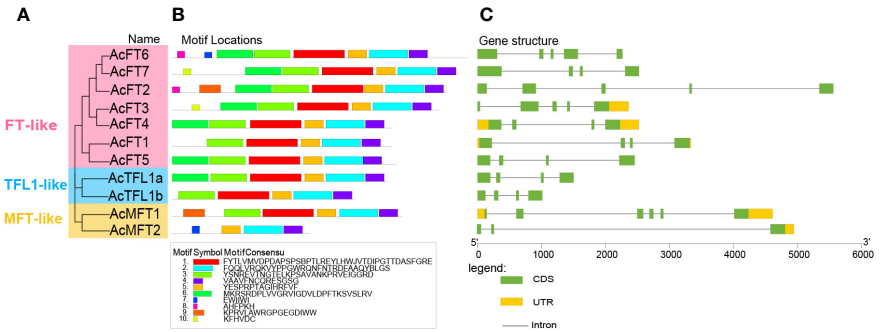



FIGURE 2
Phylogenetic connections and structural characteristics of the 11 *AcPEBP* genes. **(A)** Phylogenetic analysis using the Neighbor-Joining method for the PEBP family in pineapple. **(B)** The patterns of predicted motifs within the AcPEBP proteins. **(C)** Gene structures of PEBP family genes in pineapple. Yellow icons represent untranslated regions (UTR), green icons represent coding sequences (CDS), and lines connecting UTR and CDS represent introns (a noncoding segments of a gene or mRNA).

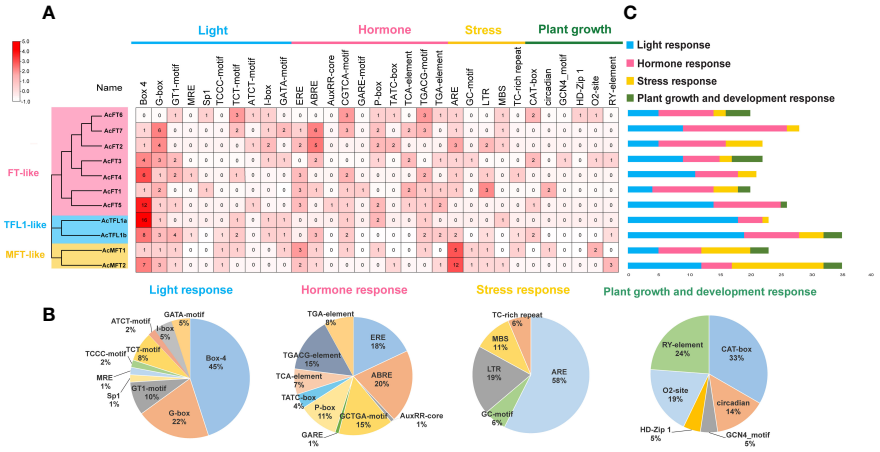



FIGURE 3
Analysis of cis-elements within the promoters of *AcPEBP* genes. **(A)** The *cis*-element analysis of *PEBP* promoter region. **(B)** The count of various types of elements in *AcPEBP* promoters was tallied and is represented using different colors. **(C)** The pie chart shows the proportion of each cis-element of the four types of response elements.

abscisic acid-responsive elements (ABRE), as well as CGTCA and TGACG motifs involved in MeJA (methyl jasmonate) responses, are the most abundant. Additionally, the promoters of 10 *AcPEBP* gene family members contain light-responsive elements Box 4 and G-box. This observation suggests that the AcPEBP gene family may exhibit responsiveness to both hormonal and light signals.

## Chromosomal position and gene duplication of the *AcPEBP* gene

In order to comprehend the distribution characteristics and gene replication of the PEBP gene, the genomic annotation information of pineapple was employed to ascertain the chromosomal localization of the *AcPEBP* gene. As depicted in Figure 4, the 11 genes exhibit an uneven distribution across 6 chromosomes, constituting approximately 1/4 of the 25 chromosomes present in pineapple. Notably, chromosome 5 harbors the highest number of three *AcPEBP* genes; whereas chromosomes 1, 10 and 17 each contain two *AcPEBP* genes, there is only one *AcPEBP* gene on chromosome 16.

To unravel the expansion mechanism of the pineapple PEBP gene family, gene collinearity analysis was performed. The results, as shown in Figure 5, revealed that 11 *AcPEBP* genes were involved in 4 segmental duplication events (*AcFT1*/*AcFT5*, *AcFT6*/*AcFT7*, *AcFT2*/*AcFT7*, *AcFT2*/*AcTFL2*), as well as 1 tandem duplication event (AcFT3/AcFT4) (Figure 5A). With the exception of *AcFT2*/*AcTFL2*, the majority of genes implicated in the duplication events belong to the FT-like subfamily. This observation suggests that gene duplication appears to be the primary mechanism driving the expansion of the pineapple FT-like subfamily. To enhance comprehension of the evolutionary trajectory and source of the AcPEBP gene family, collinearity maps were generated comparing pineapple with three dicotyledonous plants (*Arabidopsis*, grape and tomato), as well as three monocotyledonous plants (banana, maize

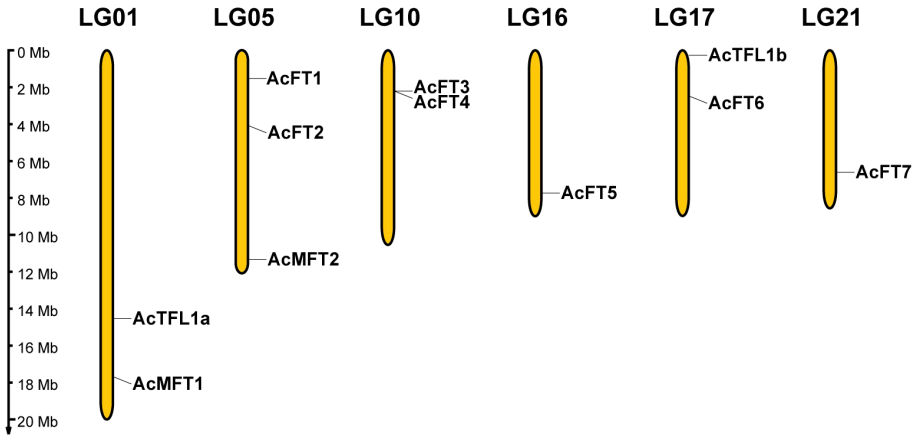



FIGURE 4
Distribution of AcPEBP genes on the pineapple chromosomes. The yellow vertical lines indicate the pineapple chromosome (Chr), the number is above the Chr. The left scale bar is the length of the chromosome.

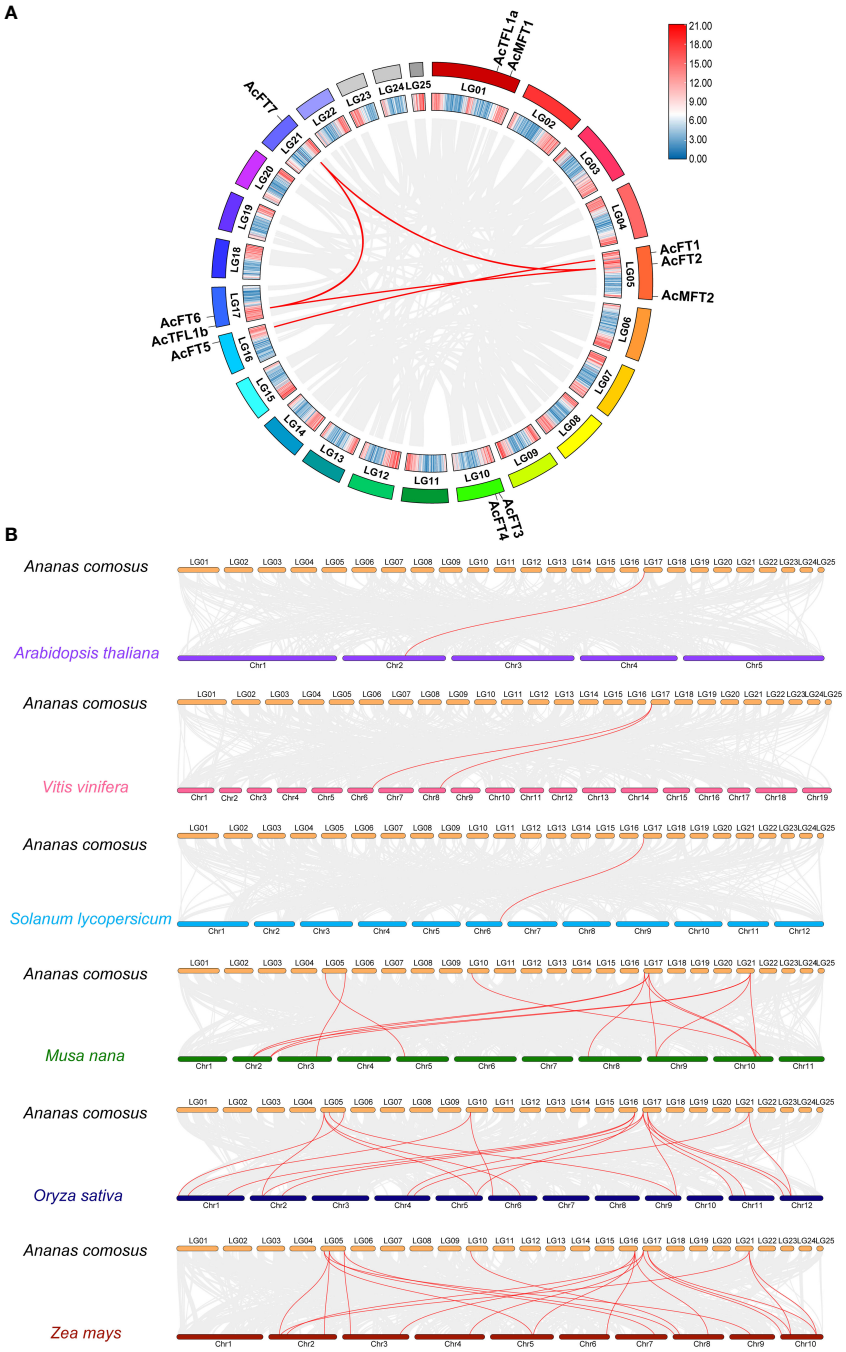



FIGURE 5
Duplication analysis and synlinearity analysis. **(A)** Duplicate analysis of *AcPEBP* gene in pineapple. **(B)** Synteny analysis of *AcPEBP* gene between pineapple and other plants (*Arabidopsis*, grape, tomato, banana, rice and corn). The red line indicates homologous gene pairs.

and rice) (Figure 5B). Pineapple displays 13, 20 and 19 homologous gene pairs with banana, maize and rice, respectively, notably higher count compared to the homologous gene pairs observed between pineapple and *Arabidopsis* (1), tomato (1) and grape (2). This phenomenon can be attributed to the fact that pineapples pertain to the category of monocotyledonous plants, thereby establishing a more intimate evolutionary connection with other species falling under the same category, while exhibiting a comparatively distant evolutionary association with dicotyledonous species.

## AcPEBP protein interaction network

The construction of the protein-protein interaction network of AcPEBP was based on the interactions of homologous proteins in *Arabidopsis thaliana* (Figure 6). Noteworthy findings were observed within this network. Specifically, in the FT interaction network (Figure 6A), CO, FD, and AP1 as crucial components in the *Arabidopsis* FT floral-dependent regulation (Komeda, 2004). Additionally, AcTFL1 may interact with VIN3, CAL, AGL8,

AGL24, SOC1, and LFY (Figure 6B). The regulation of these genes likely entails specific nodes within the flowering regulatory network that synchronize the timing of flowering and orchestrate the progression of the flowering process (Azpeitia et al., 2021). DECOY and MLE2.7 appear in the interaction networks of three subfamilies (Figure 6), suggesting their potential significance in plant growth, development, and other biological processes. Consequently, conducting a thorough investigation into the pivotal factors governing the regulation of FT will yield a more comprehensive comprehension of the flowering regulatory mechanism exhibited by PEBP family constituents in pineapple.

## Expression profiles of *AcPEBP* in different tissues in pineapple

To explore the expression patterns of the AcPEBP gene family across various tissues, transcriptomic data to determine the expression levels of *PEBP* genes in pineapple buds, flowers, fruits, leaves and roots. Our findings revealed distinct tissue specificity in the expression of AcPEBP family members (Figure 7). Notably, *AcFT3* and *AcFT4* exhibited significantly higher expression levels in flowers and buds compared to other tissue sites. Additionally, *AcFT6* and *AcMFT1* displayed the highest expression levels in fruits and leaves, respectively;*AcMFT1* and *AcFT7* also exhibited elevated expression levels in roots. In addition to these members, *AcFT1*, *AcFT2*, *AcFT5*, *AcTFL1a*, *AcTFL1b* and *AcMFT2* are almost not expressed. Consequently, it can be inferred that *AcFT3* and *AcFT4* likely assume a significant function in the process of flowering and flower development in pineapple.

## Expression patterns of *AcPEBP* during flower induction and the flowering process in pineapple

To enhance the credibility of *AcPEBP* in flower induction and flowering processes, a thorough investigation was conducted to assess the expression patterns of *AcPEBP* at different stages of flower development. As illustrated in Figure 8, the *AcPEBP* gene exhibited distinct patterns of expression. However, of greater interests to us is the observation that the members exhibiting elevated expression levels during the late stages of flower development in pineapple are predominantly concentrated within the FT-like subfamily. For example, during the course of the flowering process, the expression levels of *AcFT3* and *AcFT4* exhibited a gradual increment in conjunction with the development of flower buds (Figure 8A). qRT-PCR findings indicated that the expression levels of *AcFT3* and *AcFT4* genes in flower bud differentiation gradually increased after 4 days of ethethylene treatment and culminating in a peak at 5 to 7 weeks (Figure 8C). Furthermore, it is noteworthy that the fifth week following ethylene catalysis emerges as a pivotal phase for flower organ differentiation. Subsequently, an in-depth examination of the expression patterns of the *AcPEPB* gene in various flower organs of the pineapple was conducted (Figure 8B). The expression levels of *AcFT7*, *AcFT3* and *AcFT4* in different flower organs of pineapple were significantly increased.The concordance between the RNA-seq data and the qRT-PCR analysis results provides additional validation for the experimental findings. These findings suggest that the FT-like subfamily within the PEBP family may have a significant role in the pineapple flowering process.

## Localization within cells and transcriptional activity of AcFT4

To elucidate the potential functions of *AcFT4* in the transcriptional regulatory system, a recombinant *AcFT4*-GFP protein was synthesized and temporarily expressed in the leaves of tobacco plant (*Nicotiana benthamiana*). GFP originating from the *AcFT4*-GFP fusion protein were observed in both the nucleus and peripheral cytoplasm (Figure 9), which is consistent with the results observed in other species (Harig et al., 2012; Liu et al., 2022). To investigate whether the AcFT4 protein has transcriptional activity, the coding sequence of AcFT4 was integrated with the GAL4 DNA-binding domain encoding sequence in the pGBKT7 vector. The resulting recombinant construct was introduced into

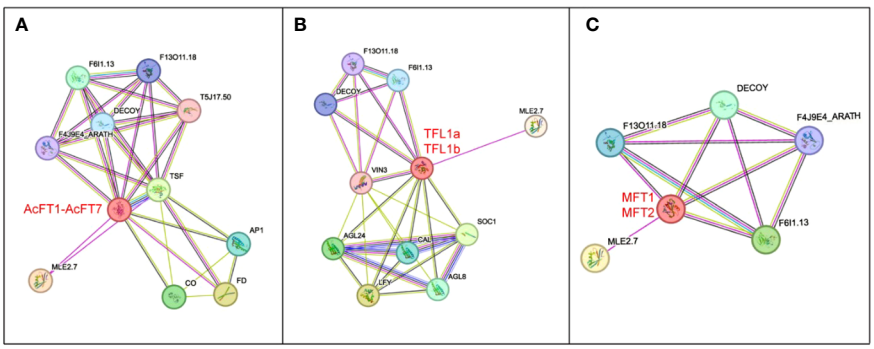



FIGURE 6
Depicts the predicted protein-protein interaction regulatory network of AcPEBP. **(A-C)** Respectively represent the network predictions for AcFTs, AcTFLs and AcMFTs, along with their interacting proteins.

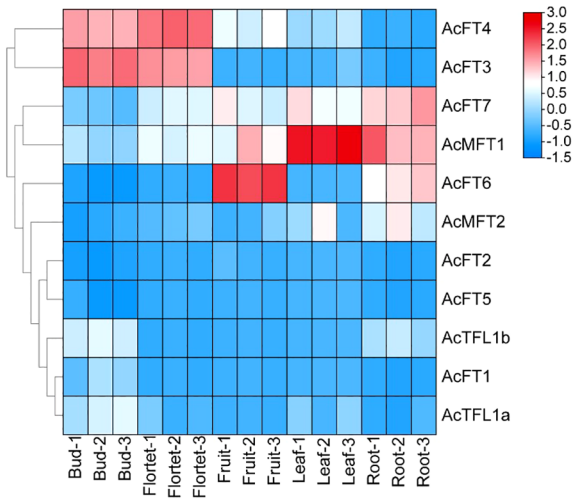



FIGURE 7
Expression profile of *AcPEBP* genes in different tissues of pineapple. The labels of the X-axis represent the different organs of the pineapple. The Y-axis refers to the deduced FPKM value normalized with Log2.

the yeast strain AH109. The findings revealed robust growth of yeast colonies on the SD/-Trp medium (Figure 10). In addition to negative controls, blue yeast colonies transformed with pGBT7-*AcFT4* were observed on SD/-Trp/-His/-Ade medium. This observation indicates that AcFT4 may indeed participate in the transcriptional regulatory system.

## Over-expression of *AcFT4* in *Arabidopsis* strongly accelerated flowering

The 35S::*AcFT4* construct was introduced into wild-type *Arabidopsis thaliana* to assess whether *AcFT4* induces flower development. The empty vector with the CaMV35S promoter was transformed into *Arabidopsis* plants as the negative control (35S-COL). After kanamycin screening and PCR identification, three T3 transgenic plants were selected for further analysis, taking into consideration the expression level of *AcFT4*. Compared with the controls, all transgenic strains showed different degrees of early flowering (Figures 11A–C). Transgenic plants overexpressing *AcFT4* had an earlier flowering time of 6-7 days and had 5-6 fewer rosette leaves at flowering (Figures 11B, C). These data suggest that overexpression of *AcFT4* can promote flowering in transgenic *Arabidopsis thaliana*.

# Discussion

The PEBP family in plants ethanolamine-binding proteins that are highly conserved. These genes are crucial in regulating various processes such as the transition to flowering, seed development and dormancy, and the formation of inflorescence structure in higher plants (Yue et al., 2021; Chen F. et al., 2021; Périlleux et al., 2019). Despite the extensive research on the PEBP family in other plant species, the function of this family in pineapples remains inadequately investigated. Therefore, there is a pressing need for

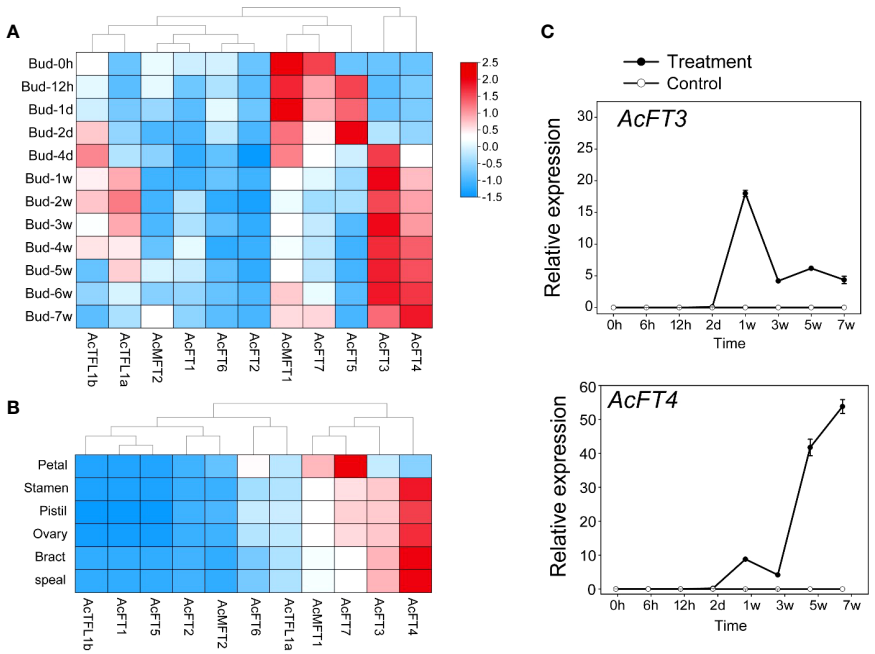



FIGURE 8
Expression profile of AcPEBP genes in flower induction and flowering process. **(A)** A heatmap illustrating the transcriptional levels of AcPEBP genes as determined by RNA-seq. **(B)** Expression pattern analysis of AcPEBP genes across various floral organs in pineapple. **(C)** Validation of AcPEBP gene expressions using qRT-PCR.

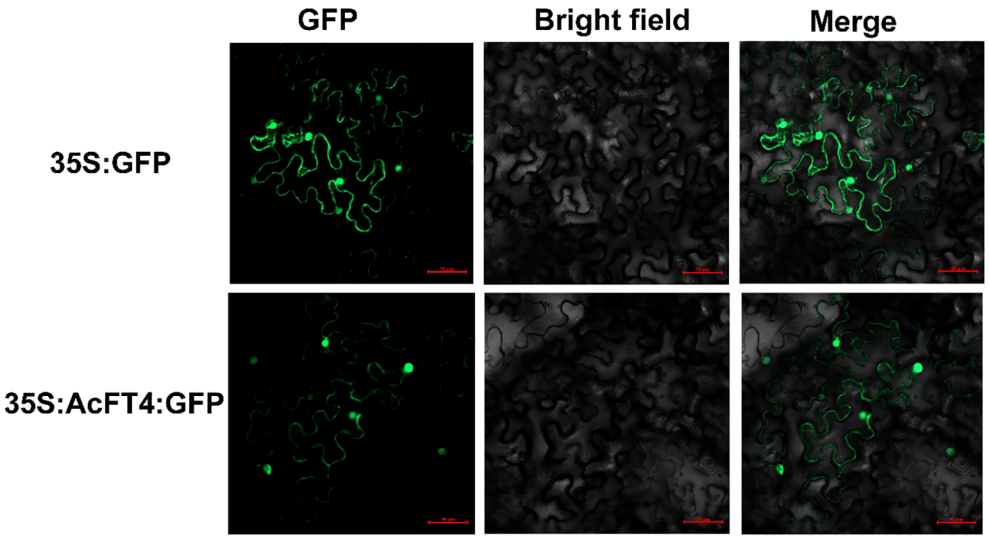



FIGURE 9
Subcellular localization of *AcPEBP* genes. Each candidate gene was individually inserted into the pCAMBIA2300-GFP vector. The empty vector was used as the control.

further investigation into the function of the AcPEBP family in pineapples. In this study, 11 PEBP genes were discovered within the pineapple genome database. Similar to *Arabidopsis* and rice, the PEBP gene family within the pineapple genome was also divided into three subfamilies: FT-like, TFL1-like and FT-like. This is also consistent with the classification of fruit tree crops such as grapes (Carmona et al., 2007). However, there are significant differences in the number of PEBP family members between species, potentially attributed to gene duplication or deletion events during the course of evolution. A previous study suggested that exon and intron structures were involved in the evolution of plant species (Zhang and Li, 2017). In particular, most genes in the FT/TFL1 subfamily predominantly features a gene structure comprising four exons and three introns (Figure 2C), and they show a similar pattern of conserved motif distribution. This suggests that these genes are conserved in evolution. However, in contrast to the FT/TFL1 subfamily, the *AcMFT2* gene had five introns, while the *AcMFT1* gene had only two introns. The presence of exon-intron acquisition and deletion within the MFT-like subgroup of pineapple, along with substantial alterations in the conserved pattern, suggests potential functional diversity among MFT-like genes. (Xi and Yu, 2010; Song et al., 2020).

At present, it is generally believed that the expression pattern of genes is typically associated with their biological functions (Dong et al., 2020). Previous studies have demonstrated that *Arabidopsis AtFT* gene exhibits expressed in both reproductive and vegetative organs (Kobayashi et al., 1999). *RoFT* gene of rose is expressed in flower bud tips and buds (Remay et al., 2009). The expression of the *VvFT* gene was detected during the development of grape inflorescence (Carmona et al., 2007). In apples, *MdFT1* primarily assumes a significant role in apple fruit development, of whereas *MdFT2* is predominantly expressed in the reproductive organs of apple, including flower buds and young fruits. The *NtFT5* gene in tobacco is expressed in tobacco flowers (Wang et al., 2018). In this

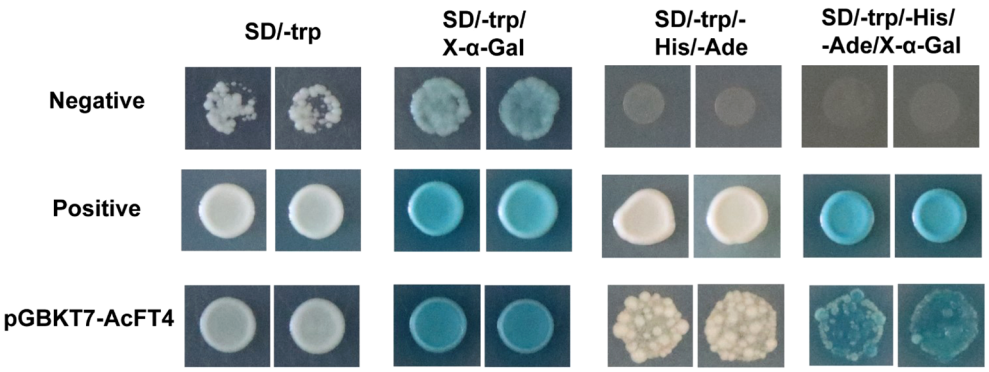



FIGURE 10
Assessment of the transcriptional activity of AcFT4 in yeast. The negative control involved the transformation of the empty vector pGBKT7 into yeast. The positive control consisted of the co-transformation of pGADT7-T and pGBKT7-53 into yeast.

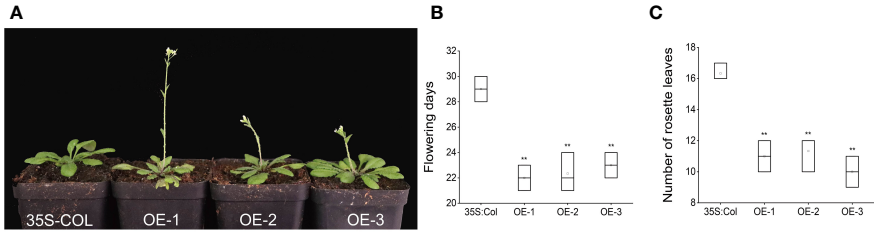



FIGURE 11
*AcPEBP* plants display an early flowering phenotype. **(A)** Transgenic lines. **(B)** Flowering days. **(C)** Number of rosette leaves. Values represent the mean ± SD of three independent biological replicates. The asterisk denotes a significantly different compared with controls (**P < 0.05, based on t-test).

study, the expression pattern of the AcPEBP gene family exhibited certain tissue specificity. Specifically, *AcFT7* and *AcFT6* were highly expressed in the root and fruit of pineapple, respectively, and *AcMFT1* exhibited high expression levels in the root and leaf tissues. These genes likely play pivotal roles within the contexts of roots, leaves and fruits. It is worth noting that *AcFT3* and *AcFT4* not only exhibited higher transcription levels in flower buds compared to other tissues, but also displayed elevated expression levels in flowers (Figure 7). After that, the expression of *AcFT3* and *AcFT4* was observed to be significantly specific across five distinct flower organs of pineapple (Figure 8B). This suggests that the *AcFT* genes potentially exert a significant regulatory on the process of flower bud formation and flowering in pineapple. In order to further verify the results, the expression of the *FT* gene in pineapple flower buds was analyzed assessed through qRT-PCR. The results showed that the expression level of *AcFT4* began to increase within just 4 days after ethylene treatment and reached a peak in late flower development (Figure 8C). This indicates that members of the pineapple PEBP family might play an important role in the flowering process of pineapple.

Simultaneously, examination of promoter cis-acting elements has suggested the possible participation of light-responsive and hormone-responsive within the intricate flowering regulatory network. (Figure 6). This phenomenon has been previously observed in various plant species. The peanut FT-like gene exhibits higher elevated expression levels in short-day conditions and is involved in the regulation of flowering time under short-day conditions (Jin et al., 2019).In *Dendrobium huoshanense*, gibberellin acid (GA) induces strong expression of *DhFT3* and *DhFT10*, while the expression of *DhTFL1* swiftly declines following GA treatment. This may have different regulatory roles in flowering control (Song et al., 2021). The short-day plant *Pharbitis nil FT* homolog *PnFT2* may be stimulated through the application of salicylic acid (SA) (Wada et al., 2010). In a recent groundbreaking discovery, Chen Y. et al. revealed that *ERF1* functions as a regulatory factor in ethylene signaling, directly suppressing the expression of *FT* in *Arabidopsis* during a complex flowering cascade, delaying flowering (Chen Y. et al., 2021). Interestingly, treatment of pineapple plants with the ethylene-releasing agent ethephon has been demonstrated to induce early flowering of pineapples (Espinosa et al., 2017). Therefore, we posit that the flowering pathway in pineapples may differ from that of the model plant *Arabidopsis*.

FT-like genes in plants play a crucial role in regulating flowering (Takagi et al., 2023). This study demonstrated that the overexpression of *AcFT4* in *Arabidopsis* resulted in early flowering (Figure 11). Transgenic lines showed flowering time 6-7 days earlier than empty vector plants with CaMV35S promoters. This finding suggests that *AcFT4* acts as a flower promoting factor, consistent with other species. For instance, the ectopic expression of *ZCN8* in maize accelerates flowering in transgenic *Arabidopsis* (Danilevskaya et al., 2008). Overexpression of *MiFT1*, *MiFT2* and *MiFT3* in mango has been found to induce early flowering phenotypes in *Arabidopsis* (Fan et al., 2020). In strawberries, *FveFT2* is a non-photoperiodic florigen that allows short-day plants to flower, and its excessive expression greatly enhances the flowering phenomenon (Gaston et al., 2021). In addition to the function of FT itself, key components CO, FD and LFY in the known flowering regulation network of *FT* genes were found in the protein interaction prediction network (Figure 6) (Blümel et al., 2015). These genes collaborate within signaling pathway, collectively orchestrating the process of flowering. Nevertheless, whether the *FT* gene in pineapple functions through these flowering regulatory factors still requires further investigation. This will contribute to a more comprehensive understanding of the molecular mechanisms of PEBP members in the pineapple flowering signaling pathway.

## Conclusions

In this study, 11 members of the PEBP family were identified from the pineapple genome annotation file. Analysis of *cis*-acting elements suggests that *AcPEBP* genes may be influenced by a complex network of hormonal and light regulation. The analysis of gene expression showed that *AcFT3* and *AcFT4* were specifically highly expressed in the later stages of flower development and in floral organs, suggesting that these genes may be involved in the process of pineapple flowering induction. Following this, the subcellular localization and transcriptional activation activities of AcFT4 were subsequently confirmed. AcFT4 exhibits

transcriptional activity and is localized in various cellular compartments, including the nucleus and peripheral cytoplasm. The overexpression of *AcFT4* in transgenic *Arabidopsis thaliana* resulted in an expedited flowering process. In addition, important components of floral-dependent regulation of FT, CO, LFY and AP1, were found in the protein interaction network. In conclusion, the genomic analysis of AcPEBP family provides a new basis for the study of flower development related functions of PEBP family members in pineapple.

## 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

HZ: Writing – review & editing, Funding acquisition. XZ: Data curation, Software, Writing – original draft. YO: Conceptualization, Writing – review & editing. LZ: Data curation, Investigation, Resources, Software, Writing – review & editing. ZL: Investigation, Resources, Writing – review & editing. YW: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The project was funded by the National Natural Science Fund of China (32160687 and 32360723), the major science and technology project of Hainan Province (ZDKJ2021014), the Natural Science Foundation of Hainan Province (321RC467), and the Scientific Research Start-up Fund Project of Hainan University (KYQD-ZR-20090).

## Conflict of interest

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

## Publisher's note

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

## References

Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., et al. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309, 1052–1056. doi: 10.1126/science.1115983

Adeyemo, O. S., Chavarriaga, P., Tohme, J., Fregene, M., Davis, S. J., and Setter, T. L. (2017). Overexpression of Arabidopsis FLOWERING LOCUS T (FT) gene improves floral development in cassava (Manihot esculenta, Crantz). *PloS One* 12, e0181460. doi: 10.1371/journal.pone.0181460

Adeyemo, O. S., Hyde, P. T., and Setter, T. L. (2019). Identification of FT family genes that respond to photoperiod, temperature and genotype in relation to flowering in cassava (Manihot esculenta, Crantz). *Plant Reprod.* 32, 181–191. doi: 10.1007/s00497-018-00354-5

Azpeitia, E., Tichtinsky, G., Le Masson, M., Serrano-Mislata, A., Lucas, J., Gregis, V., et al. (2021). Cauliflower fractal forms arise from perturbations of floral gene networks. *Science* 373, 192–197. doi: 10.1126/science.abg5999

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Blümel, M., Dally, N., and Jung, C. (2015). Flowering time regulation in crops—what did we learn from Arabidopsis? *Curr. Opin. Biotechnol.* 32, 121–129. doi: 10.1016/j.copbio.2014.11.023

Carmona, M. J., Calonje, M., and Martínez-Zapater, J. M. (2007). The FT/TFL1 gene family in grapevine. *Plant Mol. Biol.* 63, 637–650. doi: 10.1007/s11103-006-9113-z

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

Chen, F., Li, Y., Li, X., Li, W., Xu, J., Cao, H., et al. (2021). Ectopic expression of the Arabidopsis florigen gene FLOWERING LOCUS T in seeds enhances seed dormancy via the GA and DOG1 pathways. *Plant J.* 107, 909–924. doi: 10.1111/tpj.15354

Chen, Y., Zhang, L., Zhang, H., Chen, L., and Yu, D. (2021). ERF1 delays flowering through direct inhibition of FLOWERING LOCUS T expression in Arabidopsis. *J. Integr. Plant Biol.* 63, 1712–1723. doi: 10.1111/jipb.13144

Danilevskaya, O. N., Meng, X., Hou, Z., Ananiev, E. V., and Simmons, C. R. (2008). A genomic and expression compendium of the expanded PEBP gene family from maize. *Plant Physiol.* 146, 250–264. doi: 10.1104/pp.107.109538

Dong, L., Lu, Y., and Liu, S. (2020). Genome-wide member identification, phylogeny and expression analysis of PEBP gene family in wheat and its progenitors. *PeerJ* 8, e10483. doi: 10.7717/peerj.10483

Espinosa, M. E. Á., Moreira, R. O., Lima, A. A., Ságio, S. A., Barreto, H. G., Luiz, S. L. P., et al. (2017). Early histological, hormonal, and molecular changes during pineapple (Ananas comosus (L.) Merrill) artificial flowering induction. *J. Plant Physiol.* 209, 11–19. doi: 10.1016/j.jplph.2016.11.009

Fan, Z. Y., He, X. H., Fan, Y., Yu, H. X., Wang, Y. H., Xie, X. J., et al. (2020). Isolation and functional characterization of three MiFTs genes from mango. *Plant Physiol. Biochem.* 155, 169–176. doi: 10.1016/j.plaphy.2020.07.009

Gafni, I., Rai, A. C., Halon, E., Zviran, T., Sisai, I., Samach., A., et al. (2022). Expression profiling of four mango FT/TFL1-encoding genes under different fruit load conditions, and their involvement in flowering regulation. *Plants (Basel)* 11, 2409. doi: 10.3390/plants11182409

Garcia-Hernandez, M., Berardini, T. Z., Chen, G., Crist, D., Doyle, A., Huala, E., et al. (2002). TAIR: a resource for integrated Arabidopsis data. *Funct. Integr. Genomics* 2, 239–253. doi: 10.1007/s10142-002-0077-z

Gaston, A., Potier, A., Alonso, M., Sabbadini, S., Delmas, F., Tenreira, T., et al. (2021). The FveFT2 florigen/FveTFL1 antiflorigen balance is critical for the control of seasonal flowering in strawberry while FveFT3 modulates axillary meristem fate and yield. *New Phytologist* 232, 372–387. doi: 10.1111/nph.17557

Harig, L., Beinecke, F. A., Oltmanns, J., Muth, J., Müller, O., Rüping, B., et al. (2012). Proteins from the FLOWERING LOCUS T-like subclade of the PEBP family act antagonistically to regulate floral initiation in tobacco. *Plant J.* 72, 908–921. doi: 10.1111/j.1365-313X.2012.05125.x

Jin, S., Nasim, Z., Susila, H., and Ahn, J. H. (2021). Evolution and functional diversification of FLOWERING LOCUS T/TERMINAL FLOWER 1 family genes in plants. *Semin. Cell Dev. Biol.* 109, 20–30. doi: 10.1016/j.semcdb.2020.05.007

Jin, H., Tang, X., Xing, M., Zhu, H., Sui, J., Cai, C., et al. (2019). Molecular and transcriptional characterization of phosphatidyl ethanolamine-binding proteins in wild peanuts Arachis duranensis and Arachis ipaensis. *BMC Plant Biol.* 19, 484. doi: 10.1186/s12870-019-2113-3

Kaneko-Suzuki, M., Kurihara-Ishikawa, R., Okushita-Terakawa, C., Kojima, C., Nagano-Fujiwara, M., Ohki, I., et al. (2018). TFL1-like proteins in rice antagonize rice FT-like protein in inflorescence development by competition for complex

formation with 14-3-3 and FD. *Plant Cell Physiol.* 59, 458–468. doi: 10.1093/pcp/pcy021

Kim, G., Rim, Y., Cho, H., and Hyun, T. K. (2022). Identification and functional characterization of FLOWERING LOCUS T in Platycodon grandiflorus. *Plants (Basel)* 11, 325. doi: 10.3390/plants11030325

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., and Araki, T. (1999). A pair of related genes with antagonistic roles in mediating flowering signals. *Science* 286, 1960–1962. doi: 10.1126/science.286.5446.1960

Komeda, Y. (2004). Genetic regulation of time to flower in Arabidopsis thaliana. *Annu. Rev. Plant Biol.* 55, 521–535. doi: 10.1146/annurev.arplant.55.031903.141644

Komiya, R., Ikegami, A., Tamaki, S., Yokoi, S., and Shimamoto, K. (2008). Hd3a and RFT1 are essential for flowering in rice. *Development* 135, 767–774. doi: 10.1242/dev.008631

Kotoda, N., Hayashi, H., Suzuki, M., Igarashi, M., Hatsuyama, Y., Kidou, S., et al. (2010). Molecular characterization of FLOWERING LOCUS T-like genes of apple (Malus x domestica Borkh.). *Plant Cell Physiol.* 51, 561–575. doi: 10.1093/pcp/pcq021

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Li, H., Ran, K., Dong, Q., Zhao, Q., and Shi, S. (2020). Cloning, sequencing, and expression analysis of 32 NAC transcription factors (MdNAC) in apple. *PeerJ* 8, e8249. doi: 10.7717/peerj.8249

Liu, J., Wang, J., Wang, M., Zhao, J., Zheng, Y., Zhang, T., et al. (2021). Genome-wide analysis of the R2R3-MYB gene family in Fragaria × ananassa and its function identification during anthocyanins biosynthesis in pink-flowered strawberry. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.702160

Liu, X., Zhao, D., Ou, C., Hao, W., Zhao, Z., and Zhuang, F. (2022). Genome-wide identification and characterization profile of phosphatidy ethanolamine-binding protein family genes in carrot. *Front. Genet.* 13. doi: 10.3389/fgene.2022.1047890

Oda, A., Narumi, T., Li, T., Kando, T., Higuchi, Y., Sumitomo, K., et al. (2012). CsFTL3, a chrysanthemum FLOWERING LOCUS T-like gene, is a key regulator of photoperiodic flowering in chrysanthemums. *J. Exp. Bot.* 63, 1461–1477. doi: 10.1093/jxb/err387

Périlleux, C., Bouché, F., Randoux, M., and Orman-Ligeza, B. (2019). Turning meristems into fortresses. *Trends Plant Sci.* 24, 431–442. doi: 10.1016/j.tplants.2019.02.004

Rao, X., Huang, X., Zhou, Z., and Lin, X. (2013). An improvement of the 2ˆ (-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostat. Bioinform. Biomath.* 3, 71–85. doi: 10.1089/cmb.2012.0279

Remay, A., Lalanne, D., Thouroude, T., Le Couviour, F., Hibrand-Saint Oyant, L., and Foucher, F. (2009). A survey of flowering genes reveals the role of gibberellins in floral control in rose. *Theor. Appl. Genet.* 119, 767–781. doi: 10.1007/s00122-009-1087-1

Song, C., Li, G., Dai, J., and Deng, H. (2021). Genome-wide analysis of PEBP genes in Dendrobium huoshanense: unveiling the antagonistic functions of FT/TFL1 in flowering time. *Front. Genet.* 12. doi: 10.3389/fgene.2021.687689

Song, S., Wang, G., Wu, H., Fan, X., Liang, L., Zhao, H., et al. (2020). OsMFT2 is involved in the regulation of ABA signaling-mediated seed germination through interacting with OsbZIP23/66/72 in rice. *Plant J.* 103, 532–546. doi: 10.1111/tpj.14748

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkaa1074

Takagi, H., Hempton, A. K., and Imaizumi, T. (2023). Photoperiodic flowering in Arabidopsis: Multilayered regulatory mechanisms of CONSTANS and the florigen FLOWERING LOCUS T. *Plant Commun.* 4, 100552. doi: 10.1016/j.xplc.2023.100552

Venail, J., da Silva Santos, P. H., Manechini, J. R., Alves, L. C., Scarpari, M., Falcão, T., et al. (2022). Analysis of the PEBP gene family and identification of a novel FLOWERING LOCUS T orthologue in sugarcane. *J. Exp. Bot.* 5, 2035–2049. doi: 10.1093/jxb/erab539

Wada, K. C., Yamada, M., Shiraya, T., and Takeno, K. (2010). Salicylic acid and the flowering gene FLOWERING LOCUS T homolog are involved in poor-nutrition stress-induced flowering of Pharbitis nil. *J. Plant Physiol.* 167, 447–452. doi: 10.1016/j.jplph.2009.10.006

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wang, G., Wang, P., Gao, Y., Li, Y., Wu, L., Gao, J., et al. (2018). Isolation and functional characterization of a novel FLOWERING LOCUS T homolog (NtFT5) in Nicotiana tabacum. *J. Plant Physiol.* 231, 393–401. doi: 10.1016/j.jplph.2018.10.021

Wen, C., Zhao, W., Liu, W., Yang, L., Wang, Y., Liu, X., et al. (2019). CsTFL1 inhibits determinate growth and terminal flower formation through interaction with CsNOT2a in cucumber. *Development* 146, dev180166. doi: 10.1242/dev.180166

Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U., et al. (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. *Science* 309, 1056–1059. doi: 10.1126/science.1114358

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., et al. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol. Biol.* 112, 531–552. doi: 10.1385/1-59259-584-7:531

Xi, W., and Yu, H. (2010). MOTHER OF FT AND TFL1 regulates seed germination and fertility relevant to the brassinosteroid signaling pathway. *Plant Signal Behav.* 5, 1315–1317. doi: 10.4161/psb.5.10.13161

Xu, H., Guo, X., Hao, Y., Lu, G., Li, D., Lu, J., et al. (2022). Genome-wide characterization of PEBP gene family in Perilla frutescens and PfFT1 promotes flowering time in Arabidopsis thaliana. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1026696

Xu, H., Yu, Q., Shi, Y., Hua, X., Tang, H., Yang, L., et al. (2018). PGD: pineapple genomics database. *Hortic. Res.* 5, 66. doi: 10.1038/s41438-018-0078-2

Yan, X., Cao, Q. Z., He, H. B., Wang, L. J., and Jia., G. X. (2021). Functional analysis and expression patterns of members of the FLOWERING LOCUS T (FT) gene family in Lilium. *Plant Physiol. Biochem.* 163, 250–260. doi: 10.1016/j.plaphy.2021.03.056

Yue, L., Li, X., Fang, C., Chen, L., Yang, H., Yang, J., et al. (2021). FT5a interferes with the Dt1-AP1 feedback loop to control flowering time and shoot determinacy in soybean. *J. Integr. Plant Biol.* 63, 1004–1020. doi: 10.1111/jipb.13070

Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W., and Chua, N. H. (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nat. Protoc.* 1, 641–646. doi: 10.1038/nprot.2006.97

Zhang, Q., and Li, C. (2017). Comparisons of copy number, genomic structure, and conserved motifs for α-amylase genes from barley, rice, and wheat. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01727

Zhang, B., Li, C., Li, Y., and Yu, H. (2020). Mobile TERMINAL FLOWER1 determines seed size in Arabidopsis. *Nat. Plants.* 6, 1146–1157. doi: 10.1038/s41477-020-0749-5

Zhang, H., and Zhang, Y. (2020). Molecularcloning and functional characterization of CmFT (FLOWERING LOCUS T) from Cucumis melo L. *J. Genet.* 99, 1–8. doi: 10.1007/s12041-020-1191-1

Zhu, Y., Klasfeld, S., Jeong, C. W., Jin, R., Goto, K., Yamaguchi, N., et al. (2020). TERMINAL FLOWER 1-FD complex target genes and competition with FLOWERING LOCUS T. *Nat. Commun.* 11, 5118. doi: 10.1038/s41467-020-18782-1

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 January 2024
DOI 10.3389/fpls.2023.1318383

Check for updates

OPEN ACCESS

EDITED BY
Eleni Tani,
Agricultural University of Athens, Greece

REVIEWED BY
Panagiotis Madesis,
University of Thessaly, Greece
Parviz Heidari,
Shahrood University of Technology, Iran

*CORRESPONDENCE
Chenjing Shang
cjshang@szu.edu.cn

RECEIVED 11 October 2023
ACCEPTED 14 December 2023
PUBLISHED 04 January 2024

CITATION
Hussain Q, Ye T, Shang C, Li S, Khan A, Nkoh JN, Mustafa AE-ZMA and Elshikh MS (2024) *NRAMP* gene family in *Kandelia obovata*: genome-wide identification, expression analysis, and response to five different copper stress conditions.
*Front. Plant Sci.* 14:1318383.
doi: 10.3389/fpls.2023.1318383

COPYRIGHT
© 2024 Hussain, Ye, Shang, Li, Khan, Nkoh, Mustafa and Elshikh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# *NRAMP* gene family in *Kandelia obovata*: genome-wide identification, expression analysis, and response to five different copper stress conditions

Quaid Hussain [1,2], Ting Ye [1], Chenjing Shang [1]*, Sihui Li [1], Asadullah Khan [1,2], Jackson Nkoh Nkoh [1,2], Abd El-Zaher M. A. Mustafa [3] and Mohamed S. Elshikh [3]

[1]Shenzhen Engineering Laboratory for Marine Algal Biotechnology, Shenzhen Public Service Platform for Collaborative Innovation of Marine Algae Industry, Guangdong Engineering Research Center for Marine Algal Biotechnology, College of Life Science and Oceanography, Shenzhen University, Shenzhen, China, [2]College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China, [3]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh, Saudi Arabia

Natural resistance-associated macrophage proteins (NRAMPs) are a class of metal transporters found in plants that exhibit diverse functions across different species. Transporter proteins facilitate the absorption, distribution, and sequestration of metallic elements within various plant tissues. Despite the extensive identification of *NRAMP* family genes in various species, a full analysis of these genes in tree species is still necessary. Genome-wide identification and bioinformatics analysis were performed to understand the roles of *NRAMP* genes in copper ($CuCl_2$) stress in *Kandelia obovata* (Ko). In *Arachis hypogaea* L., *Populus trichocarpa*, *Vitis vinifera*, *Phaseolus vulgaris* L., *Camellia sinensis*, *Spirodela polyrhiza*, *Glycine max* L. and *Solanum lycopersicum*, a genome-wide study of the NRAMP gene family was performed earlier. The domain and 3D structural variation, phylogenetic tree, chromosomal distributions, gene structure, motif analysis, subcellular localization, cis-regulatory elements, synteny and duplication analysis, and expression profiles in leaves and $CuCl_2$ were all investigated in this research. In order to comprehend the notable functions of the NRAMP gene family in *Kandelia obovata*, a comprehensive investigation was conducted at the genomic level. This study successfully found five *NRAMP* genes, encompassing one gene pair resulting from whole-genome duplication and a gene that had undergone segmental duplication. The examination of chromosomal position revealed an unequal distribution of the *KoNRAMP* genes across chromosomes 1, 2, 5, 7, and 18. The *KoNRAMPs* can be classified into three subgroups (I, II, and SLC) based on phylogeny and synteny analyses, similar to *Solanum lycopersicum*. Examining cis-regulatory elements in the promoters revealed five hormone-correlated responsive elements and four stress-related responsive elements. The genomic architecture and properties of 10 highly conserved motifs are similar among members of the NRAMP gene family. The conducted investigations

demonstrated that the expression levels of all five genes exhibited alterations in response to different levels of $CuCl_2$ stress. The results of this study offer crucial insights into the roles of *KoNRAMPs* in the response of *Kandelia obovata* to $CuCl_2$ stress.

KEYWORDS

***Kandelia obovata*, NRAMP gene family, copper stress, phylogenetic analysis, gene expression profiling**

## 1 Introduction

Plants have developed adaptive mechanisms that facilitate the uptake of necessary nutrients and mitigate the adverse effects of heavy metal (HM) toxicity. One important process is the existence of membrane transport systems in plants, which are essential for numerous aspects of vital mineral and nutrient homeostasis as well as the accumulation, translocation, and detoxification of hazardous metals (Wu et al., 2021; Liu et al., 2022; Ma et al., 2023). Previous studies have demonstrated that NRAMPs play a crucial role as integral membrane transporters in various species, including plants. These proteins are responsible for metal uptake, transport, accumulation, and detoxification (Zhang et al., 2020; Tian et al., 2021). The *NRAMP* gene was initially identified in rats in 1993 (Vidal et al., 1993). The *NRAMPs* are a significantly conserved family of integral membrane proteins that play a crucial role in facilitating the transport of divalent metals across the cellular membrane. *NRAMP* genes have been detected in several organisms, including bacteria, fungi, plants, and mammals (Cellier et al., 1996; Segond et al., 2009). The NRAMP proteins have a high degree of conservation across various organisms, from bacteria to mammals. These proteins typically consist of 10-12 transmembrane domains (TMD) and possess a consensus transport signature (Cellier et al., 1995; Tan et al., 2023).

However, the majority of NRAMP proteins can transport several metal ions, including cobalt (Co), nickel (Ni), copper (Cu), iron (Fe), manganese (Mn), zinc (Zn), and cadmium (Cd) (Nevo and Nelson, 2006; Ishida et al., 2018). The roles of NRAMP proteins in plants have been investigated in *Arabidopsis thaliana* (Segond et al., 2009), *Camellia sinensis* (Li et al., 2021), *Spirodela polyrhiza* (Chen et al., 2021), *Glycine Max* L. (Qin et al., 2017), *Solanum tuberosum* (Tian et al., 2021), *Populus trichocarpa* (Ma et al., 2023), *Phaseolus vulgaris* L. (Ishida et al., 2018), *Arachis hypogaea* L. (Tan et al., 2023). Understanding the physiological role of these elements in plant development requires identifying a particular collection of transporters that supply the delicate balance of metal concentration across the cellular membrane (Ishida et al., 2018; Kravchenko et al., 2018). Nevertheless, the function of the NRAMP protein family and the Cu transport pathways in *Kandelia obovata* is still unclear.

Mangroves serve as a significant reservoir for HMs, with Cu, lead (Pb), and zinc being the predominant HM pollutants (Cheng et al., 2017) (Bodin et al., 2013). HMs represent a significant class of anthropogenic harmful substances within mangrove ecosystems (Huang and Wang, 2010). Among them, Cu, Zn, and Pb are often encountered pollutants (Qiu et al., 2011; Li et al., 2015). Shen et al. (2021) reported that HM enrichment on the root surface of *Kandeliaobovata* is higher than some other mangrove plants in South China, including *Acanthus ilicifolius*, *Aeagiceras corniculatum*, *Bruguiera gymnorrhiza*, and *Heritiera littorlis*. A viviparous mangrove species found in the intertidal zones of tropical and subtropical coasts, *Kandelia obovata* is a member of the Rhizophoraceae family in the Malpighiales (Sheue et al., 2003). The distribution of this species spreads from northern Vietnam to southeast China and south Japan in the East Asian region (Sheue et al., 2003; Yang et al., 2019). The native distribution of this species encompasses the Hainan, Guangdong (including Hong Kong and Macau), Guangxi, Fujian, and Taiwan Provinces in China (Li and Lee, 1997). *Kandelia obovata*, a mangrove species found in China, exhibits notable cold resistance and salt tolerance, enabling its survival in northern regions (Yang et al., 2019). Prior research has demonstrated that this species could collect HMs (Weng et al., 2012; Weng et al., 2014). *Kandeliaobovata* is a woody plant found in tidal salt wetlands in tropical and subtropical regions from East to Southeast Asia (Sheue et al., 2003; Hu et al., 2020). *Kandelia obovata* adapts to transitional ecosystems characterized by the interface between land and water by effectively coping with periodic and aperiodic tidal influences. These tidal impacts give rise to elevated salinity levels, substantial erosion, and anaerobic conditions (Giri et al., 2011; Hu et al., 2020).

Copper is a protein cofactor that participates in electron transfer activities, hence playing a crucial role in these biochemical processes. Additionally, Cu is an essential component for plants. Cu fulfills several functions in biological processes, such as photosynthesis, respiration, ethylene sensing, reactive oxygen metabolism, and cell wall remodeling (Burkhead et al., 2009). Cu is a vital plant element, but excessive quantities can lead to phytotoxicity. Soils can get contaminated by Cu and other HMs due to human activities, such as industrial, mining, and agricultural practices. These activities include sewage sludges, organic wastes,

fertilizers, and fungicides. Therefore, it is crucial to investigate the correlation between HMs and plants, given their impact on crop yield and plant development (Mourato and Martins, 1900; Fageria, 2001).

As far as current information is concerned, the presence of the NRAMP gene family in *Kandelia obovata* has not been documented. Therefore, this study is the first instance in which a genome-wide analysis was conducted to identify *NRAMP* genes inside the genome of *Kandelia obovata*. This study aimed to identify and characterize five *NRAMP* genes and then investigated their expression levels in response to five $CuCl_2$ treatments. In order to gain a deeper understanding of the evolution of *NRMAP* genes in *Kandelia obovata*, various aspects were examined using bioinformatics methodologies. These included the analysis of gene structures, physicochemical properties, chromosomal distribution, synteny and duplication structures, conserved motifs, cis-elements, phylogenetic relationships, subcellular localization, and the expression profiles of *NRAMP* homologs. The present study investigates the characterization and expression analysis of the *NRAMP* family in *Kandelia obovata*. This research seeks to establish a theoretical foundation for future investigations on the response of the *NRAMP* family to $CuCl_2$ treatment in *Kandelia obovata* plants.

# 2 Materials and methods

## 2.1 Characterization and identification of the *NRAMP* genes in *Kandelia obovata*

For *Kandelia obovata*, the genome sequences were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/; BioProject/GWH, Accession codes: PRJCA002330/GWHACBH00000000) and the *Kandelia obovata* protein database (https://www.omicsclass.com/article/310) (Hu et al., 2020). Two databases confirm hypothetical proteins: Pfam (http://pfam.xfam.org/) and NCBI CDD (https://www.omicsclass.com/article/310, E-value $1.2 \times 10^{-28}$). The protein sequence analysis of NRAMP linked with the domain profile was conducted utilizing the Pfam database (http://pfam.xfam.org). The *Kandelia obovata* genome database (https://www.omicsclass.com/article/310) and the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) were employed to identify and validate five NRAMP family genes, namely *KoNRAMP1*, *KoNRAMP2*, *KoNRAMP3*, *KoNRAMP4*, and *KoNRAMP5*. Protparam (http://web.expasy.org/protparam/) was utilized to obtain physicochemical properties (Si et al., 2023).

## 2.2 Chromosomal distribution

Using the NCBI database and https://www.omicsclass.com/article/310, the genomic locations and protein sequences of every *NRAMP* gene in *Kandelia obovata* were identified. The distribution positions of *NRAMP* genes on chromosomes were also evaluated. The MapGene2Chromosome (MG2C) tool was utilized to determine the chromosomal locations of *NRAMP* genes in *Kandelia obovata*. The MG2C tool, version 2.0, was used at the URL: http://mg2c.iask.in/mg2c (Hussain et al., 2022a).

## 2.3 Phylogenetic tree construction

Protein sequences of *NRAMP* genes from the following species were used in the phylogenetic analysis: *Vitis vinifera* (Vv), *Aegiceras corniculatum* (Ac), *Solanum tuberosum* (St), *Populus trichocarpa* (Pt), *Kandelia obovata* (Ko), and *Arabidopsis thaliana* (At). The MEGA11 (V 6.06) software, available at www.megasoftware.net, was commonly employed to align protein sequences (Hashemipetroudi et al., 2023). The phylogenetic tree was constructed using the neighbor-joining (NJ) method, using 1000 bootstrap replicates. Using Fig Tree V1.4.4, the phylogenetic tree was examined and modified (Hussain et al., 2023).

## 2.4 Gene structure and significant motif analyses

The genome of *Kandelia obovata* has been revealed to include five genes belonging to the NRAMP family. The exon/intron arrangements of the five *NRAMP* genes were exhibited along with the structural analyses of the genes via web software (http://gsds.cbi.pku.edu.cn) (Shang et al., 2023). More conserved strings or sections were found in the protein sequences of the five NRAMP proteins, according to the online tool MEME v5.4.1, which can be accessed at https://meme-suite.org/meme/tools/glam2scan (Shang et al., 2023). The application employed the following settings: sequencing of alphabet DNA, RNA, or protein; site distribution limited to zero or one occurrence per sequence (zoops); motif finding mode set to classic mode; and 10 motifs. Using the TBtools program, the MEME findings were displayed following downloading the relevant mast file (Hussain et al., 2022b).

## 2.5 Synteny and duplication analysis

The Minspan (Available online) generated synteny associations of *NRAMP* genes from the following species: *Vitis vinifera* (Vv), *Aegiceras corniculatum* (Ac), *Solanum tuberosum* (St), *Populus trichocarpa* (Pt), and *Kandelia obovata* (Ko). With the KaKs Calculator 2.0 (https://sourceforge.net/projects/kakscalculator2/), the synonymous (Ks), non-synonymous (Ka), and Ka/Ks ratios were computed in order to examine the evolutionary constraints of each pair of *NRAMP* genes (Hussain et al., 2023).

## 2.6 Cis-regulatory elements

The NRAMP family members' 2,500 bp upstream sequences were gathered using the *Kandelia obovata* genome assembly database. The PlantCARE tool (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was employed to identify CREs from the

obtained sequences. TBtools was used to generate the most prevalentCREs identified for the *NRAMP* genes, following a comprehensive analysis of the frequency of each CRE motif (Yang et al., 2022; Si et al., 2023; Yaghobi and Heidari, 2023).

## 2.7 3D structure and subcellular localization

The three-dimensional (3D) structure can be estimated using SWISS-MODEL (https://swissmodel.expasy.org/interactive) (Shang et al., 2023). The subcellular localization of the NRAMP family genes was predicted using two online tools.

ProtComp 9.0, available at http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc, is a software tool used (Si et al., 2023).

The CELLO server, which can be visited at cello.life.nctu.edu.tw (Hussain et al., 2023; Sajjad et al., 2023).

## 2.8 Plant material and environmental conditions

In the experiments, one-year-old *Kandelia obovata* seedlings were planted in the mangrove conservation site (109.22° N, 21.42° E) of Golden Bay Mangrove Reserve in Beihai, Guangxi Province. The soil was irrigated with $CuCl_2$ irrigation every six months and was watered with the nearby seawater every morning and evening as part of semi-natural cultivation techniques. Five different concentrations of $CuCl_2$—0, 50, 100, 200, and 400 mg/L—were utilized for the Cu0, Cu50, Cu100, Cu200, and Cu400 treatments, respectively, throughout two years. Cu0 mg/L was used as the starting concentration for the control treatment, which only used local seawater. The soil sample used for this experiment had a background Cu concentration of < 1.0%, which was classified as non-polluted (Liao et al., 2023). After two years of treatment, plant samples were gathered to evaluate the parameters (Hussain et al., 2023).

## 2.9 Quantitative real-time PCR assays

Total RNA extraction from the leaves was performed using TRIzol (Invitrogen, http://www.invitrogen.com). The ABI PRISM 7500 Real-time PCR Systems, manufactured by Applied Biosystems, were utilized in this study to perform quantitative real-time PCR (qRT-PCR) assays. The $2^{-\Delta\Delta CT}$ approach, as previously outlined (Hussain et al., 2022c; Sun et al., 2022), was employed for data analysis. The actin gene of *Kandelia obovata* (KoActin) was used as a reference gene using the sequence given by Sun et al., 2022 (forward primer: CAATGCAGCAGTTGAAGGAA, reverse primer: CTGCTGGAAGGAACCAAGAG). The *KoNRAMPs* gene primers utilized in real-time PCR are listed in Table 1 and were created with the PRIMER 5.0 software (http://www.premierbiosoft.com) (Hussain et al., 2023).

TABLE 1 The primers utilized in this study's qRT-PCR gene expression investigation.

| Gene name | Primer name | Sequence (5′ to 3′) |
|---|---|---|
| *KoNRAMP1* | Forward | AATTGATCGTGTGGCCGGAA |
| | Reverse | CCAGACACACTCTTCCCCAC |
| *KoNRAMP2* | Forward | GCATGGATGGTGGACAGTGA |
| | Reverse | TTTCTCGCAACTCCCCGTAG |
| *KoNRAMP3* | Forward | AGACTCAAACCCTGTGGCTG |
| | Reverse | CTGGTCGGCTTTCCATGAGT |
| *KoNRAMP4* | Forward | CTCCCAATCCCGACAAGAGC |
| | Reverse | TTCATCTGTGTCCCCTCCCT |
| *KoNRAMP5* | Forward | GGGAGCCAGCAACCTAATGA |
| | Reverse | GGTCCGGCTACAAGCACTAA |

## 2.10 Statistical analysis

The data was analyzed using Statistix 8.1, an analytical software developed in Tallahassee, FL, USA. A one-way ANOVA was employed to evaluate the data, and the results were presented as the mean and standard deviation (SD) of the three replicates. Five distinct Cu stress plants (Cu0, Cu50, Cu100, Cu200, and Cu400 mg/L) were compared for differences in leaf mean values using an LSD (least significant difference) test at $p < 0.05$ (Khan et al., 2022). The statistical tool GraphPad Prism version 9.0.0 for Windows, developed by GraphPad Software, located in San Diego, California, USA (https://www.graphpad.com), was used to construct the graphs (Asim et al., 2022).

# 3 Results

## 3.1 Identification of NRAMP family members in *Kandelia obovata*

The total number of *NRAMP* genes found in the genome of *Kandelia obovata* is five, which is significantly more than the number of *NRAMPs* found in other plant species such as *Arabidopsis thaliana*, peanut (*Arachis hypogaea* L.), common bean (*Phaseolus vulgaris*), potato (*Solanum tuberosum*), soybean (*Glycine Max* L.), *Spirodela polyrhiza*, and tea plant (*Camellia sinensis*). The molecular weight of the NRAMP family exhibited ranging from 49.4 to 139, with an average value of 84.12 kilodaltons (kDa). The protein KoNRAMP5 exhibited the highest isoelectric point (pI) of 7.61, while *KoNRAMP* displayed the lowest pI of 5.01. The isoelectric point (pI) of the NRAMP family exhibited an average range of 5.01 to 7.61. The hydrophobic nature of five *NRAMPs* was demonstrated by the range of their grand average hydropathy index (GRAVY) values, which varied from -0.082 to 0.608. Identifying the subcellular localization of NRAMP proteins can enhance the accessibility of comprehending their molecular activity. Five *NRAMPs* were probably located in the plasma membrane, based on the subcellular localization prediction of

NRAMP proteins (Table 2). The NRAMP family exhibited an average amino acid length of 452 to 1272, an aliphatic index of 89.87 to 120.7, and an instability range of 30.07 to 48.94. The MapChart web tool was utilized to map the genomic chromosomal distribution of the discovered *NRAMP* genes in *Kandelia obovata* by utilizing their respective chromosomal positions and assigning them to the corresponding chromosomes. The chromosomal locations of the five *NRAMP* genes, namely *KoNRAMP1*, *KoNRAMP2*, *KoNRAMP3*, *KoNRAMP4*, and *KoNRAMP5*, were identified as chromosome 01 (Chr01), chromosome 02 (Chr02), chromosome 05 (Chr05), chromosome 07 (Chr07), and chromosome 18 (Chr18) (Figure 1; Table 2).

## 3.2 Variation across the NRAMP family in terms of domain and 3D structure

Aligning all *KoNRAMPs*, BLAST in NCBI revealed that their amino acid sequences are 55–75% similar (Figure 2). *NRAMPs* were discovered in three transmembrane domains (TMD) of *Kandelia obovata*, according to domain analysis (PF01566) (Figure 2). The protein structures of *KoNRAMPs* were confirmed using the SWISS-MODEL workspace and SOSUI tool, as depicted in Figure 3. The significant conservation seen in all investigated species suggests that the crystal structure of a divalent metal transporter. The NRAMP proteins of *Kandelia obovata* exhibit secondary structures that are analogous to NRAMP proteins seen in other species, as depicted in Figure 3. These proteins also demonstrate a sequence identity and similarity ranging from 55% to 75%.

## 3.3 NRAMP protein phylogenetic relationships

An unrooted NJ tree was constructed by aligning five KoNRAMP, six AtNRAMP, five StNRAMP, twelve PtNRAMP, eight VvNRAMP, and eight AcNRAMP from *Kandelia obovata* (Ko), *Arabidopsis thaliana* (At), *Solanum tuberosum* (St), *Populus trichocarpa* (Pt), *Vitis vinifera* (Vv), and *Aegiceras corniculatum* (Ac). This allowed for the phylogenetic relationships among the NRAMP proteins from these five species. Three subgroups (I, II, and SLC) of NRAMP proteins could be identified based on the phylogenetic tree: Subgroup I consisted of 18 NRAMP proteins, which were classified into several species-specific groups. Specifically, this subgroup comprised two KoNRAMP proteins (KoNRAMP2/4), four AtNRAMP proteins (AtNRAMP2/4/5/7), three StNRAMP proteins (StNRAMP1/2/3), three PtNRAMP proteins (PtNRAMP4/9/10), two VvNRAMP proteins (VvNRAMP1/7), and four AcNRAMP proteins (AcNRAMP10/13/16/18). Subgroup II comprised of a total of 26 NRAMP proteins, including one KoNRAMP protein (KoNRAMP5), three AtNRAMP proteins (AtNRAMP1/3/6), one StNRAMP protein (StNRAMP5), four PtNRAMP proteins (PtNRAMP1/2/3/6), five VvNRAMP proteins (VvNRAMP3/4/5/6/8), and 12 AcNRAMP proteins (AcNRAMP1/2/3/4/5/6/7/8/9/12/14/17). The subgroup SLC comprised a total of eleven NRAMP proteins, including two KoNRAMP proteins (KoNRAMP1/3), one StNRAMP protein (StNRAMP4), five PtNRAMP proteins (PtNRAMP5/7/8/11/12), one VvNRAMP protein (VvNRAMP2), and two AcNRAMP proteins (AcNRAMP11/15). Hence, it can be observed from Figure 4 that Subgroup I and SLC had a higher number of NRAMP members compared to Subgroup II.

## 3.4 *NRAMP* genes structure and conserved motifs investigation

A phylogenetic tree was constructed utilizing the individual sequences of the NRAMP protein. The NRAMP proteins were categorized into three categories: I, II, and SLC. This classification scheme aligns with the subsequent evolutionary groups that are elaborated upon. The homologous gene pairs exhibiting significant

TABLE 2 Comprehensive details of the NRAMP gene family found in *Kandelia obovata*.

| Name | Gene ID | Location | AA[1] | Chains[2] | MW[3]/ kDd | pI[4] | GRAVY[5] | Aliphatic Index | Instability | Subcellular localization |
|---|---|---|---|---|---|---|---|---|---|---|
| KoNRAMP1 | geneMaker00017339 | Chr1 9445760-9454057 | 1272 | + | 139 | 5.49 | 0.022 | 94.54 | 42 | Plasma membrane |
| KoNRAMP2 | geneMaker00008219 | Chr2 2033753-2037135 | 516 | + | 56.6 | 5.01 | 0.608 | 117.8 | 30.7 | Plasma membrane |
| KoNRAMP3 | geneMaker00005552 | Chr5 2870003-2878317 | 1097 | – | 119 | 6.11 | -0.082 | 89.87 | 48.94 | Plasma membrane |
| KoNRAMP4 | geneMaker00003484 | Chr7 3102352-3106139 | 452 | – | 49.4 | 5.28 | 0.457 | 113.1 | 40.59 | Plasma membrane |
| KoNRAMP5 | geneMaker00018842 | Chr18 3057367-3065595 | 523 | – | 56.6 | 7.61 | 0.602 | 120.7 | 30.07 | Plasma membrane |

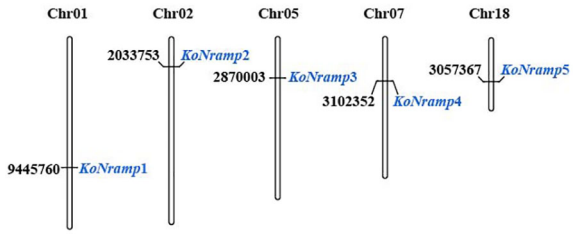



FIGURE 1
The schematic representation displays the distribution of the *NRAMP* gene on the five chromosomes of *Kandelia obovata*. The gene's name is indicated in blue on the right side. The chromosomal regions where the *NRAMP* genes are located are denoted by black letters. The chromosomal numbers can be observed at the uppermost region of each chromosome. The chromosomal numbers can be found at the uppermost region of each chromosome (Chr).

sequence similarities demonstrated strong evolutionary connections and similar exon-intron organizations, as illustrated in Figure 5. In order to further our understanding of the evolutionary dynamics of the NRAMP gene family, an investigation was conducted on the exon-intron arrangements of the *NRAMP* genes. Subgroup I consists of two constituents from the *KoNRAMPs* group, accounting for two out of four members. The presence of four exons characterizes each constituent within this subgroup. It is worth noting that the *KoNRAMP* constituent comprises four introns, while the *KoNRAMP2* constituent contains three introns.

Similarly, just one *KoNRAMP5* gene with 13–14 exons and 12–13 introns is present in subgroup II. Within the subgroup SLC, it has been observed that most *NRAMP* genes had an elongated coding sequence (CDS), accompanied by five comparatively shorter CDSs within the SLC subtribe. Furthermore, these genes have a remarkably conserved structure. Most genes in each of the three categories contained a surprising number of CDSs and introns, suggesting their activities might be very similar.

The protein sequences of *NRAMPs* were analyzed, identifying 10 conserved motifs (Figure 5). The conserved motifs seen in all *NRAMP* genes exhibited a range of two to eight; similarly, the *KoNRAMP* genes displayed motifs ranging from two to eight. The KoNRAMP3 protein has two distinct motifs at positions 6 and 9, whereas KoNRAMP1 and KoNRAMP4 proteins possess five motifs. The protein KoNRAMP5 exhibited a total of six distinct motifs, whereas KoNRAMP2 displayed a larger repertoire with a total of eight motifs. Most KoNRAMP proteins are anticipated to possess motifs one and three, except for KoNRAMP3.

Conversely, motif 4 has been detected in KoNRAMP1/2/5 proteins. Motifs 1 and 3 were identified in all genes except for KoNRAMP3 and StNramp4, while motif 2 was identified in all genes except KoNRAMP1/3 and StNRAMP4. Similarly, motif 4 was observed in all genes except for KoNRAMP3/4.

## 3.5 Cis-regulatory elements in the promoters of five *NRAMP* genes

After looking at the cis-element research, which may provide insights into regulatory gene expression pathways, we looked at the 2500-bp upstream promoter sequences of *NRAMP* genes. Concerning the *NRAMP* genes, it is noteworthy that the light-responsiveness gene exhibits the greatest abundance of cis-elements, with a total of 91. Furthermore, the promoter sequences of *NRAMP* genes were found to contain 53 cis-elements associated with phytohormones, including Abscisic acid (18), Gibberellin (8), Auxin (8), Salicylic acid (3), and MeJA (16). Additionally, three cis-elements related to defense and stress responses, namely low temperature (4), drought (7), and Endosperm expression (2), as well as two cis-elements associated with Zein metabolism (2), one

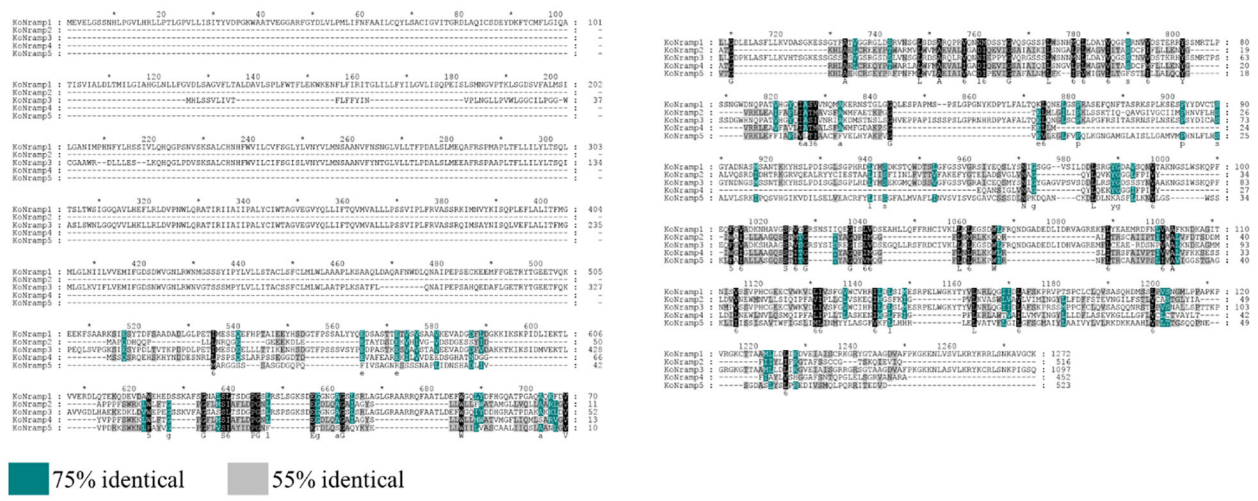



FIGURE 2
Multiple alignments of the amino acid sequence were performed utilizing information from each *KoNRAMP* transporter. The predicted transmembrane regions are displayed as black boxes. The symbols "*" that appear above the sequence indicate every 10 residues of amino acids.

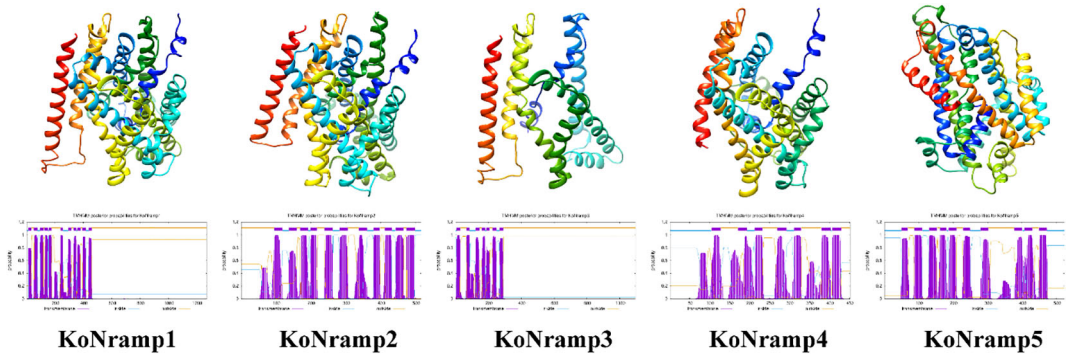



FIGURE 3
The 3D and transmembrane structures of *KoNRAMP*. The SWISS-MODEL software is utilized for the prediction of three-dimensional structural homology models. The SOSUI program validated the transmembrane structures.

cis-element linked to Cell cycle (1), and nine cis-elements involved in Anaerobic responses were also identified (Figure 6). The observed variability in the response components provides evidence for the regulatory roles of *NRAMP* genes in a wide range of physiological and biological processes.

## 3.6 Synteny and duplication analysis of NRAMP gene family

A collinearity analysis revealed significant orthologs of the *NRAMP* genes in *Kandelia obovata* and the other six inherited plant species: *Arabidopsis thaliana*, *Solanum tuberosum*, *Populus trichocarpa*, *Vitis vinifera*, and *Aegiceras corniculatum* (Figure 7). In chromosome 01, a gene from *Kandelia obovata* had syntenic associations with two genes from *Populus trichocarpa*, as well as with one gene each from *Arabidopsis thaliana*, *Solanum tuberosum*, *Vitis vinifera*, and *Aegiceras corniculatum*. In contrast, it is seen that within chromosome 02, a specific gene of *Kandelia obovata* exhibits syntenic relationships with two genes of *Arabidopsis thaliana* and *Vitis vinifera*. This gene demonstrates a syntenic link with one gene each from *Solanum tuberosum*, *Populus trichocarpa*, and *Aegiceras corniculatum*.

Similar syntenic relationships were seen in chromosome 18, where one gene from *Kandelia obovata* showed up with two genes from *Arabidopsis thaliana*, one gene from *Populus trichocarpa*, and

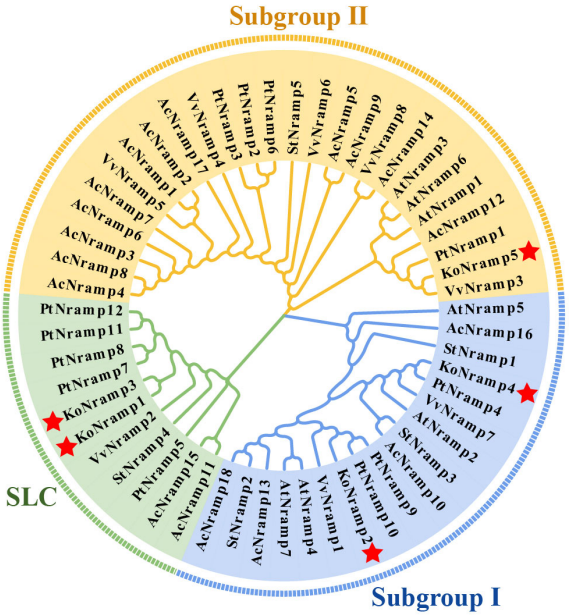



FIGURE 4
The most significant likelihood technique was employed to conduct a phylogenetic analysis of NRAMP proteins derived from *Kandelia obovata* (Ko), *Arabidopsis thaliana* (At), *Solanum tuberosum* (St), *Populus trichocarpa* (Pt), *Vitis vinifera* (Vv), and *Aegiceras corniculatum* (Ac). Subgroup I, Subgroup II, and SLC are the three groups of NRAMP proteins; a distinct color denotes each. The red asterisk symbol (*) was used to represent the *Kandelia obovata* NRAMP proteins.

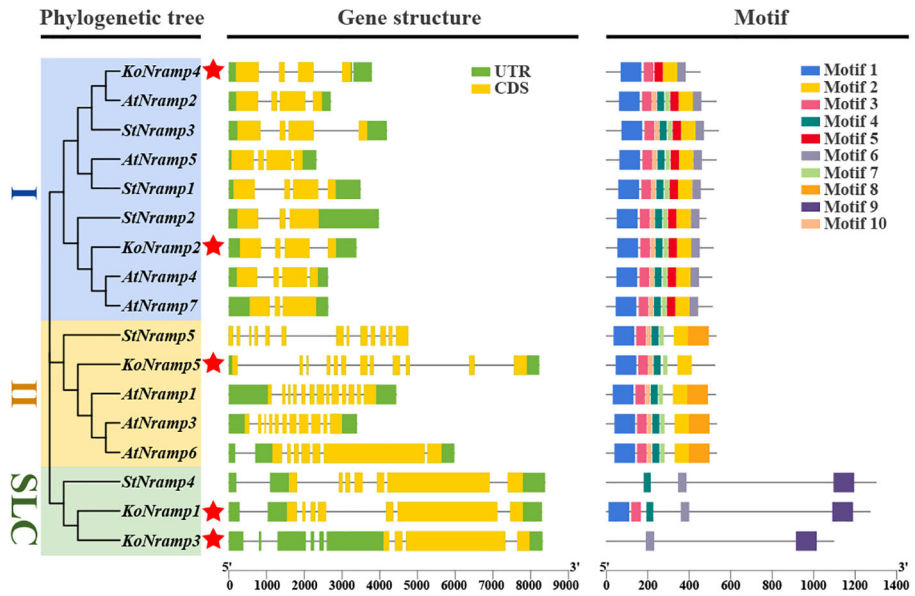



FIGURE 5
*Kandelia obovata's NRAMP* family genes were examined for their gene structure and motif makeup. The *NRAMP* genes of both genomes were classified into two distinct categories according to their evolutionary ties—the gene structure of the *NRAMPs*. The color green visually represents the UTR regions, while the CDS or exons are depicted in yellow. A black horizontal line indicates the presence of introns. Additionally, the conserved motif structures identified in the *NRAMPs* are denoted by the letter (Ko). Various colored boxes display distinct motifs. The red asterisk symbol (*) was used to represent the *Kandelia obovata* NRAMP proteins.

one gene from *Aegiceras corniculatum* (Figure 7). It is worth mentioning that multiple homologs of *Kandelia obovata* (referred to as *KoNRAMPs*) have demonstrated their ability to persistently coexist with *Arabidopsis thaliana*, *Solanum tuberosum*, *Populus trichocarpa*, *Vitis vinifera*, and *Aegiceras corniculatum* through a syntenic association. This finding suggests that both segmental repetition and whole-genome duplication played significant roles in the evolutionary development of the KoNRAMPs gene family.

Segmental and tandem duplication promotes the emergence of new gene families and plant genomes. In order to gain a deeper comprehension of the duplication events of the *Kandelia obovata NRAMP* gene, an examination was conducted on both segmental and tandem duplications within the KoNRAMP gene family. The chromosomal dispersals of five *KoNRAMP* genes were evaluated. The findings of this study indicate that a single segmental duplication event involving the *KoNRAMP2* and *KoNRAMP4* gene pairs of chromosomes A02 and A07 was observed, as depicted in Figure 8. It is worth noting that the *KoNRAMP* gene was not present in the remaining chromosomes. Tandem repeats paralogous genes were not identified in regions A01, A05, and A18,

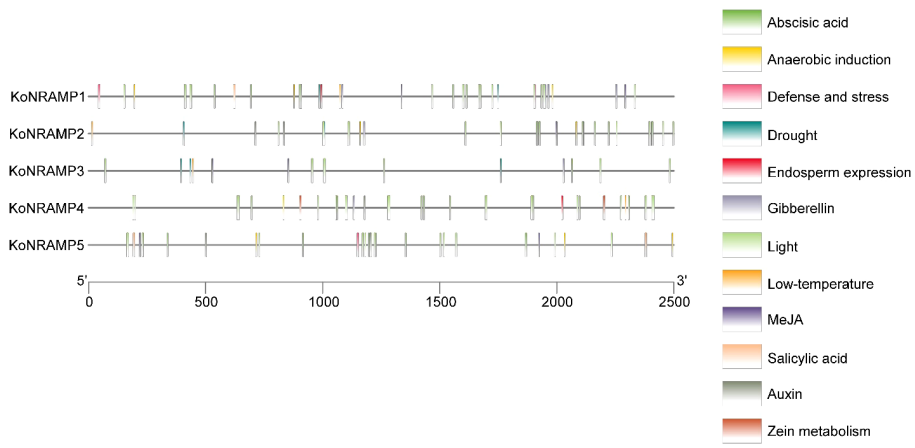



FIGURE 6
The promoters of the *KoNRAMP* gene have been found to include regulatory elements referred to as CREs. The positional distribution of the predicted CREs on the *KoNRAMP* promoters is shown by vertical bars. The promoter sequences (2500 bp) of five *KoNRAMP* genes were analyzed using PlantCARE. In the context of this legend, distinct cis-elements were symbolically represented by various colors.

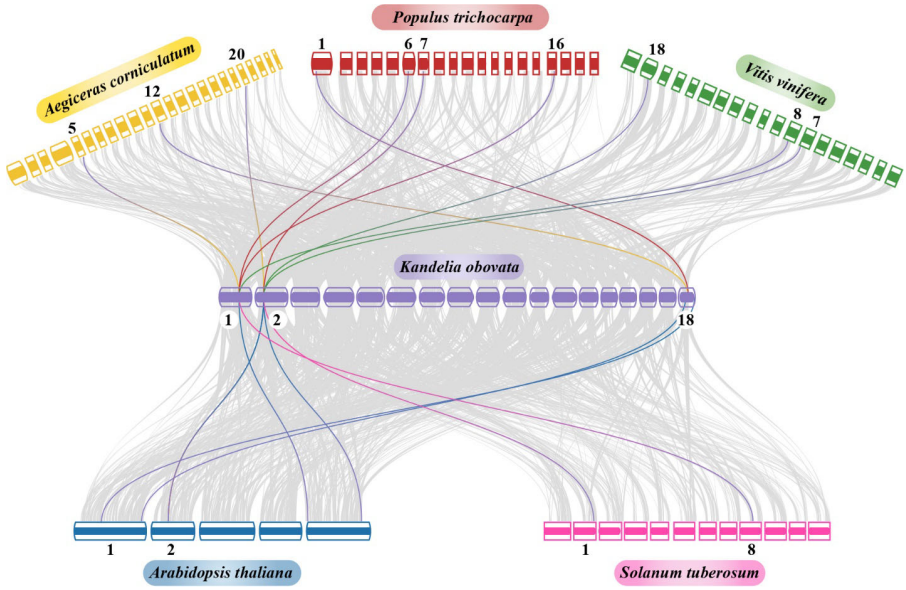



FIGURE 7
Synteny analysis of *NRAMP* genes in the chromosomes of *Aegiceras corniculatum*, *Vitis vinifera*, *Populus trichocarpa*, *Solanum tuberosum*, *Arabidopsis thaliana*, and *Kandelia obovata* was examined. The grey lines in the background show the syntenic NRAMP gene pairs, and the collinear blocks in the genomes of *Kandelia obovata* and the other five plant species are highlighted by the red lines. The box's distinct colors represented plant species. The chromosomal number of each species is indicated at the upper and lower ends of each chromosome.

each with a solitary gene irrespective of the chromosome. The results of this study have illustrated the importance of duplication events in the expansion of the *KoNRAMP* family genes.

In order to further the understanding of the evolutionary constraints on the KoNRAMP gene family, an analysis was conducted to determine the values of Ka (non-synonymous substitution rate), Ks (synonymous substitution rate), and the Ka/Ks ratio for *Kandelia obovata*. The duplicated *KoNRAMP* gene pairs displayed a Ka/Ks ratio of 0.86, indicating that the *NRAMP* family genes in *Kandelia obovata* may have experienced selection pressure or a discriminatory load during their evolutionary development (Table 3).

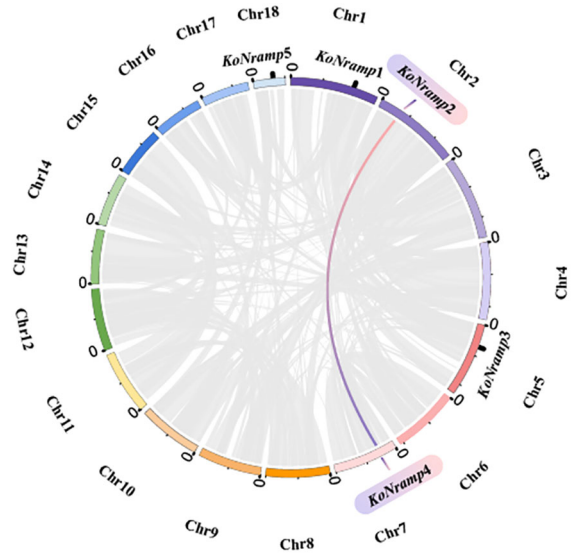



FIGURE 8
Circles represent the distribution of the *KoNRAMP* gene chromosomes and the interactions between chromosomes. The red and blue lines represent the syntenic NRAMP gene pair, whereas the syntenic blocks in the *Kandelia obovata* genome are depicted by the grey lines in the background.

### 3.7 Expression analysis of *KoNRAMP* genes under $CuCl_2$ treatment

The expression profiling of five *KoNRAMP* genes (*KoNRAMP1*, *KoNRAMP2*, *KoNRAMP3*, *KoNRAMP4*, and *KoNRAMP5*) was conducted using qRT-PCR. This analysis was carried out in the presence of five distinct $CuCl_2$ treatments, namely Cu0, Cu50, Cu100, Cu200, and Cu400 mg/L, as depicted in Figure 9. In the present investigation, the expression levels of *KoNRAMP1*, *KoNRAMP4*, and *KoNRAMP5* were significantly reduced across all five $CuCl_2$ treatments compared to the control group (Cu0). The expression levels of all five genes were shown to be significantly down-regulated in the Cu400 condition compared to the Cu0 condition. This observation can be attributed to the increased *KoNRAMPs* expression levels in the leaf when Cu availability was limited. However, when Cu was abundant (Cu400), the transcript levels of these genes decreased in comparison to both the Cu0 condition and other levels of Cu. The expression levels of *KoNRAMP2* and *KoNRAMP3* genes did not show statistical significance when comparing Cu50 with the control group. However, these genes exhibited statistically significant up-regulation in the Cu100 and Cu200 groups compared to the Cu0 group.

## 4 Discussion

Numerous *NRAMP* genes have been found and functionally described as a result of their significant role in the absorption and transport of metal ions (Thomine et al., 2003; Cailliatte et al., 2010; Sasaki et al., 2012; Peris-Peris et al., 2017; Chang et al., 2020). However, *Kandelia obovata* does not have comprehensive information about this family. The genome of *Kandelia obovata* was used in this investigation to identify five *NRAMP* genes that are relevant to reported plant species, including *Arabidopsis thaliana* (six) (Segond et al., 2009), *Camellia sinensis* (11) (Li et al., 2021), *Spirodela polyrhiza* (three) (Chen et al., 2021), *Glycine Max* L. (13) (Qin et al., 2017), *Solanum tuberosum* (five) (Tian et al., 2021), *Populus trichocarpa* (11) (Ma et al., 2023), *Phaseolus vulgaris* L. (seven) (Ishida et al., 2018), *Arachis hypogaea* L. (15) (Tan et al., 2023). Based on the findings of previously published investigations, it has been observed that numerous plant species possess a range of three to fifteen *NRAMP* genes. Our research outcome aligns with a prior investigation in this regard.

Fifteen (15) *NRAMP* family genes were found in the genome of *Arachis hypogaea* and subsequently designated as *AhNRAMP1-AhNRAMP15*. The upregulation of the majority of *AhNRAMPs* is triggered by iron deficiency in the roots of *Arachis hypogaea*, and there is a positive correlation between their expression and the accumulation of cadmium. This suggests that *AhNRAMPs* play a significant role in the interactions between iron and cadmium in *Arachis hypogaea* (Tan et al., 2023). Thirteen (13) *NRAMP* family genes have been discovered from *Glycine Max* L. Among these genes, *GmNRAMP1-13* exhibit distinct regulatory patterns in response to shortages in nitrogen, phosphorus, potassium, Fe, and sulfur, as well as toxicities caused by excessive amounts of Fe, Cu, Cd, and Mn. According to the results of the expression study, it has been indicated that *GmNRAMP* genes exhibit functionality across many tissues and developmental stages (Qin et al., 2017).

A comprehensive set of 11 *NRAMP* family genes was discovered in *Populus trichocarpa* and *Camellia sinensis* genomes. These genes were subsequently designated as *PtNRAMP1-PtNRAMP11* and *CsNRAMP1-CsNRAMP11*. According to the findings of gene expression research, it was observed that the *PtNRAMP* genes exhibited varying responses to metal stress conditions, such as deficiencies in Fe and Mn, as well as toxicities caused by Fe, Mn, Zn, and Cd (Ma et al., 2023). To evaluate the effects of Pb treatment on roots and leaves, the expression levels of *CsNRAMPs* were found (Li et al., 2021). A comprehensive analysis has shown the identification of *NRAMP* family genes in *Phaseolus vulgaris* L., *Arabidopsis thaliana*, *Solanum tuberosum*, and *Spirodela polyrhiza*. Specifically, 7, 6, 5, and 3 *NRAMP* family genes have been identified from these respective species. Notably, the products *PvNRAMP1-7* have been found to possess transporters involved in metal homeostasis in *Phaseolus vulgaris* L. (Ishida et al., 2018). The induction of *StNRAMPs* expression appears to be unique to HMs, suggesting a potential involvement in the transportation and uptake of HMs such as (Cu, Cd, Zn, nickel (Ni), and Pb (Tian et al., 2021). The SpNRAMP gene's expression level was affected by Cd stress, particularly in the context of Fe or Mn shortage (Chen et al., 2021).

The NRAMP proteins were grouped into three subgroups by phylogenetic analysis, which aligns with recent research that made a similar division among seven *P. vulgaris* NRAMP proteins (Ishida et al., 2018). A comprehensie set of eleven NRAMP members (*PtNRAMP1–11*) were discovered in *Populus trichocarpa*, a widely used woody model plant. These members were further categorized into three groups through phylogenetic analysis (Ma et al., 2023). Additional investigation of the *PtNRAMP* gene structure revealed that members of closely related subgroups typically exhibited comparable exon-intron arrangements, suggesting a shared evolutionary lineage. Additionally, it was shown that the *PtNRAMP* genes in the various subgroups had distinct exon-intron architectures. These modest variations might significantly impact the evolutionary process of these genes. Hence, there is speculation on the involvement of distinct gene architectures in disparate tasks

TABLE 3 Comprehensive data regarding the Ka, Ks, and Ka/Ks ratio in *Kandelia obovata*.

| Name | Method | *Ka* | *Ks* | *Ka/Ks* | Divergence-time (MYB) | Duplicated type |
|---|---|---|---|---|---|---|
| *KoNRAMP2*, *KoNRAMP4* | MS | 0.06 | 0.07 | 0.86 | 2.2 | Segmental |

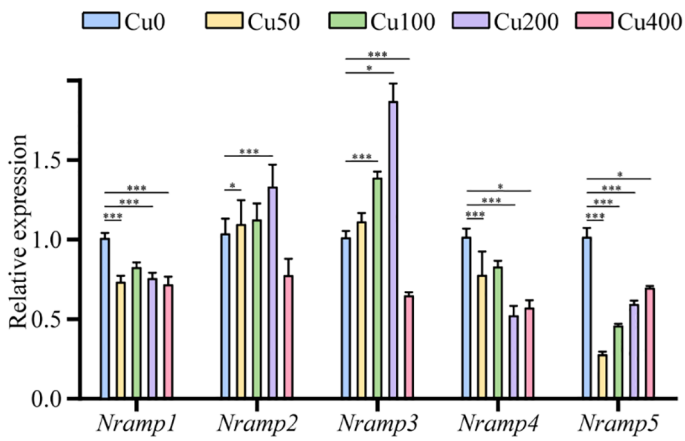



FIGURE 9
Expression of *KoNRAMPs* in leaves of seedling-stage *Kandelia obovata* plants under different Cu stress conditions (Cu0, Cu50, Cu100, Cu200, and Cu400 mg/L) as determined by qRT-PCR analysis. There is a statistically significant difference ($p < 0.05$) seen between the control group and all experimental conditions, as determined by the Least Significant Difference (LSD) test. *$p < 0.05$, *** means $p < 0.001$. The vertical axis shows the relative gene expression, and the horizontal axis shows *NRAMP1–5* genes.

(Ma et al., 2023). The structure of genes has undergone changes during evolutionary processes, leading to an increase in the number of exons/introns. Generally, genes with fewer introns go through the editing process and leave the nucleus sooner (Hashemipetroudi et al., 2023; Yaghobi and Heidari, 2023). Based on this, the different gene structures may affect the expression pattern of *NRAMP* genes.

There was an unequal distribution of the 11 *PtNRAMP* genes over the six chromosomes. Only one *PtNRAMP* gene was found on chromosomes 1, 6, and 16, two on chromosome 7, and three on chromosomes 2 and 5. *PtNRAMP5*, *PtNRAMP6*, and *PtNRAMP7* may have been tandemly duplicated during evolution, as shown by their clustering on chromosome 5 (Ma et al., 2023). The distribution of the five *StNRAMP* genes is observed throughout five chromosomes, specifically Chr00, Chr02, Chr04, Chr09, and Chr1(Tian et al., 2021). Cis-acting elements distinctly regulate gene transcription and expression in plants (Chen et al., 2021). The promoter regions of *NRAMPs* were shown to possess several stress response and hormone response components. Since numerous stressors can affect a gene's expression, *NRAMP* genes may be involved (Chen et al., 2021). Examining cis-acting elements has indicated that *StNRAMP2* possesses more elements responsive to MYB transcription factors. The sensitivity of *StNRAMP1* to the five HMs in the stem tissue was negligible, resulting in no statistically significant variation in relative expression. Furthermore, it has been noted that the gene's 2000 bp upstream region exhibits supplementary elements that are receptive to phytohormones, including methyl jasmonate, gibberellin, salicylic acid, and auxin. Consequently, it is plausible that this gene may have played a role in plant growth and development throughout evolution (Tian et al., 2021). According to studies, many phytohormones, including JA, ABA, and SA, have been linked to the plant's response to various metal stresses (Ma et al., 2023). Additionally, our investigation revealed the presence of several cis-elements linked with HMs in the promoters of the *KoNRAMP* gene. These cis-elements might play a role in the transportation and uptake of metallic substances. The findings of this study align with prior research, demonstrating the presence of multiple cis-acting elements associated with hormone responsiveness within the putative promoter regions of *Spirodela polyrhiza* (Chen et al., 2021), *Solanum tuberosum* (Tian et al., 2021), *Populus trichocarpa* (Ma et al., 2023), *Phaseolus vulgaris* L. (Ishida et al., 2018), and *Arachis hypogaea* L. *NRAMP* genes (Tan et al., 2023).

As several earlier experiments have demonstrated, the entire genome, segmental, tandem, and gene duplication events are necessary for the growth and evolution of gene families (Vision et al., 2000; Grassi et al., 2008; Chen et al., 2021). According to the study conducted by Force et al., 1999 it was found that the likelihood of tandem duplicate events was greater when the complementarity level was higher than segmental duplicates. Furthermore, the duplication events involving tandem and segmental duplications were identified as the primary types of duplication (Kong et al., 2007). According to earlier findings, *Theobroma cacao* exhibited the existence of a single tandem duplicated pair and one pair of segmental duplications (Ullah et al., 2018). *Glycine max* L. showed six duplicated blocks (Qin et al., 2017), while *Oryza sativa* L. demonstrated one syntenic block with the paralogous pair (Wang et al., 2019). Additionally, *A. thaliana* showcased two pairs of segmental duplications (Segond et al., 2009). Nevertheless, the results indicated the absence of tandem duplication in *Kandelia obovata*. Previous investigations have shown similar findings; no tandem duplication was found in *Arachis hypogaea* L. (Tan et al., 2023), and *S. polyrhiza* (Chen et al., 2021). Thus, the number of *KoNRAMP* genes in *Kandelia obovata* may be related to whole-genome duplication during allopolyploid speciation.

In order to get an additional understanding of the transcript levels, we conducted a quantitative real-time polymerase chain reaction (qRT-PCR) to assess the expression of *KoNRAMP* genes

in response to various Cu stress conditions. Our findings show that the expression levels of *KoNRAMPs* varied significantly in response to five distinct $CuCl_2$ treatments, exhibiting upregulation and downregulation. In leaves, the transcript abundance of *KoNRAMP1*, *KoNRAMP4*, and *KoNRAMP5* decreased in response to $CuCl_2$ treatment. The experimental findings revealed that the expression of all *SpNRAMP* gene families was downregulated in response to 50 μM Cd treatments (Chen et al., 2021). The expression patterns of *StNRAMP* exhibited significant variations across different tissues in response to various HM treatments. The expression pattern of *StNRAMP2* is comparable under stress induced by Pb and Cu, as evidenced by the peak in relative expression seen at 24 hours (Tian et al., 2021). As Tian et al. (2021) reported, these findings supported the notion that *NRAMPs* were responsible for the intake and transport of HMs. The results of our study indicate that the expression levels of the *KoNRAMP2* and *KoNRAMP3* genes were significantly increased in response to Cu concentrations of 100 and 200 mg/L. The results of this study were in agreement with prior research. Specifically, the expression levels of *CsNRAMP1*, *CsNRAMP2*, *CsNRAMP9*, and *CsNRAMP10* were shown to be up-regulated in leaves. Additionally, *CsNRAMP2* exhibited a highly responsive reaction to Pb treatment, indicating its heightened sensitivity (Li et al., 2021). Under the stress conditions of Cd and Fe, most *PtNRAMPs* genes exhibited up-regulation in both leaves and roots, except for *PtNRAMP3* and *PtNRAMP8*. In addition, it was shown that *PtNRAMP1*, *PtNRAMP3*, *PtNRAMP6*, *PtNRAMP7*, and *PtNRAMP10* exhibited up-regulation in both leaves and roots in response to Zn and Mn stress, as reported in the study (Ma et al., 2023). The findings of this study align with other research, suggesting that *NRAMPs* play a crucial role in the absorption and transportation of HMs (Tian et al., 2021). In order to enhance the validation of the functionality of the *KoNRAMP* genes, it is important to conduct a series of experiments involving gene overexpression and knockdown.

## 5 Conclusions

This work identified five *KoNRAMP* genes by doing a thorough genome-wide analysis of the NRAMP gene family in *Kandelia obovata*. On the five chromosomes of *Kandelia obovata*, these *KoNRAMP* genes were located. The expression patterns of five *KoNRAMP* genes were validated using qRT-PCR. Comparing the Cu400 condition to the Cu0 condition showed that the expression levels of all five genes were considerably down-regulated. Based on phylogenetic analysis, the *KoNRAMP* genes were divided into three subgroups (subgroup I, II, SLC). To better understand the evolution of the NRAMP gene family in the *Kandelia obovata* genome, various analyses were conducted, including domain and 3D structural variation, gene structure, phylogenetic and synteny, chromosomal distributions, motif analysis, subcellular localization, cis-regulatory elements, and expression profiling against $CuCl_2$ stress treatments (Figure 10). Multiple promoter investigations have revealed the presence of many stress and hormone response elements, which play a crucial role in the response to Cu and other metal-induced stress. In general, this research offers significant insights for future functional investigations about the biological functions of *NRAMP* genes in *Kandelia obovata*.

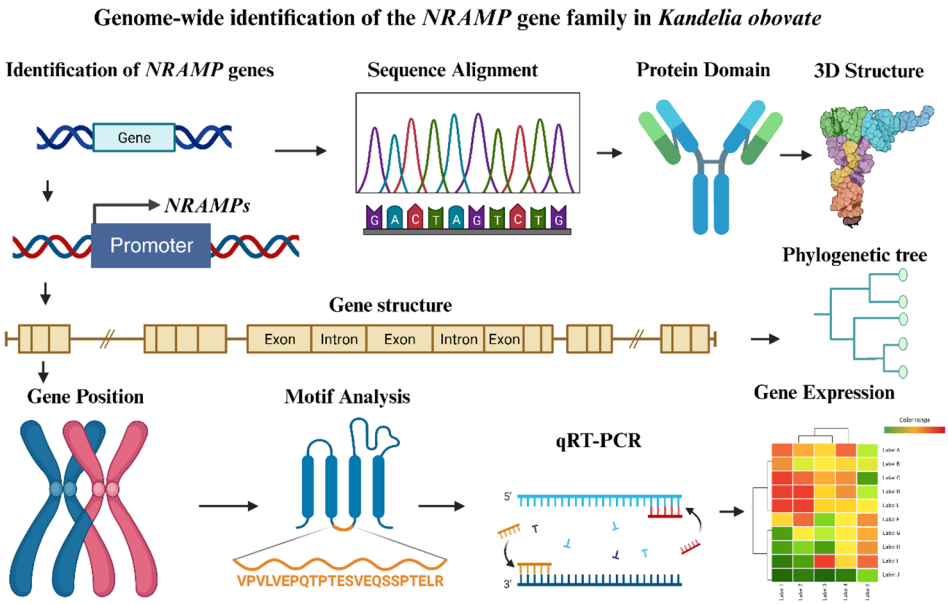



FIGURE 10
An overview of the thorough investigation of the NRAMP gene family in *Kandelia obovata*. The figure was created using the BioRender (https://www.biorender.com/) tool, which was modified from Sajjad et al. (2023). Different colors were used to illustrate the sketches from the various analyses. The arrow symbol represents the several analyses conducted to comprehend the characterization of the *KoNRMAP* genes.

## 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

QH: Conceptualization, Writing – original draft, Writing – review & editing. TY: Formal analysis, Software, Methodology, Writing – review & editing. CS: Funding acquisition, Supervision, Writing – review & editing. SL: Formal analysis, Methodology, Writing – review & editing. AK: Writing – review & editing. JN: Writing – review & editing. A-ZM: Writing – review & editing. ME: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Key R&D Program of China (2018YFA0902500), the National Natural Science Foundation of China (42376219, 41706137), the CAS Key Laboratory of Science and Technology on Operational Oceanography (No. OOST2021-07), the Graduate Education Innovation Project (Guangdong Province, 2020JGXM094).

## Acknowledgments

We thank the Instrument Analysis Center of Shenzhen University. We are also grateful for the support from the public service platform of instruments and equipment at the College of Life Sciences and Oceanography at Shenzhen University. The authors extend their appreciation to the Researchers Supporting Project number (RSPD2023R941), King Saud University, Riyadh, Saudi Arabia.

## Conflict of interest

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

## Publisher's note

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

## References

Asim, M., Hussain, Q., Wang, X., Sun, Y., Liu, H., Khan, R., et al. (2022). Mathematical modeling reveals that sucrose regulates leaf senescence via dynamic sugar signaling pathways. *Int. J. Mol. Sci.* 23, 6498. doi: 10.3390/ijms23126498

Bodin, N., N'Gom-Kâ, R., Kâ, S., Thiaw, O. T., Tito de Morais, L., Le Loc'h, F., et al. (2013). Assessment of trace metal contamination in mangrove ecosystems from Senegal, West Africa. *Chemosphere* 90, 150–157. doi: 10.1016/j.chemosphere.2012.06.019

Burkhead, J. L., Gogolin Reynolds, K. A., Abdel-Ghany, S. E., Cohu, C. M., and Pilon, M. (2009). Copper homeostasis. *New Phytol.* 182, 799–816. doi: 10.1111/j.1469-8137.2009.02846.x

Cailliatte, R., Schikora, A., Briat, J. F., Mari, S., and Curie, C. (2010). High-affinity manganese uptake by the metal transporter nramp1 is essential for Arabidopsis growth in low manganese conditions. *Plant Cell* 22, 904–917. doi: 10.1105/tpc.109.073023

Cellier, K., Belouchi, A., and Gros, P. (1996). Resistance to intracellular infections: Comparative genomic analysis of *Nramp*. *Trends Genet.* 12, 201–204. doi: 10.1016/0168-9525(96)30042-5

Cellier, M., Privé, G., Belouchi, A., Kwan, T., Rodrigues, V., Chia, W., et al. (1995). *Nramp* defines a family of membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* 92, 10089–10093. doi: 10.1073/pnas.92.22.10089

Chang, J. D., Huang, S., Yamaji, N., Zhang, W., Ma, J. F., and Zhao, F. J. (2020). OsNRAMP1 transporter contributes to cadmium and manganese uptake in rice. *Plant Cell Environ.* 43, 2476–2491. doi: 10.1111/pce.13843

Chen, Y., Zhao, X., Li, G., Kumar, S., Sun, Z., Li, Y., et al. (2021). Genome-wide identification of the *Nramp* gene family in *spirodela polyrhiza* and expression analysis under cadmium stress. *Int. J. Mol. Sci.* 22, 6414. doi: 10.3390/ijms22126414

Cheng, S., Tam, N. F. Y., Li, R., Shen, X., Niu, Z., Chai, M., et al. (2017). Temporal variations in physiological responses of *Kandelia obovata* seedlings exposed to multiple heavy metals. *Mar. pollut. Bull.* 124, 1089–1095. doi: 10.1016/j.marpolbul.2017.03.060

Fageria, N. K. (2001). Adequate and toxic levels of copper and manganese in upland rice, common bean, corn, soybean, and wheat grown on an oxisol. *Commun. Soil Sci. Plant Anal.* 32, 1659–1676. doi: 10.1081/CSS-100104220

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545. doi: 10.1093/genetics/151.4.1531

Giri, C., Ochieng, E., Tieszen, L. L., Zhu, Z., Singh, A., Loveland, T., et al. (2011). Status and distribution of mangrove forests of the world using earth observation satellite data. *Glob. Ecol. Biogeogr.* 20, 154–159. doi: 10.1111/j.1466-8238.2010.00584.x

Grassi, A., De Lanave, C., and Saccone, C. (2008). Genome duplication and gene-family evolution : The case of three OXPHOS gene families. *Gene* 421, 1–6. doi: 10.1016/j.gene.2008.05.011

Hashemipetroudi, S. H., Arab, M., Heidari, P., and Kuhlmann, M. (2023). Genome-wide analysis of the laccase (*LAC*) gene family in *Aeluropus littoralis* : A focus on identification, evolution and expression patterns in response to abiotic stresses and ABA treatment. *Front. Plant Sci.* 14, 1112354. doi: 10.3389/fpls.2023.1112354

Hu, M. J., Sun, W. H., Tsai, W. C., Xiang, S., Lai, X. K., Chen, D. Q., et al. (2020). Chromosome-scale assembly of the *Kandelia obovata* genome. *Hortic. Res.* 7, 75. doi: 10.1038/s41438-020-0300-x

Huang, G. Y., and Wang, Y. S. (2010). Physiological and biochemical responses in the leaves of two mangrove plant seedlings (*Kandelia candel* and *Bruguiera gymnorrhiza*) exposed to multiple heavy metals. *J. Hazard. Mater.* 182, 848–854. doi: 10.1016/j.jhazmat.2010.06.121

Hussain, Q., Ye, T., Nkoh, N., Zhou, Q., and Shang, C. (2023). Genome-wide identification and expression analysis of the copper transporter (*COPT/ctr*) gene family in *kandelia obovata*, a typical mangrove plant. *Int. J. Mol. Sci.* 24, 15579. doi: 10.3390/cimb44110381

Hussain, Q., Zhan, J., Liang, H., Wang, X., Liu, G., Shi, J., et al. (2022c). Key genes and mechanisms underlying natural variation of silique length in oilseed rape (*Brassica napus* L .) germplasm. *Crop J.* 10, 617–626. doi: 10.1016/j.cj.2021.08.010

Hussain, Q., Zheng, M., Chang, W., Ashraf, M. F., Khan, R., Asim, M., et al. (2022a). Genome-wide identification and expression analysis of *snRK2* gene family in dormant vegetative buds of *liriodendron chinense* in response to abscisic acid, chilling, and photoperiod. *Genes (Basel).* 13, 1–19. doi: 10.3390/genes13081305

Hussain, Q., Zheng, M., Furqan, M., Khan, R., Yasir, M., Farooq, S., et al. (2022b). Scientia Horticulturae Genome-wide identification , characterization and expression analysis of the ABA receptor *PYL* gene family in response to ABA , photoperiod , and chilling in vegetative buds of Liriodendron chinense. *Sci. Hortic. (Amsterdam).* 303, 111200. doi: 10.1016/j.scienta.2022.111200

Ishida, J. K., Caldas, D. G. G., Oliveira, L. R., Frederici, G. C., and Leite, L. M. P. (2018). and mui, T Genome-wide characterization of the NRAMP gene family in *Phaseolus vulgaris* provides insights into functional implications during common bean development. *S. Genet. Mol. Biol.* 41, 820–833. doi: 10.1590/1678-4685-gmb-2017-0272

Khan, R., Ma, X., Hussain, Q., Asim, M., Iqbal, A., Ren, X., et al. (2022). Application of 2,4-epibrassinolide improves drought tolerance in tobacco through physiological and biochemical mechanisms. *Biology.* 11, 1192. doi: 10.3390/biology11081192

Kong, H., Landherr, L. L., Frohlich, M. W., Leebens-Mack, J., Ma, H., and DePamphilis, C. W. (2007). Patterns of gene duplication in the plant *SKP1* gene family in angiosperms: Evidence for multiple mechanisms of rapid gene birth. *Plant J.* 50, 873–885. doi: 10.1111/j.1365-313X.2007.03097.x

Kravchenko, Y., Kursitys, I., and Kravchenko, D. (2018). Architecture and method of integrating information and knowledge on the basis of the ontological structure. *Adv. Intell. Syst. Comput.* 658, 93–103. doi: 10.1007/978-3-319-67349-3_8

Li, J., Duan, Y., Han, Z., Shang, X., Zhang, K., Zou, Z., et al. (2021). Genome-wide identification and expression analysis of the *NRAMP* family genes in tea plant (*Camellia sinensis*). *Plants* 10, 1055. doi: 10.3390/plants10061055

Li, M. S., and Lee, S. Y. (1997). Mangroves of China: A brief review. *For. Ecol. Manage.* 96, 241–259. doi: 10.1016/S0378-1127(97)00054-6

Li, R., Li, R., Chai, M., Shen, X., Xu, H., and Qiu, G. (2015). Heavy metal contamination and ecological risk in Futian mangrove forest sediment in Shenzhen Bay, South China. *Mar. pollut. Bull.* 101, 448–456. doi: 10.1016/j.marpolbul.2015.09.048

Liao, J., Xu, Y., Zhang, Z., Zeng, L., Qiao, Y., Guo, Z., et al. (2023). Effect of Cu addition on sedimentary bacterial community structure and heavy metal resistance gene abundance in mangrove wetlands. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1157905

Liu, W., Huo, C., He, L., Ji, X., Yu, T., Yuan, J., et al. (2022). The *NtNRAMP1* transporter is involved in cadmium and iron transport in tobacco (*Nicotiana tabacum*). *Plant Physiol. Biochem.* 173, 59–67. doi: 10.1016/j.plaphy.2022.01.024

Ma, X., Yang, H., Bu, Y., Zhang, Y., Sun, N., Wu, X., et al. (2023). Genome-wide identification of the NRAMP gene family in *Populus trichocarpa* and their function as heavy metal transporters. *Ecotoxicol. Environ. Saf.* 261, 115110. doi: 10.1016/j.ecoenv.2023.115110

Mourato, M. P., Martins, L. L., and Cuypers, A. (2009). Effect of copper on antioxidant enzyme activities and mineral nutrition of white lupin plants grown in nutrient solution. *J. Plant Nutr.* 32, 1882–1900. doi: 10.1080/01904160903242375

Nevo, Y., and Nelson, N. (2006). The NRAMP family of metal-ion transporters. *Biochim. Biophys. Acta - Mol. Cell Res.* 1763, 609–620. doi: 10.1016/j.bbamcr.2006.05.007

Peris-Peris, C., Serra-Cardona, A., Śanchez-Sanuy, F., Campo, S., Ariño, J., and Segundo, B. S. (2017). Two *NRAMP6* isoforms function as iron and manganese transporters and contribute to disease resistance in rice. *Mol. Plant-Microbe Interact.* 30, 385–398. doi: 10.1094/MPMI-01-17-0005-R

Qin, L., Han, P., Chen, L., Walk, T. C., Li, Y., Hu, X., et al. (2017). Genome-wide identification and expression analysis of NRAMP family genes in soybean (*Glycine max* L.). *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01436

Qiu, Y. W., Yu, K. F., Zhang, G., and wang, W. X. (2011). Accumulation and partitioning of seven trace metals in mangroves and sediment cores from three estuarine wetlands of Hainan Island, China. *J. Hazard. Mater.* 190, 631–638. doi: 10.1016/j.jhazmat.2011.03.091

Sajjad, M., Ahmad, A., Riaz, M. W., Hussain, Q., Yasir, M., and Lu, M. (2023). Recent genome resequencing paraded *COBRA-Like* gene family roles in abiotic stress and wood formation in Poplar. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1242836

Sasaki, A., Yamaji, N., Yokosho, K., and Ma, J. F. (2012). *Nramp5* is a major transporter responsible for manganese and cadmium uptake in rice. *Plant Cell* 24, 2155–2167. doi: 10.1105/tpc.112.096925

Segond, D., Dellagi, A., Lanquar, V., Rigault, M., Patrit, O., Thomine, S., et al. (2009). NRAMP genes function in *Arabidopsis thaliana* resistance to Erwinia chrysanthemi infection. *Plant J.* 58, 195–207. doi: 10.1111/j.1365-313X.2008.03775.x

Shang, C., Ye, T., Zhou, Q., Chen, P., Li, X., Li, W., et al. (2023). Genome-wide identification and bioinformatics analyses of host defense peptides snakin/GASA in mangrove plants. *Genes.* 14, 923. doi: 10.3390/genes14040923

Shen, X., Li, R., Chai, M., Cheng, S., Tam, N. F. Y., and Han, J. (2021). Does combined heavy metal stress enhance iron plaque formation and heavy metal bioaccumulation in *Kandelia obovata*? *Environ. Exp. Bot.* 186, 104463. doi: 10.1016/j.envexpbot.2021.104463

Sheue, C. R., Liu, H. Y., and Yong, J. W. H. (2003). *Kandelia obovata* (Rhizophoraceae), a new mangrove species from Eastern Asia. *Taxon* 52, 287–294. doi: 10.2307/3647398

Si, X., Lyu, S., Hussain, Q., Ye, H., Huang, C., Li, Y., et al. (2023). Analysis of Delta (9) fatty acid desaturase gene family and their role in oleic acid accumulation in *Carya cathayensis* kernel. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1193063

Sun, M. M., Liu, X., Huang, X. J., Yang, J. J., Qin, P. T., Zhou, H., et al. (2022). Genome-wide identification and expression analysis of the NAC gene family in *kandelia obovata*, a typical mangrove plant. *Curr. Issues Mol. Biol.* 44, 5622–5637. doi: 10.3390/cimb44110381

Tan, Z., Li, J., Guan, J., Wang, C., Zhang, Z., and Shi, G. (2023). Genome-wide identification and expression analysis reveals roles of the NRAMP gene family in iron/cadmium interactions in peanut. *Int. J. Mol. Sci.* 24, 1713. doi: 10.3390/ijms24021713

Thomine, S., Lelièvre, F., Debarbieux, E., Schroeder, J. I., and Barbier-Brygoo, H. (2003). AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant J.* 34, 685–695. doi: 10.1046/j.1365-313X.2003.01760.x

Tian, W., He, G., Qin, L., Li, D., Meng, L., Huang, Y., et al. (2021). Genome-wide analysis of the NRAMP gene family in potato (*Solanum tuberosum*): Identification, expression analysis and response to five heavy metals stress. *Ecotoxicol. Environ. Saf.* 208, 111661. doi: 10.1016/j.ecoenv.2020.111661

Ullah, I., Wang, Y., Eide, D. J., and Dunwell, J. M. (2018). Evolution, and functional analysis of Natural Resistance-Associated Macrophage Proteins (*NRAMPs*) from *Theobroma cacao* and their role in cadmium accumulation. *Sci. Rep.* 8, 14412. doi: 10.1038/s41598-018-32819-y

Vidal, S. M., Malo, D., Vogan, K., Skamene, E., and Gros, P. (1993). Natural resistance to infection with intracellular parasites: Isolation of a candidate for Bcg. *Cell* 73, 469–485. doi: 10.1016/0092-8674(93)90135-D

Vision, T. J., Brown, D. G., and Tanksley, S. D. (2000). The origins of genomic duplications in arabidopsis. *Science* 80. 290, 2114–2117. doi: 10.1126/science.290.5499.2114

Wang, J., Li, Y., Fu, Y., Xie, H., Song, S., Qiu, M., et al. (2019). Mutation at Different Sites of Metal Transporter Gene OsNramp5 Affects Cd Accumulation and Related Agronomic Traits in Rice (Oryza sativa L.). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01081

Weng, B., Huang, Y., Liu, J., Lu, H., and Yan, C. (2014). Alleviated toxicity of cadmium by the rhizosphere of *kandelia obovata* (S., L.) Yong. *Bull. Environ. Contam. Toxicol.* 93, 603–610. doi: 10.1007/s00128-014-1372-9

Weng, B., Xie, X., Weiss, D. J., Liu, J., Lu, H., and Yan, C. (2012). *Kandelia obovata* (S., L.) Yong tolerance mechanisms to Cadmium: Subcellular distribution, chemical forms and thiol pools. *Mar. pollut. Bull.* 64, 2453–2460. doi: 10.1016/j.marpolbul.2012.07.047

Wu, D., He, G., Tian, W., Saleem, M., Li, D., Huang, Y., et al. (2021). OPT gene family analysis of potato (*Solanum tuberosum*) responding to heavy metal stress: Comparative omics and co-expression networks revealed the underlying core templates and specific response patterns. *Int. J. Biol. Macromol.* 188, 892–903. doi: 10.1016/j.ijbiomac.2021.07.183

Yaghobi, M., and Heidari, P. (2023). Genome-wide analysis of aquaporin gene family in triticum turgidum and its expression profile in response to salt stress. *Genes (Basel).* 14, 202. doi: 10.3390/genes14010202

Yang, J., Ling, C., Liu, Y., Zhang, H., Hussain, Q., Lyu, S., et al. (2022). Genome-wide expression profiling analysis of kiwifruit *GolS* and *RFS* genes and identification of *AcRFS4* function in raffinose accumulation. *Int. J. Mol. Sci.* 23, 8836. doi: 10.3390/ijms23168836

Yang, Y., Zhang, Y., Chen, Y., Gul, J., Zhang, J., Liu, Q., et al. (2019). Complete chloroplast genome sequence of the mangrove species *Kandelia obovata* and comparative analyses with related species. *PeerJ* 7, e7713. doi: 10.7717/peerj.7713

Zhang, J., Zhang, M., Song, H., Zhao, J., Shabala, S., Tian, S., et al. (2020). A novel plasma membrane-based *NRAMP* transporter contributes to Cd and Zn hyperaccumulation in Sedum alfredii Hance. *Environ. Exp. Bot.* 176, 104121. doi: 10.1016/j.envexpbot.2020.104121

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 January 2024
DOI 10.3389/fpls.2023.1301164

Check for updates

OPEN ACCESS

EDITED BY
Christos Bazakos,
Max Planck Institute for Plant Breeding Research, Germany

REVIEWED BY
Nunzio D'Agostino,
University of Naples Federico II, Italy
Xiujun Zhang,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Qing Li
liqing130811@126.com

RECEIVED 24 September 2023
ACCEPTED 18 December 2023
PUBLISHED 09 January 2024

CITATION
Lu G and Li Q (2024) Complete mitochondrial genome of *Syzygium samarangense* reveals genomic recombination, gene transfer, and RNA editing events. *Front. Plant Sci.* 14:1301164. doi: 10.3389/fpls.2023.1301164

COPYRIGHT
© 2024 Lu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Complete mitochondrial genome of *Syzygium samarangense* reveals genomic recombination, gene transfer, and RNA editing events

Guilong Lu[1,2] and Qing Li[1]*

[1]Institute of Vegetables, Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa, China, [2]College of Horticulture and Landscape Architecture, Henan Institute of Science and Technology, Xinxiang, China

Wax apple (*Syzygium samarangense*) is a commercial fruit that belongs to one of the most species-rich tree genera in the world. We report here the first complete *S. samarangense* mitogenome obtained using a hybrid assembly strategy. The mitogenome was a 530,242 bp circular molecule encoding 61 unique genes accounting for 7.99% of the full-length genome. Additionally, 167 simple sequence repeats, 19 tandem repeats, and 529 pairs of interspersed repeats were identified. Long read mapping and Sanger sequencing revealed the involvement of two forward repeats (35,843 bp and 22,925 bp) in mediating recombination. Thirteen homologous fragments in the chloroplast genome were identified, accounting for 1.53% of the mitogenome, and the longest fragment was 2,432 bp. An evolutionary analysis showed that *S. samarangense* underwent multiple genomic reorganization events and lost at least four protein-coding genes (PCGs) (*rps2*, *rps7*, *rps11*, and *rps19*). A total of 591 RNA editing sites were predicted in 37 PCGs, of which *nad1*-2, *nad4L*-2, and *rps10*-2 led to the gain of new start codons, while *atp6*-1156, *ccmFC*-1315 and *rps10*-331 created new stop codons. This study reveals the genetic features of the *S. samarangense* mitogenome and provides a scientific basis for further studies of traits with an epistatic basis and for germplasm identification.

KEYWORDS

***Syzygium samarangense*, mitogenome, gene loss, evolution analysis, RNA editing**

## Introduction

Wax apple (*Syzygium samarangense*, also known as java apple, rose apple, wax jambu, and bell fruit) is a small evergreen tree in the family Myrtaceae that is native to the Andaman and Nicobar islands and to the Malaysian Archipelago (Khandaker and Boyce, 2016; Banadka et al., 2022). The genus is extremely rich in germplasm resources, with 1,193

recognized species worldwide. It is mainly grown in tropical and subtropical regions and is distributed from Africa to India, across Southeast Asia, and extending to Hawaii in the Pacific (Govaerts et al., 2008). Wax apple fruit is rich in proteins, dietary fiber, sugar, vitamins, flavonoids, phenolic acids, and other nutrients. Phenolic acids antioxidant, anti-cancer, anti-bacterial, anti-diabetic, and other biological activities (Banadka et al., 2022). In addition, the extracts of leaves, flower buds, bark, and roots of wax apple are effective in removing fire and toxins, drying dampness, and relieving itching (Chinese Herbalism Editorial Board, 1999; Gurib-Fakim, 2006). The species also has high ornamental value, mostly owing to its beautiful shape, fragrant and beautiful flowers, bright fruit color, beautiful fruit shape, and long hanging period. In recent years, the rapid development of the wax apple industry has increased the demand for high quality, high yield, and high resistance varieties. Genome and transcriptome research on wax apple has laid a scientific foundation for the utilization of excellent germplasm resources and improvement of varietal traits (Chen et al., 2017; Liu et al., 2018; Low et al., 2022).

Mitochondria, which provide energy for various physiological activities, are important organelles in eukaryotic cells (Møller et al., 2021). Originally as an endosymbiotic form of alphaproteobacteria (Lang et al., 1999), mitochondria are semi-autonomous organelles exhibiting predominantly matrilineal inheritance (Greiner et al., 2015). Previous studies have shown that mitochondria play an important role in plant growth and development and are closely associated with growth vigor, chloroplast function (Chevigny et al., 2020), stress tolerance (Liberatore et al., 2016). Compared to animal mitogenomes, plant mitogenomes exhibit complex and unique genetic features. Their structures are remarkably diverse, with circular, linear, highly branched, sigma-like, and networked types; and the genome sizes vary widely among species, ranging from 66 Kb (*Viscum scurruloideum*) (Skippington et al., 2015) to 11.7 Mb (*Larix sibirica* Ledeb.) (Putintseva et al., 2020). During long-term evolution, the number of mitochondrial genes in extant plants is very low, and most of them have been transferred into the nucleus. There are generally 32-67 genes that are sparsely distributed and have a highly conserved coding sequence, with a large number of repetitive sequences distributed among them (Mower et al., 2012; Richardson et al., 2013). Accumulation of repetitive sequences may lead to frequent recombination events and structural variation in the mitogenome (Sloan, 2013; Wu et al., 2022), whereas active genomic rearrangements can even lead to cytoplasmic male sterility (Kubo et al., 2011; Liao et al., 2020; Wang et al., 2022). The plant mitogenome can integrate plastome and nuclear DNA fragments and even exogenous mitogenomes, which is a major driver of the enlargement and rapid evolution of plant mitogenomes (Rice et al., 2013; Prasad et al., 2022). In addition, RNA editing events are widespread in plant mitochondrial transcripts, where they are critical for the production of functional proteins and adaptive evolution (Gommans et al., 2009; Ichinose and Sugita, 2017) and are closely associated with cytoplasmic male sterility (Gallagher et al., 2002; Wei et al., 2010). In summary, plant mitogenomes have become an important tool for species classification, evolutionary analyses, and parental traceability (Sibbald et al., 2021; Wang et al., 2022; Zhang et al., 2023). However, related research lags far behind studies on chloroplast and plastid genomes owing to the complexity of the mitogenome. The complete mitochondrial genome of *S. samarangense* has not been reported in public databases (https://www.ncbi.nlm.nih.gov/genome/browse/#!/organelles/).

In this study, the main *S. samarangense* cultivar in China, Black Diamond (Zheng, 2011), was selected for the following investigations: 1) the mitogenome was assembled, annotated, and comparatively analyzed; 2) the repetitive sequences present in this genome were identified, and the possible structure of the repeat-mediated genomic recombination was verified using Sanger sequencing and polymerase chain reaction (PCR) validation; 3) the chloroplast genome of the cultivar was assembled, and the transfer of homologous fragments between this genome and the mitogenome was identified; 4) phylogeny and collinearity analyses were conducted to identify the relationships between this species and closely related species; and 5) RNA editing events were predicted to be present in the protein-coding gene (PCG) transcripts, and editing sites that could create new start and stop codons were validated. This study aims to provide a scientific and theoretical basis for the in-depth understanding of the genetic characteristics of *S. samarangense* and further germplasm identification.

## Materials and methods

### Plant material, DNA and RNA preparation, and sequencing

The young leaves of wax apple (Black Diamond) were harvested from the National Agricultural Science and Technology Park, located in Lhasa, Tibet Autonomous Region, China (Coordinates: 91°2'8"E, 29°38'15"N; Altitude: 3650 m). The gDNA and RNA were separately extracted using the TianGen Plant Genomic DNA Kit and the RNAprep Pure Plant Kit, respectively (Beijing, China). The NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) was used to ensure the integrity of the samples. Samples meeting quality thresholds were then sent to Wuhan Benagen Tech Solutions Company Limited (Wuhan, China) for sequencing, after packing in dry ice. Using DNBSEQ-T7 (developed by Shenzhen Huada Intelligent Technology Co., Ltd., Shenzhen, China), short reads were sequenced, followed by a quality control filtering process on the raw reads using fastp v0.21.0 (Chen, 2023). The Nanopore PromethION sequencer from Oxford Nanopore Technologies (Oxford, UK) was employed to sequence long reads, and NanoFilt v2.8.0 (De Coster et al., 2018) was used for quality control. Long non-coding RNA (lncRNA) sequencing was performed via MGISEQ-2000 (by Shenzhen Huada Intelligent Technology Co., Ltd., Shenzhen, China), and the raw reads were subjected to quality control filtering using SOAPnuke v2.0 (Chen et al., 2018).

### Mitogenome assembly

Initially, Flye v2.9.2 software (University of California, San Diego, USA) (Kolmogorov et al., 2019) was utilized to assemble the long-read sequencing data. Following this, the *Arabidopsis*

*thaliana* mitogenome was used as a reference sequence to identify contig fragments containing mitochondrial genomes via the BLASTn program. Both long-read and short-read data were then aligned to mitogenome contigs using BWA v0.7.17 (Li and Durbin, 2010), after which matched reads were filtered, extracted, and stored separately for subsequent assembly. In the final step, short-read and long-read sequencing data were merged using Unicycler v0.4.7 (The University of Melbourne, Victoria, Australia) (Wick et al., 2017), facilitating hybrid assembly with parameters "- -kmers 57,67,77", and visualized graphics were obtained using Bandage v0.8.1 (Wick et al., 2015).

## Mitogenome annotation and analysis

GeSeq v2.03 (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al., 2017) was utilized to annotate PCGs of the *S. samarangense* mitogenome, with the mitogenomes of *A. thaliana* (NC_037304), *Liriodendron tulipifera* (NC_021152.1), and *Rhodomyrtus tomentosa* (NC_071968.1) used as reference genomes. tRNA and rRNA genes were annotated using tRNAscan-SE v.2.0.11 (Chan et al., 2021) and BLASTN v2.13.0 (Chen et al., 2015), respectively, and errors in annotation were manually fixed using Apollo v1.11.8 (Dunn et al., 2019). For the extraction of protein-coding gene sequences of the *S. samarangense* mitogenome, PhyloSuite v1.2.2 (Xiang et al., 2023) was employed. Mega software (Tamura et al., 2021) was utilized for the codon preference analysis of PCGs and the calculation of the relative synonymous codon usage values (RSCU). RSCU > 1 signifies that the codon is preferentially used by amino acids, whereas RSCU < 1 denotes the opposite.

## Mitogenome repeat elements

The simple sequence repeats (SSRs) were identified using MISA v2.1 (https://webblast.ipk-gatersleben.de/misa/) with the parameter "1-10 2-5 3-4 4-3 5-3 6-3" (Beier et al., 2017). Tandem repeats were recognized using TRF v4.09 (https://tandem.bu.edu/trf/trf.unix.help.html) with the parameter "2 7 7 80 10 50 500 -f -d -m" (Behboudi et al., 2023). Interspersed repeats were detected using the REPuter web server (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) (Kurtz and Schleiermacher, 1999) with the minimum repeat size set at 30 bp.

## Validation of repeat-mediated recombination

To confirm the proposed genomic structure of the *S. samarangense* mitogenome, each set of repeats and their sequences extending 500 bp upstream and downstream were extracted as reference points. Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast) was used to design primers for the four pathways of the dual bifurcation structure, and PCR amplification along with Sanger sequencing (You et al., 2023) was used to verify the validity of the junction sequences. The PCR amplification procedure was carried out in a 50 μL total volume, containing 2 μL of gDNA template, 2 μL each of upstream and downstream primer (10 μmol/L), 25 μL of 2× Rapid Taq Master Mix (Vazyme Biotech Co., Ltd., Nanjing, China), and 19 μL of $ddH_2O$. The PCR steps involved a 95°C pre-denaturation step for 3 min, 35 cycles of 95°C denaturation for 15 s, 55°C annealing for 15 s, and 72°C extension for 30 s, and a final 72°C extension for 15 min.

## Chloroplast genome assembly and gene transfer analysis

GetOrganelle v1.7.7.0 (Freudenthal et al., 2020) was used to assemble the *S. samarangense* chloroplast genome with parameters set to "-R 15 -k 21,45,65,85,105 -F embplant_pt". The chloroplast genome was then annotated with CPGAVAS2 (Liu et al., 2012) and the annotation results were corrected using CPGView software (Liu et al., 2023). Homologous fragments in the mitochondrial and chloroplast genomes were examined using BLASTN v2.13.0 (Chen et al., 2015) with an e-value threshold of 1*e*-5. The results were visualized via Excel v2021 and Circos v0.69-9 (Zhang et al., 2013).

## Phylogenetic and collinearity analysis

The mitogenomes of 30 species closely related to *S. samarangense* were downloaded from NCBI for a phylogenetic analysis (Supplementary Table 1), with *Tribulus terrestris* (MK431825.1) and *Zygophyllum fabago* (MK431827.1) of Zygophyllales assigned as outgroups. PhyloSuite software was used to extract shared genes in the mitogenomes. A multiple sequence alignment was generated using MAFFT v7.505 (Katoh et al., 2019). The "GTR +F+I+I+R2" model was selected for the phylogenetic analysis using IQ-TREE v2.1.3 (Minh et al., 2020), and the maximum likelihood tree was visualized using iTOL v6 (https://itol.embl.de/) (Letunic and Bork, 2007). The mitogenome sequence information for *S. samarangense* and seven closely related species within the same order were analyzed via BLASTn. Homologous sequences exceeding 500 bp were maintained as conserved co-linear blocks, and a multiple synteny plot was drafted using MCscanX (Tang et al., 2008).

## Identification and validation of RNA-editing sites

By using TopHat2 (Mehmood et al., 2020), mitochondrial DNA sequences from the *S. samarangense* mitogenome were mapped to transcriptomic data to obtain transcripts. The potential RNA editing sites were identified by REDItools v2.0 (Flati et al., 2020) by a comparison of DNA and RNA sequences and the selection of positions with a coverage depth ≥ 100× and editing frequency ≥ 0.1. Primer-BLAST was used to design primers for the RNA editing sites that generate start and stop codons (Supplementary Table 2). A

HiScript III 1st Strand cDNA Synthesis Kit (Vazyme, Nanjing, China) was employed to convert the extracted wax apple RNA into cDNA. Both gDNA and cDNA were used as templates for PCR amplification, employing the same system and procedure described in the validation of repeat-mediated recombination. To confirm the RNA editing sites, the amplified products were sequenced and compared.

# Results

## Mitogenome assembly, gene annotation, and analysis

The *S. samarangense* mitogenome was assembled using 10.39 Gb short reads and 11.23 Gb long reads by a hybrid assembly strategy. It contained six contigs, including two double bifurcating structures (Figure 1A). The longest contig (221,760 bp) and the shortest contig (12,514 bp) were single-copy regions. The repetitive sequences were determined based on long-read data and one master circular structure was obtained (Figure 1B) with a total length of 530,242 bp (27.43% Adenine, A; 27.53% Thymine, T; 22.63% Cytosine, C; and 22.41% Guanine, G). The size of this genome differs slightly from those of other species of the same family: *Rhodomyrtus tomentosa* (400,547 bp, NC_071968.1), *Eucalyptus grandis* (478,813 bp, NC_040010.1), and *E. polybractea* (580,440 bp, CM048836.1).

The *S. samarangense* mitogenome was annotated with 37 unique PCGs, including 24 core genes and 13 non-core genes, 21 tRNA genes (four tRNAs were multiple copies), and three rRNA genes (Figure 2, Table 1). *S. samarangense* lost four genes (*rps2*, *rps7*, *rps11*, and *rps19*) compared to the PCGs of the "fossilized" *L. tulipifera* mitogenome (Richardson et al., 2013). The total length of coding sequences in the *S. samarangense* mitochondrial genome (PCGs, tRNA, and rRNA genes) was 42,391 bp. This accounted for 7.99% of the entire genome, while over 90% belonged to intergenic regions.

RSCU, which reflects the optimized results of the evolution of plant ecological adaptations, is the "dialect" of a gene or genome (Tyagi et al., 2023). The amino acids encoded by 37 PCGs from the *S. samarangense* mitogenome generally showed codon usage bias, except for the RSCU of 1.00 for the initiation codon (AUG) and tryptophan (UGG) (Supplementary Figure 1, Supplementary Table 3). For example, alanine (Ala) had a high codon usage preference for GCU with the highest RSCU value of 1.56. Histidine (His) had a usage preference for CAU, while tyrosine (Tyr) had a usage preference for UAU, both of which had an RSCU value of 1.52. In addition, cysteine (Cys), lysine (Lys), phenylalanine (Phe), and valine (Val) had maximum RSCU values below 1.20 and did not have strong codon usage preference. However, the third codon of these amino acids has a greater preference for the use of U or A than other nucleotides.

## Repeat sequence analysis and repeat-mediated recombination

There were many repetitive sequences in the *S. samarangense* mitogenome (Supplementary Figure 2). A total of 167 simple sequence repeats (SSRs) were identified throughout the genome (Figure 3A, Supplementary Table 4), and monomeric and dimeric forms of SSRs accounted for 49.10% of the total SSRs. T monomeric repeat sequences (No.35) accounted for 57.38% of the 61 monomeric SSRs, and the TA repeat was the most common type of dimeric SSR, accounting for 38.10% of dimeric SSRs; no hexamer SSRs were detected. Nineteen tandem repeats with lengths between 15 bp and 33 bp and ≥76% match were identified (Supplementary Table 5). In addition, 529 pairs of dispersed repeats with ≥30 bp were detected (Figure 3B, Supplementary Table 6), including 249 pairs of palindromic repeats, 280 pairs of forward repeats, and no reverse and complementary repeats. The total length of the above three types of SSRs, tandem repeats, and interspersed repeats were 1,951 bp, 799 bp and 177,672 bp, accounting for 0.37%, 0.15%, and 33.51% of the mitogenome, respectively.

The structure of the plant mitogenome is dynamic owing to long repetitive sequences that can mediate genome recombination causing conformational changes (Gualberto and Newton, 2017; Kozik et al., 2019). It can be reasonably deduced that a single circular molecule is insufficient for the comprehensive display of the *S. samarangense* mitogenome structure. There may be multiple connections at branch nodes, and potential secondary conformations mediated by R1 and/or R2 (Table 2), as shown in

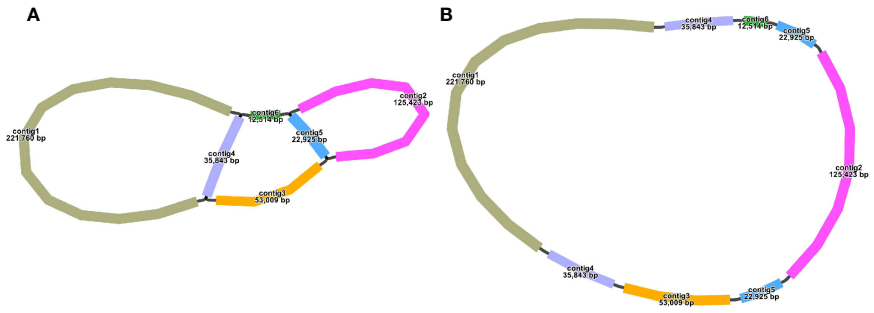



FIGURE 1
Mitogenome sketch **(A)** and master circular structure **(B)** in *S. samarangense*.

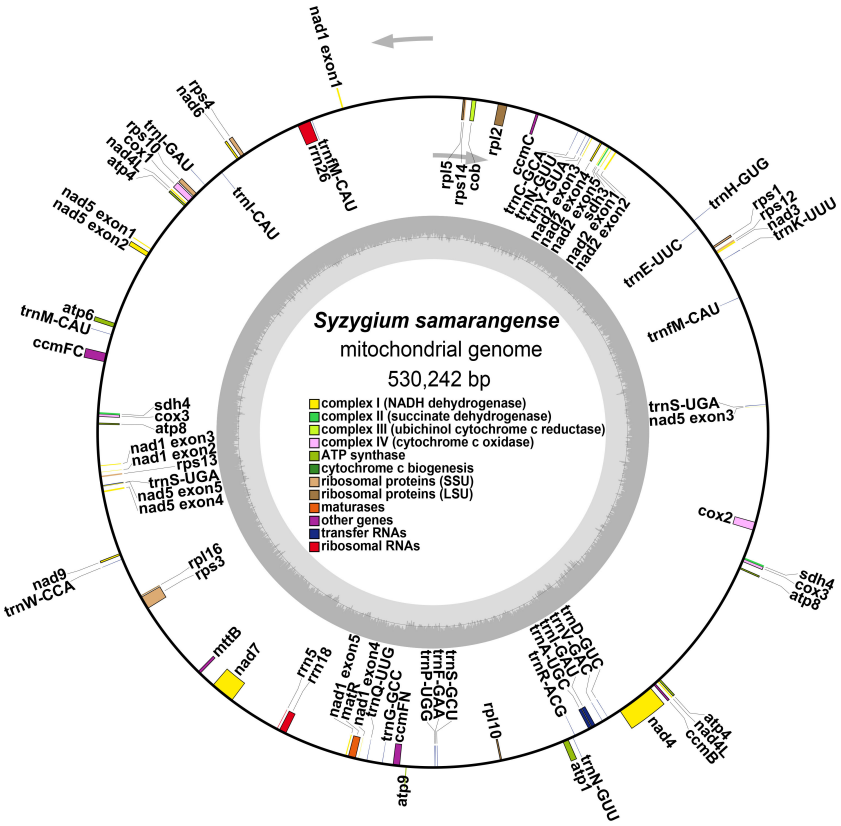



FIGURE 2
Circular maps of the *S. samarangense* mitogenome.

TABLE 1 Genes in the *S. samarangense* mitogenome.

| Group of genes | | Name of genes |
|---|---|---|
| Core genes | ATP synthase | *atp1*, *atp4* (×2), *atp6*, *atp8* (×2), *atp9* |
| | NADH dehydrogenase | *nad1*, *nad2*, *nad3*, *nad4*, *nad4L* (×2), *nad5*, *nad6*, *nad7*, *nad9* |
| | Cytochrome b | *cob* |
| | Cytochrome c biogenesis | *ccmB*, *ccmC*, *ccmFC*, *ccmFN* |
| | Cytochrome c oxidase | *cox1*, *cox2*, *cox3* (×2) |
| | Maturases | *matR* |
| | Protein transport subunit | *mttB* |
| Variable genes | Ribosomal protein large subunit | *rpl2*, *rpl5*, *rpl10*, *rpl16* |
| | Ribosomal protein small subunit | *rps1*, *rps3*, *rps4*, *rps10*, *rps12*, *rps13*, *rps14* |
| | Succinate dehydrogenase | *sdh3*, *sdh4* (×2) |
| rRNA genes | Ribosome RNA | *rrn5*, *rrn18*, *rrn26* |
| tRNA genes | Transfer RNA | *trnA-UGC*, *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU* (×2), *trnG-GCC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU* (×2), *trnK-UUU*, *trnM-CAU*, *trnN-GUU* (×2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*, *trnS-GCU*, *trnS-UGA* (×2), *trnV-GAC*, *trnW-CCA*, *trnY-GUA* |

*(Continued)*

TABLE 1 Continued

The number in brackets represents the copy number of the gene.

Figures 4A–C. R1 and R2-mediated recombination was identified using a junction approach with validation using newly designed primers (Supplementary Table 7) and electrophoresis (Figures 4D, E). Detailed sequencing comparisons are shown in Supplementary Figure 3. In summary, the *S. samarangense* mitogenome had a variety of potential recombination conformations.

## Chloroplast genome and sequence migration analysis

Sequence migration between the chloroplast and plant mitochondrial genomes has been prevalent throughout their long-

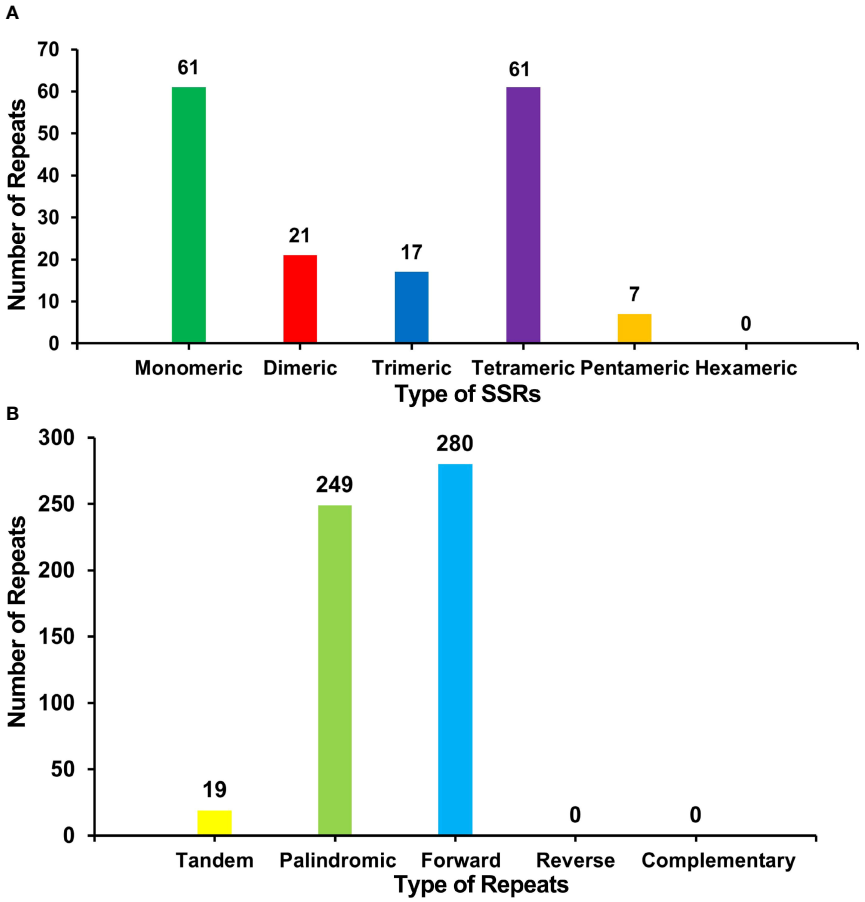



FIGURE 3
**(A)** Type and number of SSRs in the *S. samarangense* mitogenome. Green, red, blue, purple, and yellow indicate monomeric, dimeric, trimeric, tetrameric, and pentameric, respectively. **(B)** Type and number of repeat sequences in the *S. samarangense* mitogenome. Yellow, light green, and blue indicate tandem repeats, palindromic repeats, and forward repeats, respectively.

term evolution (Prasad et al., 2022). The chloroplast genome of *S. samarangense* was assembled using the same sequencing data, with a size of 159,109 bp (Figure 5A). Thirteen homologous fragments of the chloroplast genome and mitochondrial genome were identified based on sequence similarity (Figure 5B, Supplementary Table 8), with a total length of 8,100 bp accounting for 1.53% of the total mitogenome length. Mitochondrial plastid sequence seven (MTPT7) was the longest (2,432 bp). Eighteen unique genes were identified after annotating these homologous sequences, of which ten were complete, including six PCGs (*psbJ*, *psbL*, *psbF*, *psbE*, *rps19*, and *rpl2*) and four tRNA genes (*trnD-GUC*, *trnH-GUG*, *trnM-CAU*, and *trnN-GUU*). The eight incomplete genes included *ndhC*, *psaD*, *rpl22*, *rpl23*, *rpoB*, *ycf2*, *trnI-GAU*, and *trnV-UAC*. However, these chloroplast genes transferred to the mitochondrial genome have undergone the varying degrees of sequence duplication, substitution, mutation, or loss during evolution, which have hindered their ability to perform their normal functions and caused them to evolve into pseudogenes (Mower et al., 2010; Mower et al., 2012).

TABLE 2 List of two repeat sequences mediated recombination in the *S. samarangense* mitogenome.

| Repeat Name | Repeat 1 | Repeat 2 |
|---|---|---|
| Identities (%) | 100 | 100 |
| Length (bp) | 35,843 | 22,925 |
| Position-1 | 201,357-237,200 | 125,423-148,348 |
| Position-2 | 458,960-494,803 | 507,317-530,242 |
| E-value | 0 | 0 |
| Type | Forward | Forward |

## Phylogeny and collinearity analysis

The loss or gain of PCGs commonly occurs during plant mitogenome evolution (Skippington et al., 2015; Wu et al., 2022). Only 23 out of 37 PCGs from the genomes of closely related species included in the phylogenetic analysis were shared: *atp1*, *atp4*, *atp6*, *atp8*, *ccmB*, *ccmC*, *ccmFC*, *ccmFN*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad4*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl5*, *rpl16*, *rps3*, and *sdh4*. The phylogenetic tree showed that *S. samarangense* was most closely related to *Eucalyptus grandis* and *Rhodomyrtus tomentosa*

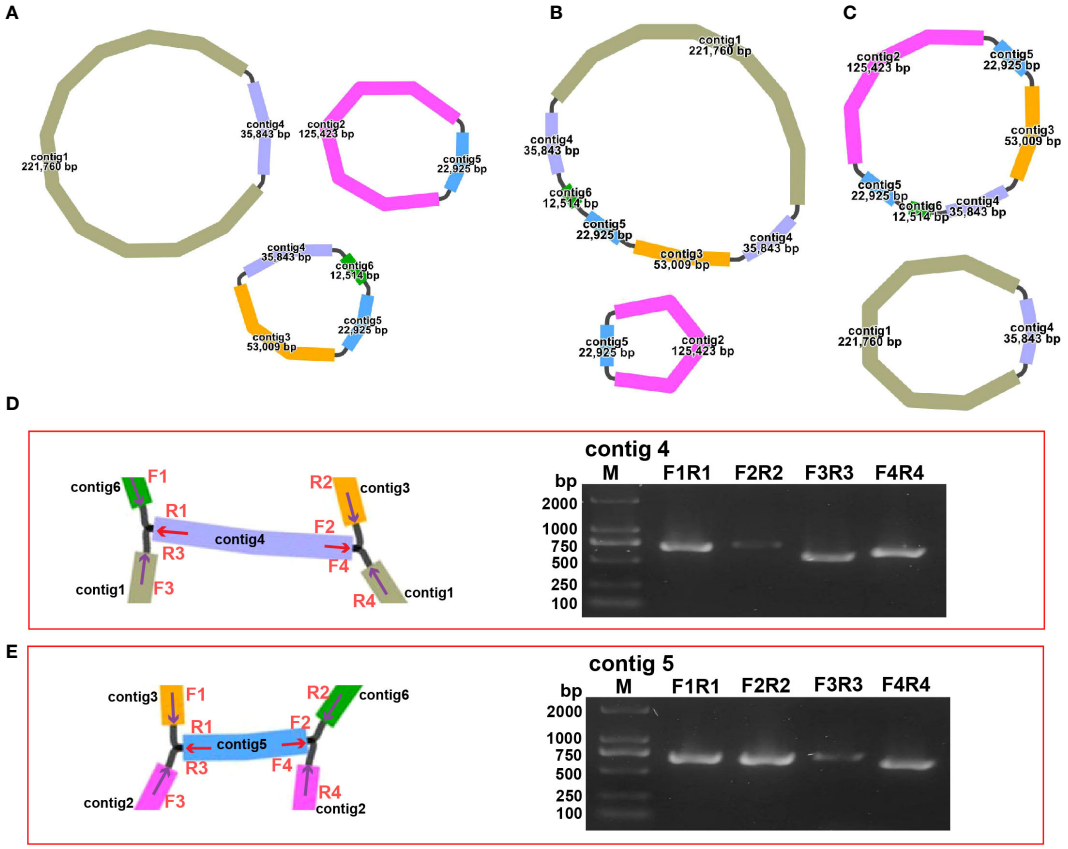



FIGURE 4
Recombination structure prediction and verification in *S. samarangense*. **(A)** Secondary conformation 1 (including 3 small circles) mediated by R1 and R2; **(B)** Secondary conformation 2 (including 2 small circles) mediated by R2; **(C)** Secondary conformation 3 (including 2 small circles) mediated by R1. **(D)** Verification of junction sequences in R1. **(E)** Verification of junction sequences in R2.

in the Myrtle family (Figure 6A). Moreover, the topology of the phylogeny of the above species (based on mitochondrial DNA) coincided with the latest classification of the Angiosperm Phylogeny Group.

A collinearity analysis of *S. samarangense* and seven Myrtales species (Figure 6B, Supplementary Table 9) revealed many homologous co-linear blocks, although the sizes of these co-linear blocks were generally short. Moreover, the co-linear block

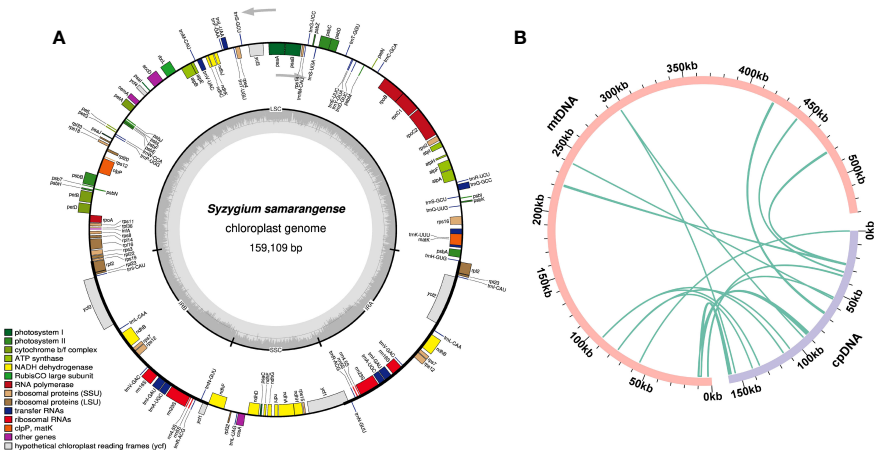



FIGURE 5
Chloroplast genome of *S. samarangense* **(A)** and sequence migration analysis between the chloroplast genome and mitochondrial genome **(B)**. Pink and light purple arcs represent the mitogenome and the chloroplast genome, respectively, and the dark green lines between the arcs correspond to the homologous fragments.

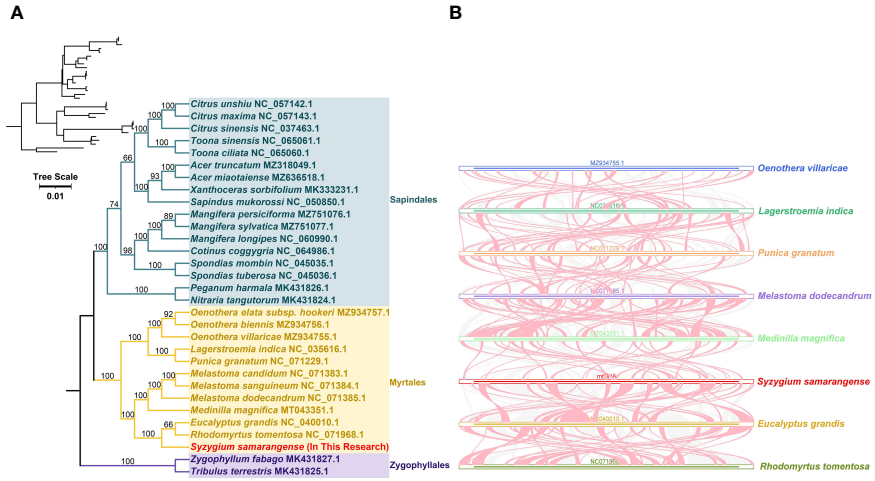



FIGURE 6
**(A)** Phylogenetic analysis of *S. samarangense* and related genera. Numbers at nodes show the bootstrap support values, and different colors indicate different Orders. **(B)** Collinearity analysis between *S. samarangense* and related genera. Blue, dark green, orange, purple, light green, red, yellow, and yellow-green lines represent *Oenothera villaricae*, *Lagerstroemia indica*, *Punica granatum*, *Melastoma dodecandrum*, *Medilla magnifica*, *Syzygium samarangense*, *Eucalyptus grandis*, and *Rhodomyrtus tomentosa*, respectively. Red curved areas indicate regions where inversions occur, and the gray areas indicate regions with high homology.

arrangement between the individual mitogenomes was inconsistent. It was hypothesized that extensive genomic reorganization occurred during the evolution of *S. samarangense* and closely related species. In addition, the identification of unique regions in species that lack homology with other species indicated species specificity in evolution.

## RNA editing events

RNA editing, one of the steps required for gene expression, is prevalent in higher plant mitochondria (Takenaka et al., 2013). RNA editing events were predicted for 37 PCGs of the *S. samarangense* mitogenome based on transcriptome data. They were observed in all of those genes, with a total of 591 RNA editing sites (Figure 7A), of which 473 had an editing frequency ≥ 0.80 (Supplementary Table 10). All were base C to U edits, and their RNA-seq mapping results were uploaded to the figshare platform (doi: 10.6084/m9.figshare.22650238). Among them, 47 RNA editing sites were identified in the *ccmB* gene, which had the highest number of edits among all mitochondrial genes. This was followed by the *nad4* gene, with 44 RNA editing sites. Furthermore, these editing sites were primarily observed in the

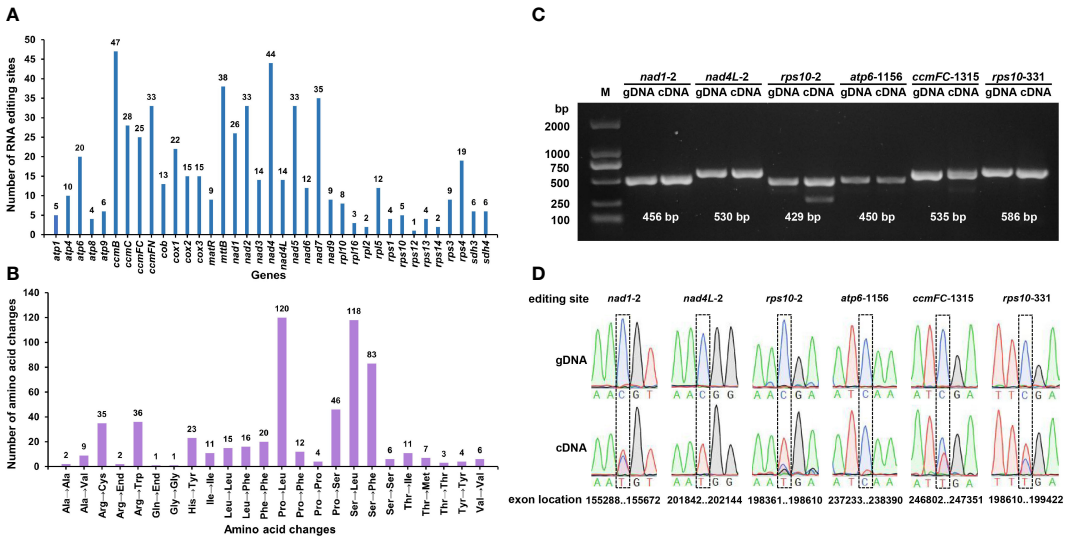



FIGURE 7
Predicted RNA editing sites **(A)** and corresponding amino acid changes **(B)** in *S. samarangense* mitogenome PCGs. PCR products **(C)** and sequence comparisons **(D)** for six specific editing sites.

second and first bases of the coding amino acids, and the number was 336 (56.85% of the total) and 187 (31.64%), respectively.

Most of the editing events resulted in non-synonymous codon changes, primarily involving the following three amino acid changes: Pro to Leu (No.120), Ser to Leu (No.118), and Ser to Phe (No.83) (Figure 7B). We predicted one ACG (Thr) to AUG (Met) editing event in *nad1*, *nad4L*, and *rps10*; one CAA (Gln) to UAA (End) in *atp6*; and one CGA (Arg) to UGA (End) in *ccmFC* and *rps10*. Further, we validated these six editing sites by PCR (Figure 7C) and Sanger sequencing (Figure 7D, Supplementary file 1). The editing frequency of the *atp6*-1156 editing site was low (the red line indicates base T in transcription). However, the specific effects of changes in the start and stop codons caused by RNA editing events in the *S. samarangense* mitogenome require in-depth studies.

## Discussion

A high-quality *S. samarangense* mitogenome with a length of 530,242 bp was assembled. The size of this genome varies slightly from that of other Myrtaceae species and differs significantly from those of *Fragaria* (275.2–351.2 Kb) (Fan et al., 2022), *Mangifera indica* (714.4–750.9 Kb) (Niu et al., 2022), and *Rhynchospora* (1.7–2.2 Mb) (Hofstatter et al., 2022). The 45.03% GC content in the *S. samarangense* mitogenome was not significantly different from those of the above species. This indicated that the GC content of the plant mitogenome was relatively conserved during evolution. *L. tulipifera* is an ancient legacy plant whose mitogenome evolved very slowly, retaining genes frequently lost in angiosperms and containing 41 PCGs (Richardson et al., 2013). At least four PCGs (*rps2*, *rps7*, *rps11*, and *rps19*) were lost in the *S. samarangense* genome. This indicated that a gene loss or transfer event occurred in this species, and ribosomal protein genes are more likely to be lost (Adams and Palmer, 2003). The coding region of the *S. samarangense* mitogenome only accounted for 7.99% of its full length with >90% non-coding regions; the gene distribution was very sparse, with highly conserved coding sequences. This proportion of the genome that is coding was similar to those in *L. tulipifera* (7.7%, excluding *cis*-spliced introns) (Richardson et al., 2013) and *Populus simonii* (8.25%) (Bi et al., 2022).

Repeats are identical or symmetrical DNA sequence fragments that frequently occur in plant genomes, occupy a large proportion of plant mitogenomes, and have important roles in genome size evolution, gene expression regulation, and responses to stress (Wendel et al., 2016). They are classified into tandem repeats and interspersed repeats according to whether they are adjacent or not. SSRs are a special kind of tandem repeat (generally not more than 6 bp) that are often used as genetic markers in plants (Khera et al., 2015). This study identified a total of 167 SSRs in the *S. samarangense* mitogenome, providing a large number of referenceable markers for germplasm diversity evaluation and species identification. Interspersed repeat sequences (also known as transposable elements) are a class of DNA sequences that can move their position on the genome, leading to gradual changes in functional genes and contributing to evolution (Bennetzen and Wang, 2014). They can regulate gene expression and affect plant phenotypic traits (Lisch, 2013), such as plant height and ear height in maize (Li et al., 2020), pigment formation in blood orange (Butelli et al., 2012), fruit shape in tomato (Xiao et al., 2008), and aluminum tolerance in wheat (Tovkach et al., 2013). In this study, 529 pairs of interspersed repeats were detected in the *S. samarangense* mitogenome; however, their effects on plant phenotypic traits require further in-depth study.

Repeat-mediated genome rearrangement is one of the main drivers of changes in mitogenome structure in plants, affecting gene organization and creating gene chimeras. This process affects plant phenotypes and evolution (Sanchez-Puerta et al., 2015; Gualberto and Newton, 2017). Large repetitive sequences over 1000 bp are more prone to recombination in the genome (Kozik et al., 2019). This study identified multiple potential conformations due to R1 (35,843 bp) and/or R2 (22,925 bp) repeat-mediated recombination in the *S. samarangense* mitogenome, although a single circle conformation containing information about the entire genome predominates (Figures 1B, 4A–C). Therefore, rearrangement events in *S. samarangense* resulted in differentiation within the *Syzygium* genome.

Horizontal gene transfer is the transmission of genes or DNA fragments within cells or across species boundaries by asexual means; it is one of the main drivers of land plant evolution (Prasad et al., 2022). However, the introduced genes usually degenerate into pseudogenes (Mower et al., 2010). This study identified a total of 13 homologous fragments between the chloroplast genome and the *S. samarangense* mitogenome, accounting for 1.53% of the total mitogenome length. Among these homologous sequences, ten complete mitochondrial genes were identified, including six PCGs and four tRNA genes. The migration of gene sequences led to mitogenome expansion, which contributed to plant mitochondrial genetic diversity, although most of these migrating sequences lost their integrity during evolution (Alverson et al., 2011). Furthermore, only 23 PCGs were identical among the 31 closely related species in this study. This further suggested that gene gain/loss events occurred in closely related plants during evolution (Allen et al., 2007; Kubo and Newton, 2008).

RNA editing is a kind of nucleotide modification at the RNA level in higher plants and plays an essential role in plant adaptation and development. Changes in nonsynonymous codons caused by RNA editing may result in significant changes in gene function, especially with the creation of new start and stop codons (Ichinose and Sugita, 2017; Hao et al., 2021). RNA editing produces an ACG (Thr) to AUG (Met) editing events that can serve as the starting point for gene transcription (Quiñones et al., 1995; Zandueta-Criado and Bock, 2004). Similarly, RNA editing may produce stop codons that cause the premature termination of protein synthesis. Mitochondrial RNA editing truncated the *orf77* chimeric open reading frame associated with *S* cytoplasmic male sterility in maize, resulting in pollen abortion (Gallagher et al., 2002). Meanwhile, RNA editing changes from CAA (Gln) to UAA (End) at the *atp6*-1003 site ensured the normal synthesis of the polypeptide encoded by *atp6* in Yunnan purple rice maintenance lines, while sterile lines had no RNA editing at this site (Wei et al., 2010). Interestingly, this study identified six editing sites that could create start or stop codons: *nad1*-2 (Thr to Met), *nad4L*-2 (Thr to

Met), *rps10*-2 (Thr to Met), *atp6*-1156 (Gln to End), *ccmFC*-1315 (Arg to End), and *rps10*-331 (Arg to End). Their specific roles in plant growth and development require in-depth exploration.

## Conclusions

In this study, we assembled the first complete mitogenome of *S. samarangense*. It was 530,242 bp in length, with a GC content of 45.03%, and encoded 61 unique genes. The many repetitive sequences and homologous segments of the chloroplast genome in the mitogenome of *S. samarangense* suggests that it may underwent multiple genomic recombination events during evolution. Moreover, two forward repeats were involved in mediating recombination. Therefore, multiple secondary conformations may exist in addition to the master circular structure. A total of 591 RNA editing sites were identified in the PCGs, among which six sites create start or stop codons and should be the focus of future studies. The newly obtained mitogenome can be used as a reference genome for analyses of other *Syzygium* species and provides important information for the molecular breeding of *S. samarangense*.

## 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. The mitogenome sequence supporting the conclusions of this article is available in GenBank (https://www.ncbi.nlm.nih.gov/) under accession number: OQ701348. The chloroplast genome sequence information has been uploaded to NCBI (accession number: OR766332). The Illumina, Nanopore, and transcriptome data of Syzygium samarangense have been deposited to the Sequence Read Archive repository under SRR26670650, SRR26670651, and SRR24058007, respectively.

## Author contributions

GL: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft. QL: Funding acquisition, Project administration, Resources, Supervision, Visualization, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Central Guided Local Development of Science and Technology Project (YDZX20195400004007), the Science and Technology Program of Tibet Autonomous Region, China (XZ2019NKNS009) and the Scientific Research Start-up Fund for High-level Introduced Talents of Henan Institute of Science and Technology.

## Acknowledgments

The authors sincerely thank the experimental personnel and bioinformatics analysis at Wuhan Benagen Technology Co., Ltd (www.benagen.com) and MitoRun research group participated in this project. The authors are thankful to Youxiong Que for reviewing the manuscript.

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1301164/full#supplementary-material

## References

Adams, K. L., and Palmer, J. D. (2003). Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* 29, 380–395. doi: 10.1016/S1055-7903(03)00194-5

Allen, J. O., Fauron, C. M., Minx, P., Roark, L., Oddiraju, S., Lin, G. N., et al. (2007). Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 177, 1173–1192. doi: 10.1534/genetics.107.073312

Alverson, A. J., Rice, D. W., Dickinson, S., Barry, K., and Palmer, J. D. (2011). Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* 23, 2499–2513. doi: 10.1105/tpc.111.087189

Banadka, A., Wudali, N. S., Al-Khayri, J. M., and Nagella, P. (2022). The role of *Syzygium samarangense* in nutrition and economy: An overview. *S. Afr. J. Bot.* 145, 481–492. doi: 10.1016/j.sajb.2022.03.014

Behboudi, R., Nouri-Baygi, M., and Naghibzadeh, M. (2023). RPTRF: A rapid perfect tandem repeat finder tool for DNA sequences. *Biosystems* 226, 104869. doi: 10.1016/j.biosystems.2023.104869

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Bennetzen, J. L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* 65, 505–530. doi: 10.1146/annurev-arplant-050213-035811

Bi, C., Qu, Y., Hou, J., Wu, K., Ye, N., and Yin, T. (2022). Deciphering the multi-chromosomal mitochondrial genome of *Populus simonii*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.914635

Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., et al. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* 24, 1242–1255. doi: 10.1105/tpc.111.095232

Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49, 9077–9096. doi: 10.1093/nar/gkab688

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2, e107. doi: 10.1002/imt2.107

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* 7, gix120. doi: 10.1093/gigascience/gix120

Chen, F., Hao, Y., Yin, Z., Wu, G., and Jiang, X. (2017). Transcriptome of wax apple (*Syzygium samarangense*) provides insights into nitric oxide-induced delays of postharvest cottony softening. *Acta Physiol. Plant* 39, 273. doi: 10.1007/s11738-017-2569-4

Chen, Y., Ye, W., Zhang, Y., and Xu, Y. (2015). High speed BLASTN: an accelerated MegaBLAST search tool. *Nucleic Acids Res.* 43, 7762–7768. doi: 10.1093/nar/gkv784

Chevigny, N., Schatz-Daas, D., Lotfi, F., and Gualberto, J. M. (2020). DNA repair and the stability of the plant mitochondrial genome. *Int. J. Mol. Sci.* 21, 328. doi: 10.3390/ijms21010328

Chinese Herbalism Editorial Board (1999). *Chinese Herbalism* Vol. 5 (Shanghai: Shanghai Scientific and Technical Publishers), 658.

De Coster, W., D'hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149

Dunn, N. A., Unni, D. R., Diesh, C., Munoz-Torres, M., Harris, N. L., Yao, E., et al. (2019). Apollo: Democratizing genome annotation. *PloS Comput. Biol.* 15, e1006790. doi: 10.1371/journal.pcbi.1006790

Fan, W., Liu, F., Jia, Q., Du, H., Chen, W., Ruan, J., et al. (2022). *Fragaria* mitogenomes evolve rapidly in structure but slowly in sequence and incur frequent multinucleotide mutations mediated by microinversions. *New Phytol.* 236, 745–759. doi: 10.1111/nph.18334

Flati, T., Gioiosa, S., Spallanzani, N., Tagliaferri, I., Diroma, M. A., Pesole, G., et al. (2020). HPC-REDItools: a novel HPC-aware tool for improved large scale RNA-editing analysis. *BMC Bioinf.* 21, 353. doi: 10.1186/s12859-020-03562-x

Freudenthal, J. A., Pfaff, S., Terhoeven, N., Korte, A., Ankenbrand, M. J., and Förster, F. (2020). A systematic comparison of chloroplast genome assembly tools. *Genome Biol.* 21, 254. doi: 10.1186/s13059-020-02153-6

Gallagher, L. J., Betz, S. K., and Chase, C. D. (2002). Mitochondrial RNA editing truncates a chimeric open reading frame associated with *S* male-sterility in maize. *Curr. Genet.* 42, 179–184. doi: 10.1007/s00294-002-0344-5

Gommans, W. M., Mullen, S. P., and Maas, S. (2009). RNA editing: a driving force for adaptive evolution? *Bioessays* 31, 1137–1145. doi: 10.1002/bies.200900045

Govaerts, R. H. A., Sobral, M., Ashton, P., Barrie, F. R., Holst, B. K., Landrum, L., et al. (2008). *World Checklist of Myrtaceae* (London: Royal Botanic Gardens).

Greiner, S., Sobanski, J., and Bock, R. (2015). Why are most organelle genomes transmitted maternally? *Bioessays* 37, 80–94. doi: 10.1002/bies.201400110

Gualberto, J. M., and Newton, K. J. (2017). Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* 68, 225–252. doi: 10.1146/annurev-arplant-043015-112232

Gurib-Fakim, A. (2006). Medicinal plants: traditions of yesterday and drugs of tomorrow. *Mol. Aspects Med.* 27, 1–93. doi: 10.1016/j.mam.2005.07.008

Hao, W., Liu, G., Wang, W., Shen, W., Zhao, Y., Sun, J., et al. (2021). RNA editing and its roles in plant organelles. *Front. Genet.* 12. doi: 10.3389/fgene.2021.757109

Hofstatter, P. G., Thangavel, G., Lux, T., Neumann, P., Vondrak, T., Novak, P., et al. (2022). Repeat-based holocentromeres influence genome architecture and karyotype evolution. *Cell* 185, 3153–3168. doi: 10.1016/j.cell.2022.06.045

Ichinose, M., and Sugita, M. (2017). RNA editing and its molecular mechanism in plant organelles. *Genes (Basel)* 8, 5. doi: 10.3390/genes8010005

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Khandaker, M. M., and Boyce, A. N. (2016). Growth, distribution and physiochemical properties of wax apple (*Syzygium samarangense*): A Review. *Aust. J. Crop Sci.* 10, 1640–1648. doi: 10.21475/AJCS.2016.10.12.PNE306

Khera, P., Saxena, R., Sameerkumar, C. V., Saxena, K., and Varshney, R. K. (2015). Mitochondrial SSRs and their utility in distinguishing wild species, CMS lines and maintainer lines in pigeonpea (*Cajanus cajan* L.). *Euphytica* 206, 737–746. doi: 10.1007/s10681-015-1504-2

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37, 540–546. doi: 10.1038/s41587-019-0072-8

Kozik, A., Rowan, B. A., Lavelle, D., Berke, L., Schranz, M. E., Michelmore, R. W., et al. (2019). The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PloS Genet.* 15, e1008373. doi: 10.1371/journal.pgen.1008373

Kubo, T., Kitazaki, K., Matsunaga, M., Kagami, H., and Mikami, T. (2011). Male sterility-inducing mitochondrial genomes: how do they differ? *Crit. Rev. Plant Sci.* 30, 378–400. doi: 10.1080/07352689.2011.587727

Kubo, T., and Newton, K. J. (2008). Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8, 5–14. doi: 10.1016/j.mito.2007.10.006

Kurtz, S., and Schleiermacher, C. (1999). REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* 15, 426–427. doi: 10.1093/bioinformatics/15.5.426

Lang, B. F., Gray, M. W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* 33, 351–397. doi: 10.1146/annurev.genet.33.1.351

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23, 127–128. doi: 10.1093/bioinformatics/btl529

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

Li, C., Tang, J., Hu, Z., Wang, J., Yu, T., Yi, H., et al. (2020). A novel maize dwarf mutant generated by *Ty1*-copia LTR-retrotransposon insertion in *Brachytic2* after spaceflight. *Plant Cell Rep.* 39, 393–408. doi: 10.1007/s00299-019-02498-8

Liao, X., Wei, M., Khan, A., Zhao, Y., Kong, X., Zhou, B., et al. (2020). Comparative analysis of mitochondrial genome and expression variation between UG93A and UG93B reveals a candidate gene related to cytoplasmic male sterility in kenaf. *Ind. Crop Prod.* 152, 112502. doi: 10.1016/j.indcrop.2020.112502

Liberatore, K. L., Dukowic-Schulze, S., Miller, M. E., Chen, C., and Kianian, S. F. (2016). The role of mitochondria in plant development and stress tolerance. *Free Radical Bio. Med.* 100, 238–256. doi: 10.1016/j.freeradbiomed.2016.03.033

Lisch, D. (2013). How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61. doi: 10.1038/nrg3374

Liu, S., Ni, Y., Li, J., Zhang, X., Yang, H., Chen, H., et al. (2023). CPGView: A package for visualizing detailed chloroplast genome structures. *Mol. Ecol. Resour.* 23, 694–704. doi: 10.1111/1755-0998.13729

Liu, J., Ni, S., Zheng, C., Shi, C., and Niu, Y. (2018). Chloroplast genome of tropical and sub-tropical fruit tree *Syzygium samarangense* (Myrtaceae). *Mitochondrial DNA B Resour.* 3, 890–891. doi: 10.1080/23802359.2018.1501296

Liu, C., Shi, L., Zhu, Y., Chen, H., Zhang, J., Lin, X., et al. (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13, 715. doi: 10.1186/1471-2164-13-715

Low, Y. W., Rajaraman, S., Tomlin, C. M., Ahmad, J. A., Ardi, W. H. , Armstrong, K., et al. (2022). Genomic insights into rapid speciation within the world's largest tree genus *Syzygium*. *Nat. Commun.* 13, 5031. doi: 10.1038/s41467-022-32637-x

Mehmood, A., Laiho, A., Venäläinen, M. S., McGlinchey, A. J., Wang, N., and Elo, L. L. (2020). Systematic evaluation of differential splicing tools for RNA-seq studies. *Brief. Bioinform.* 21, 2052–2065. doi: 10.1093/bib/bbz126

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Møller, I. M., Rasmusson, A. G., and Van Aken, O. (2021). Plant mitochondria - past, present and future. *Plant J.* 108, 912–959. doi: 10.1111/tpj.15495

Mower, J. P., Sloan, D. B., and Alverson, A. J. (2012). "Plant mitochondrial genome diversity: The genomics revolution," in *Plant Genome Diversity*, vol. 1 . Eds. J. Wendel, J. Greilhuber, J. Dolezel and I. Leitch (Vienna: Springer), 123–144. doi: 10.1007/978-3-7091-1130-7_9

Mower, J. P., Stefanović, S., Hao, W., Gummow, J. S., Jain, K., Ahmed, D., et al. (2010). Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol.* 8, 150. doi: 10.1186/1741-7007-8-150

Niu, Y., Gao, C., and Liu, J. (2022). Complete mitochondrial genomes of three *Mangifera* species, their genomic structure and gene transfer from chloroplast genomes. *BMC Genomics* 23, 147. doi: 10.1186/s12864-022-08383-1

Prasad, A., Chirom, O., and Prasad, M. (2022). Horizontal gene transfer and the evolution of land plants. *Trends Plant Sci.* 27, 1203–1205. doi: 10.1016/j.tplants.2022.08.020

Putintseva, Y. A., Bondar, E. I., Simonov, E. P., Sharov, V. V., Oreshkova, N. V., Kuzmin, D. A., et al. (2020). Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics* 21, 654. doi: 10.1186/s12864-020-07061-4

Quiñones, V., Zanlungo, S., Holuigue, L., Litvak, S., and Jordana, X. (1995). The *cox1* initiation codon is created by RNA editing in potato mitochondria. *Plant Physiol.* 108, 1327–1328. doi: 10.1104/pp.108.3.1327

Rice, D. W., Alverson, A. J., Richardson, A. O., Young, G. J., Sanchez-Puerta, M. V., Munzinger, J., et al. (2013). Horizontal transfer of entire genomes *via* mitochondrial fusion in the angiosperm *Amborella*. *Science* 342, 1468–1473. doi: 10.1126/science.1246275

Richardson, A. O., Rice, D. W., Young, G. J., Alverson, A. J., and Palmer, J. D. (2013). The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 11, 29. doi: 10.1186/1741-7007-11-29

Sanchez-Puerta, M. V., Zubko, M. K., and Palmer, J. D. (2015). Homologous recombination and retention of a single form of most genes shape the highly chimeric mitochondrial genome of a cybrid plant. *New Phytol.* 206, 381–396. doi: 10.1111/nph.13188

Sibbald, S. J., Lawton, M., and Archibald, J. M. (2021). Mitochondrial genome evolution in pelagophyte algae. *Genome Biol. Evol.* 13, evab018. doi: 10.1093/gbe/evab018

Skippington, E., Barkman, T. J., Rice, D. W., and Palmer, J. D. (2015). Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc. Natl. Acad. Sci. U.S.A.* 112, E3515–E3524. doi: 10.1073/pnas.1504491112

Sloan, D. B. (2013). One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytol.* 200, 978–985. doi: 10.1111/nph.12395

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., and Brennicke, A. (2013). RNA editing in plants and its evolution. *Annu. Rev. Genet.* 47, 335–352. doi: 10.1146/annurev-genet-111212-133519

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

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

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391

Tovkach, A., Ryan, P. R., Richardson, A. E., Lewis, D. C., Rathjen, T. M., Ramesh, S., et al. (2013). Transposon-mediated alteration of *TaMATE1B* expression in wheat confers constitutive citrate efflux from root apices. *Plant Physiol.* 161, 880–892. doi: 10.1104/pp.112.207142

Tyagi, S., Kabade, P. G., Gnanapragasam, N., Singh, U. M., Gurjar, A. K. S., Rai, A., et al. (2023). Codon usage provide insights into the adaptation of rice genes under stress condition. *Int. J. Mol. Sci.* 24, 1098. doi: 10.3390/ijms24021098

Wang, N., Li, C., Kuang, L., Wu, X., Xie, K., Zhu, A., et al. (2022). Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2206076119. doi: 10.1073/pnas.2206076119

Wei, L., Fei, Z., and Ding, Y. (2010). Mitochondrial RNA editing of ATPase *atp6* gene transcripts of Yunnan purple rice (*Oryza sativa* L.). *J. Wuhan Bot. Res.* 28, 251–256. doi: 10.3724/SP.J.1142.2010.30251

Wendel, J. F., Jackson, S. A., Meyers, B. C., and Wing, R. A. (2016). Evolution of plant genome architecture. *Genome Biol.* 17, 37. doi: 10.1186/s13059-016-0908-1

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PloS Comput. Biol.* 13, e1005595. doi: 10.1371/journal.pcbi.1005595

Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383

Wu, Z., Liao, X., Zhang, X., Tembrock, L. R., and Broz, A. (2022). Genomic architectural variation of plant mitochondria - A review of multichromosomal structuring. *J. Syst. Evol.* 60, 160–168. doi: 10.1111/jse.12655

Xiang, C., Gao, F., Jakovlić, I., Lei, H., Hu, Y., Zhang, H., et al. (2023). Using PhyloSuite for molecular phylogeny and tree-based analyses. *iMeta* 2, e87. doi: 10.1002/imt2.87

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E. J., and van der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319, 1527–1530. doi: 10.1126/science.1153040

You, C., Cui, T., Zhang, C., Zang, S., Su, Y., and Que, Y. (2023). Assembly of the complete mitochondrial genome of *Gelsemium elegans* revealed the existence of homologous conformations generated by a repeat mediated recombination. *Int. J. Mol. Sci.* 24, 527. doi: 10.3390/ijms24010527

Zandueta-Criado, A., and Bock, R. (2004). Surprising features of plastid *ndhD* transcripts: addition of non-encoded nucleotides and polysome association of mRNAs with an unedited start codon. *Nucleic Acids Res.* 32, 542–550. doi: 10.1093/nar/gkh217

Zhang, H., Meltzer, P., and Davis, S. (2013). RCircos: an R package for Circos 2D track plots. *BMC Bioinf.* 14, 244. doi: 10.1186/1471-2105-14-244

Zhang, S., Wang, J., He, W., Kan, S., Liao, X., Jordan, D. R., et al. (2023). Variation in mitogenome structural conformation in wild and cultivated lineages of sorghum corresponds with domestication history and plastome evolution. *BMC Plant Biol.* 23, 91. doi: 10.1186/s12870-023-04104-2

Zheng, D. (2011). *Cultivation Techniques for High Quality and High Yield of Wax Apple* (Beijing: China Agricultural Science and Technology Press).

# Frontiers in
# Plant Science

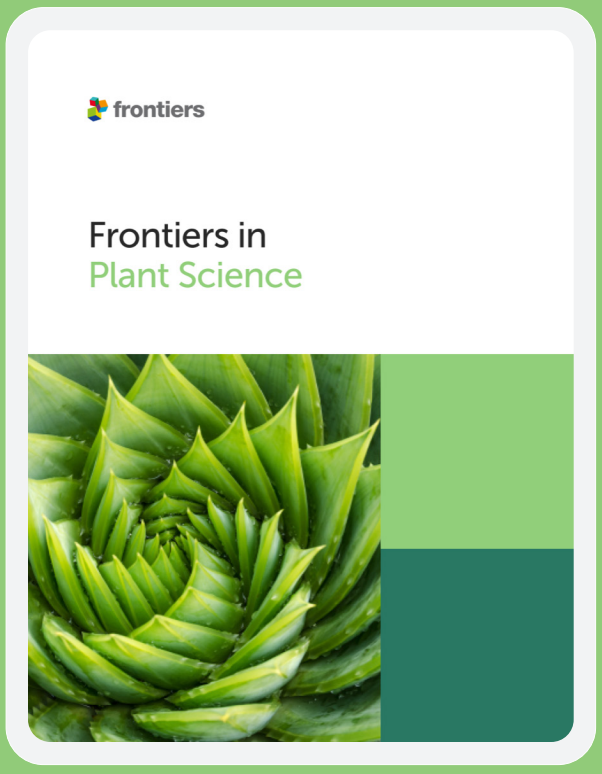
**Cultivates the science of plant biology and its applications**

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