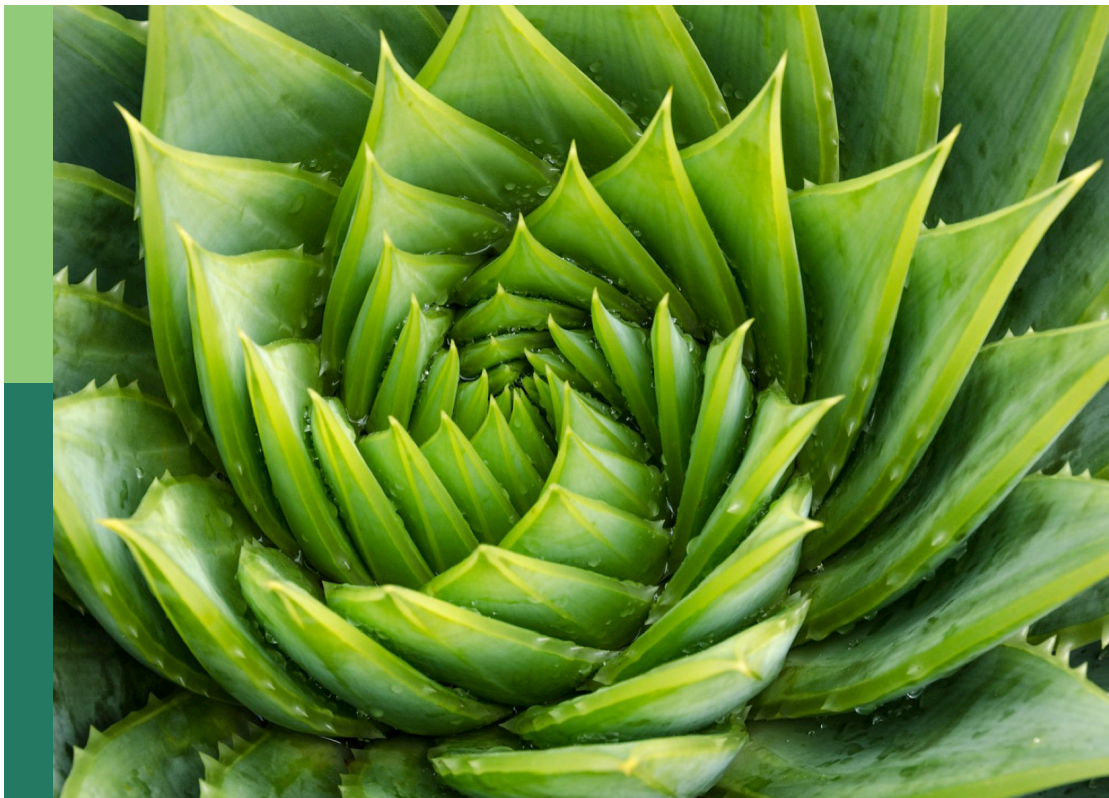
frontiers | Research Topics

# Omics and its integration: A systems biology approach to understanding plant physiology

**Edited by**
Kashif Ali and Milen I. Georgiev

**Published in**
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3978-1
DOI 10.3389/978-2-8325-3978-1

## About Frontiers

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

## Frontiers journal series

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

## Dedication to quality

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

## What are Frontiers Research Topics?

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

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

# Omics and its Integration: A Systems Biology Approach to Understanding Plant Physiology

**Topic editors**

Kashif Ali — Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Pakistan

Milen I. Georgiev — Stephan Angelov Institute of Microbiology, Bulgarian Academy of Sciences, Bulgaria

**Citation**

Ali, K., Georgiev, M. I., eds. (2023).*Omics and its integration: A systems biology approach to understanding plant physiology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3978-1

# Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 07 November 2023
DOI 10.3389/fpls.2023.1324901



OPEN ACCESS

EDITED AND REVIEWED BY
Roger Deal,
Emory University, United States

*CORRESPONDENCE
Kashif Ali
Kashif.ali@szabist.edu.pk

RECEIVED 20 October 2023
ACCEPTED 01 November 2023
PUBLISHED 07 November 2023

CITATION
Ali K and Georgiev MI (2023) Editorial: Omics and its integration: a systems biology approach to understanding plant physiology.
*Front. Plant Sci.* 14:1324901.
doi: 10.3389/fpls.2023.1324901



# Editorial: Omics and its integration: a systems biology approach to understanding plant physiology

Kashif Ali[1]* and Milen I. Georgiev[2]

[1]Department of Biosciences, Faculty of Life Sciences, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi, Pakistan, [2]Institute of Microbiology, Bulgarian Academy of Sciences, Plovdiv, Bulgaria

KEYWORDS

**omics, genomics, transcriptomics, proteomics, metabolomics, systems biology**

Editorial on the Research Topic
Omics and its integration: a systems biology approach to understanding plant physiology

The comprehension of intricate and interconnected biological processes necessitates the adoption of an integrative systems biology approach. The incorporation of diverse omics data, encompassing genomes, proteomics, transcriptomics, and metabolomics, will yield a comprehensive perspective that elucidates the interplay between different biomolecules and their roles in plant development, resilience, characteristics, and other significant attributes. The integration of multi-omics data has been made possible by recent advancements in analytical capabilities and the processing of massive data sets. As a result, researchers now possess a significantly enhanced understanding of the cell, tissue, organ, or even the entirety of an organism under investigation.

The primary objective of this Research Topic is not solely limited to the dissemination of the omics data that is of superior quality and dependability, but also encompasses the broader inquiry into the methods of integrating and establishing correlations amongst diverse data sets derived from numerous omics investigations. The studies included in this topic are focused on the areas of integration of multiple omics data for discovering gene(s) functions in autophagy, circadian clock and its connection with response against herbicides, resistance in plants, advancements in single cell and spatial transcriptomics, study of cultivars based on chemotype, and the interaction of proteins with the metabolites.

The emergence of single-cell and spatial transcriptomics has shifted the focus of researchers from the study of multicellular systems to the exploration of single-cell phenomena and spatial information. The utilization of single-cell transcriptomes enables the examination of transcriptomic profiles at the individual cell level, facilitating a deeper understanding of cellular heterogeneity. Conversely, spatial transcriptomes have the advantage of preserving and analyzing the spatial organization of gene expression patterns within tissues or organisms. Despite the usefulness and maturity of these two omics methods, additional study is required to assure their broad usage in plant studies. In a comprehensive review, Chen et al. conducted a comparative analysis of experimental methodologies employed in several plant species, focusing specifically on plant single-cell

or spatial transcriptomics. The constraints and difficulties associated with single-cell and spatial transcriptome investigations are readily apparent, including restrictions in terms of application, geographical information, and resolution. The review also proposes additional applications, including the examination of roots at the single-cell level through cross-species analysis and asserts that the integration of single-cell transcriptome analysis with other omics studies is necessary to surpass the capabilities of individual omics analyses.

Discovering gene functions has been argued to be possible through the use of a systems biology technique, which combines molecular data from many levels of genome expression (*i.e.*, multi-omics data). Ding et al. analyzed the effects of mutations in two AuTophaGy-related (ATG) genes of Arabidopsis by combining lipidomics, metabolite mass-spectral imaging, and transcriptomics data from leaves and roots. The autophagy process, which is necessary for the degradation and recycling of macromolecules and organelles, is inhibited in the atg7 and atg9 mutants studied here. A prior knowledge of the exact biochemical function of the relevant proteins greatly facilitates the development of a comprehensive physiological model.

Plants have developed circadian clock mechanisms that facilitate the synchronization of biological processes with cyclic variations in the surrounding environment. Nevertheless, the interaction between the circadian clock and the herbicide response in rice remains unexplored. An interesting study by Chen et al. employed multi-omics data to elucidate the interplay between the circadian clock and herbicide response mechanisms in rice, specifically focusing on the epigenomic, transcriptomic, and metabolomic levels. Ultimately, it has been determined that the use of herbicides has the potential to impact the expression of several key oscillator genes inside the circadian clock of rice. This study can provide a theoretical framework for the development of novel herbicides or the cultivation of herbicide-resistant crops.

An interesting study by Zhang et al. capitalizes on the power of genomics and metabolomics for screening endophytic fungi with antibacterial potential, in *Rosa roxburghii* Tratt. This study identified 54 endophytes from *R. roxburghii* using molecular identification. Genomic, non-target metabolomic, and comparative genomic investigations examined the screened-out endophytic fungus' biosynthesis.

Commercial distillation of mint produces essential oils, which are used in many consumer products. Most study on terpenoid oil constituent development has focused on leaves. Lange et al. found that mint species, including peppermint (*Mentha* $^{x}$ *piperita* L.), spearmint (*Mentha spicata* L.), and horsemint (*Mentha longifolia* (L.) Huds.), emit volatiles from stems, rhizomes, and roots. The terpenoid volatile composition of these organs differs significantly from leaves, suggesting significant chemical diversity exists. The newly sequenced *M. longifolia* CMEN 585 genome was used to identify candidate genes for monoterpene synthases (MTSs), which catalyze the first step in the biosynthesis of monoterpenoid volatiles. This was done to investigate the genetic and biochemical basis of chemical diversity. These MTSs can contribute to the production of all main monoterpene skeletons seen in volatiles from various mint organs.

Wheat is well recognized as a prominent staple crop within the global food production system. Nevertheless, the presence of stripe rust fungus has a substantial negative impact on both the productivity and quality of wheat. The findings of a study by Liu et al. provide valuable insights into the regulatory networks involved in the interactions between wheat and *Puccinia striiformis* f. sp. *tritici* (Pst). These insights have the potential to facilitate the development of durable resistance breeding strategies in wheat, which could contribute to addressing global environmental and food challenges. The findings of the study indicated that the infection caused by Pst led to the upregulation of genes and metabolites associated with the manufacture of phenylpropanoids. The unique resistance exhibited by R88 is controlled by the selective expression of genes that are involved in the precise regulation of interactions between wheat and Pst. In addition, the examination of the metabolome indicated that the accumulation of metabolites associated to lignin production was significantly influenced by Pst.

A major obstacle to widespread use of plant fibers is understanding the molecular components that affect the mechanical properties of both elementary and scutched fibers. Chabi et al. studied genome-wide transcription profiling in bast fiber-bearing tissues of seven flax cultivars and found 1041 differentially expressed genes, with 97 connected to cell wall metabolism. Three distinct clusters were identified based on gene expression levels for 32 proteins, out of which 4 proteins significantly linked with morphometric characteristics. These findings identify molecular players involved in determining in-planta fiber morphometrics and ex-planta fiber mechanical characteristics, which are crucial for determining basic and scutched fiber quality in flax.

Fernandes et al. assessed the phenotypic and chemotype of three high-THCA and one high-CBDA cultivars in their diploid, triploid, and tetraploid states using agronomic and metabolomic methods. This study reports that plant morphology showed a cultivar-dependent increase in height and leaf size with higher ploidy levels. Higher ploidy levels severely impacted cannabinoids, with concentrations of total cannabinoids, THCA, CBDA, and CBGA dropping significantly across all four cultivars. Overall, plant morphology observations match the giga phenotype in other polyploid plant species. This study indicated that tetraploidization can boost *Cannabis sativa* L. therapeutic potential, although cultivar and genotype-dependently. This effort prepares the breeding, biotechnological, and pharmaceutical industries to improve, evaluate, and use *C. sativa* chemical diversity.

Chlorophylls and carotenoids, which are generated from isoprenoid precursors, are needed for photosynthesis and plant greening. A study by Wolters et al. used *in silico* analysis, gene overexpression, transcriptomics, and metabolic profiling to study the impact of a homologous protein in the Russian dandelion (*Taraxacum koksaghyz*) on its rich isoprenoid network. Research suggests that TkPEL-like negatively controls genes involved to photosynthesis and chlorophyll in a light-dependent way, which is conserved across animals. These findings will guide future research on leaf isoprenoid biosynthesis and photomorphogenesis regulation, enabling breeding techniques to improve plant isoprenoid profiles and create plant-based industrial platforms.

The Guest Editors would like to express their gratitude to all the authors and reviewers of this Research Topic, and acknowledge their hard work and dedication toward the area of omics and systems biology. The Guest Editors believe that the presented researches will encourage the generation of more knowledge and valuable research in the fields of omics, their integration, and systems biology.

## Author contributions

KA: Writing – original draft. MG: Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

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

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

## Publisher's note

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 November 2022
DOI 10.3389/fpls.2022.1060478



OPEN ACCESS

EDITED BY
Christopher Rensing,
Fujian Agriculture and Forestry University, China

REVIEWED BY
Shihong Luo,
Shenyang Agricultural University, China
Vishnu Sukumari Nath,
Donald Danforth Plant Science Center, United States

*CORRESPONDENCE
Yu-Lan Jiang
yljchsd@163.com

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

RECEIVED 03 October 2022
ACCEPTED 28 October 2022
PUBLISHED 16 November 2022

CITATION
Zhang H, Yang M-F, Zhang Q, Yan B and Jiang Y-L (2022) Screening for broad-spectrum antimicrobial endophytes from *Rosa roxburghii* and multi-omic analyses of biosynthetic capacity. *Front. Plant Sci.* 13:1060478. doi: 10.3389/fpls.2022.1060478



# Screening for broad-spectrum antimicrobial endophytes from *Rosa roxburghii* and multi-omic analyses of biosynthetic capacity

Hong Zhang[1,2], Mao-Fa Yang[3,4], Qian Zhang[1], Bin Yan[3] and Yu-Lan Jiang[1]*

[1]Department of Plant Pathology, College of Agriculture, Guizhou University, Guiyang, China, [2]Guizhou Academy of Testing and Analysis, Guiyang, China, [3]Institute of Entomology, Guizhou University, Guiyang, China, [4]College of Tobacco Science, Guizhou University, Guiyang, China

Plants with certain medicinal values are a good source for isolating function-specific endophytes. *Rosa roxburghii* Tratt. has been reported to be a botanical source of antimicrobial compounds, which may represent a promising candidate for screening endophytic fungi with antimicrobial potential. In this study, 54 endophytes were isolated and molecularly identified from *R. roxburghii*. The preliminary screening using the plate confrontation method resulted in 15 different endophytic strains showing at least one strong inhibition or three or more moderate inhibition against the 12 tested strains. Further re-screening experiments based on the disc diffusion method demonstrated that *Epicoccum latusicollum* HGUP191049 and *Setophoma terrestris* HGUP190028 had excellent antagonistic activity. The minimum inhibitory concentration (MIC) test for extracellular metabolites finally indicated that HGUP191049 had lower MIC values and a broader antimicrobial spectrum, compared to HGUP190028. Genomic, non-target metabolomic, and comparative genomic studies were performed to understand the biosynthetic capacity of the screened-out endophytic fungus. Genome sequencing and annotation of HGUP191049 revealed a size of 33.24 megabase pairs (Mbp), with 24 biosynthetic gene clusters (BGCs), where the putative antimicrobial compounds, oxyjavanicin, patulin and squalestatin S1 were encoded by three different BGCs, respectively. In addition, the non-targeted metabolic results demonstrated that the strain contained approximately 120 antimicrobial secondary metabolites and was structurally diverse. Finally, comparative genomics revealed differences in pathogenicity, virulence, and carbohydrate-active enzymes in the genome of *Epicoccum* spp. Moreover, the results of the comparative analyses presumed that *Epicoccum* is a promising source of

antimicrobial terpenes, while oxyjavanicin and squalestatin S1 are antimicrobial compounds shared by the genus. In conclusion, *R. roxburghii* and the endophytic HGUP191049 isolated from it are promising sources of broad-spectrum antimicrobial agents.

KEYWORDS

biological activity, *Epicoccum*, genomics, non-target metabolomics, comparative genomics, *Epicoccum latusicollum*, *Setophoma terrestris*

## 1 Introduction

Endophytic fungi are generally recognized as a group of microorganisms that do not cause substantial damage to the host and live harmlessly in healthy plant tissues throughout a certain life cycle stage (Yuan et al., 2018). The existence of fungi inside the tissues of healthy plants has been known as early as the late 19th century when endophytic fungi were first successfully isolated from darnel (*Lolium temulentum*) (Freeman, 1904; Kusari et al., 2012). However, plant endophytic fungi had not attracted much attention as a new microbial resource until 1993. When an endophytic fungus, Ceriporiopsis andreanae (basionym Taxomyces andreanae), was isolated from Taxus brevifolia for the production of taxol, which initiated a surge in studies on endophytes (Stierle et al., 1993; Cheng et al., 2022). Another excellent anticancer drug, vincristine is originally reported from *Catharanthus roseus*, endophytic *Fusarium oxysporum* isolated from this plant can also produce vinblastine and vincristine in appreciable amounts (Kumar et al., 2013). The herb *Artemisia annua* L. is well known for its antimalarial properties and is the source of the antimalarial drug artemisinin (Madsen et al., 2010). Extracts of both endophytic *Penicillium* and *Talaromyces* isolated from *A. annua* exhibited significant antimalarial activity (Alhadrami et al., 2021). Therefore, endophytic fungi can produce the same natural products as their host plants. Importantly, microbial fermentation has several advantages over the use of plants for the production of bioactive substances, such as easy-to-operate, reducing the need for plants, and obtaining stronger active drug derivatives by varying the culture conditions (Kumar et al., 2013).

Various endophytic fungi have been employed in recent years to produce bioactive compounds, such as *Aspergillus*, *Epicoccum*, *Hypoxylon*, *Induratia*, *Penicillium*, *Phoma*, *Phaeosphaeria*, *Saccharomycopsis*, *Sarocladium*, *Trichoderma*, and *Wickerhamomyces*. The biologically active secondary metabolites deriving from endophytic fungi belong to diverse structural classes. The secondary metabolites include alkaloids, anthraquinones, polyketides, sterols, terpenes, and volatile organic compounds (Zhang et al., 2021a). They possess potent antimicrobial, antiviral, insecticidal, antioxidative, antidiabetic, cytotoxic, and anticancer properties (Deshmukh et al., 2015; Zhang et al., 2019; Fernando et al., 2020; Manganyi and Ateba, 2020; Pal et al., 2020; Rahaman et al., 2020; Agrawal et al., 2022). A few endophytic fungi can produce phytohormones to promote the growth of their host plants. And synthesize bioactive compounds to increase the resistance of the plants to environmental stresses. Still, they can also promote the accumulation of secondary metabolites initially produced by the plant, including pharmaceutical ingredients (Jia et al., 2016).

Research on endophytic fungi has become more accessible with the continual advancement of sequencing and omics technologies. Genetics- and genomics-based strategies have emerged as a comprehensive approach to studying natural microbial products (Walker et al., 2020). It is possible to elucidate the basic pathways of secondary metabolites isolated from organisms using these technologies. These technologies can facilitate the computational discovery of biosynthetic pathways. Producer strains for biosynthesis are investigated, silenced biosynthetic gene clusters are activated, and synthetic pathways for novel compounds are designed to increase their yields and activity (Sagita et al., 2021). Comparative multi-genome analysis significantly improves understanding of the genetic and metabolic diversity of endophytic fungi involved in different host-plant interactions (Ye et al., 2017). Additionally, the putative functional characteristics of endophytes can be elucidated by metagenome-based analyses (Gupta et al., 2020). The rapid development of omics technologies has accelerated the development of endophytic fungal resources.

*Rosa roxburghii* Tratt., a homology of medicine and food, has received considerable attention across many research fields because of its notably high vitamin C. Various phytochemicals extracted from its fruits, roots, and leaves have shown potential antimicrobial activity. When choline chloride with lactic acid or levulinic acid (molar ratio 1:2) as deep eutectic solvents, the extracts of *R. roxburghii* leaves showed incredible antibacterial activities against the five tested pathogens (*Bacillus subtilis*, *Escherichia coli*, *Listeria monocytogenes*, *Salmonella typhimurium*, and *Staphylococcus aureus*), with the minimum inhibition concentration (MIC) values ranging from 0.012 to 0.049 mg/mL

(Wang et al., 2021). Strictinin isomers, separated from the root of this plant, were excellent antimicrobial components, mainly responsible for oxidative stress and protein synthesis disorder (Ma et al., 2020). Since *R. roxburghii* is a botanical source of antimicrobial compounds, it may represent a promising target for screening endophytic fungi with antimicrobial potential. We isolated endophytic fungi from *R. roxburghii* and screened the most prospective strain by antimicrobial activity assays. Additionally, we investigated the strain's biosynthetic capacity through genomics, non-targeted metabolomics, and comparative genomics.

# 2 Materials and methods

## 2.1 Isolation and identification

### 2.1.1 Sample collection and endophyte isolation

Healthy *R. roxburghii* tissues (roots, stems, leaves, flowers, fruits, and seeds) were collected from April to August 2020 in Guizhou Province (27°4′50″ N, 106°29′50″ E and 25°52′52″ N, 104°33′59″ E), China. Endophytic fungi were isolated from different tissue parts using a surface sterilization method (Wang et al., 2019). The main steps of the procedure: Tissue segments were surface-sterilized with 75% ethanol for 1 min, rinsed thrice with sterile water, immersed in 1% (w/v) aqueous sodium hypochlorite (NaOCl) for 1–3 minutes (roots, 2 min; stems and seeds, 3min; and leaves, fruits, and flowers, 1 min), and washed thrice again with sterile distilled water. Six different media were used for fungal isolation, namely, potato dextrose agar (PDA), oatmeal agar (OA), malt extract agar (MEA), Czapek Dox agar (CDA), water agar (WA), and synthetic low nutrient agar (SNA). The media were supplemented with streptomycin sulphate (0.5 g/L) to avoid bacterial contamination. Meanwhile, the effectiveness of surface sterilization was examined according to the previous description (Singh et al., 2017; Rojas et al., 2020). All pure isolates were stored at -80°C with 30% glycerol.

### 2.1.2 Molecular identification

DNA was extracted from mycelia grown on potato dextrose agar (PDA) according to the manufacturer's instructions for a Fungal gDNA Isolation Kit (BW-GD2416, Biomiga, China). The primers used for polymerase chain reaction (PCR) amplification and sequencing included ITS5/ITS4 for ITS (White et al., 1990), LR0R/LR5 for LSU (Vilgalys and Hester, 1990), Bt2a/Bt2b for TUB (Glass and Donaldson, 1995), and fRPB2-5F/fRPB2-7cR for RPB2 (Liu et al., 1999). Successful amplification is generally obtained by annealing at 55°C for 35 cycles. The PCR products were sequenced by Sangon Biotech (Shanghai, China).

Endophytic fungi were identified based on multigene phylogenetic analyses. Consensus sequences were edited with BioEdit v. 7.0.9.0 (Hall, 1999). Multiple sequence alignment was performed using MAFFT v. 7 (Katoh et al., 2019), manually adjusted in BioEdit, and concatenated in PhyloSuite v. 1.2.2 (Zhang et al., 2020). Phylogenetic analyses were inferred from maximum likelihood (ML) and Bayesian inference (BI). ModelFinder determined the substitution models based on the Bayesian Information Criteria (BIC) and Akaike information criterion (AIC) (Kalyaanamoorthy et al., 2017). BIC was used for ML analyses, while AIC was used for BI analyses. ML tree inference was constructed using 10,000 ultrafast bootstraps (Minh et al., 2013) under the edge-linked partition model implemented in IQ-TREE (Nguyen et al., 2014). BI analyses were carried out in MrBayes 3.2.6 (Ronquist et al., 2012) under the partition models, with two independent runs of four chains that were run for five million generations using the Markov chain Monte Carlo algorithm. Finally, the resulting trees were visualized using Figtree v.1.4.3 (Rambaut, 2014).

## 2.2 Antimicrobial activity

### 2.2.1 Tested strains

To evaluate the antimicrobial activity of endophytic fungi, the following microorganisms were used: Six tested fungi, including kiwifruit soft rot pathogens *Lasiodiplodia theobromae* and *Botryosphaeria dothidea*, pepper anthracnose fungus *Colletotrichum capsici*, rice blast fungus *Pyricularia oryzae*, rice sheath blight fungus *Rhizoctonia solani*, and root rot fungus *Fusarium oxysporum* (causing *Pseudostellaria heterophylla* and *Zanthoxylum schinifolium* diseases). Six tested bacteria, namely kiwifruit bacterial canker pathogen *Pseudomonas syringae* pv. *actinidiae*; peach bacterial shot hole pathogen *Pantoea agglomerans*; other bacteria *Bacillus subtilis* CMCC (B) 63501, *Escherichia coli* CMCC (B) 44102, *Pseudomonas aeruginosa* ATCC 27853, and *Staphylococcus aureus* ATCC 6538.

### 2.2.2 Preliminary screening of antimicrobial activity assay

Endophytic strains with antagonistic ability were screened out by the plate confrontation method (Gao et al., 2021). The width of the zone of inhibition ($I$) between tested fungi (or bacteria) and endophytes was determined according to the previously described method (Gao et al., 2021). The definition of the inhibition intensity is based on the previously described method (Gashgari et al., 2016; Zhao et al., 2019). The intensity is divided into four levels, which are indicated by 0, 1, 2, and 3 for no inhibition, weak inhibition, moderate inhibition, and strong inhibition, respectively. For evaluating antifungal activity: 0 ($I$ = 0 mm), 1 (0 mm < $I \leq$ 1 mm), 2 (1 mm < $I \leq$ 3 mm), and 3 ($I >$ 3 mm); for antibacterial activity: 0 ($I \leq$ 1 mm), 1 (1 mm < $I \leq$ 2 mm), 2 (2 mm < $I \leq$ 10 mm), and 3 ($I >$ 10 mm). Through phylogenetic analyses and preliminary screening, strains were selected for re-screening antimicrobial activity under the

following principles: with the best inhibition effect in the same species and strong inhibition of at least one or moderate inhibition of three or more against the tested strains.

### 2.2.3 Secondary metabolites extraction

To further investigate the antimicrobial activity of the initially screened-out strains, crude extracts of the secondary metabolites were prepared using the method described previously (Zhang et al., 2021b). The endophytic strains were fermented in Erlenmeyer flasks (250 mL) containing 100 mL potato dextrose broth (PDB) (potato: 200 g/L, glucose: 20 g/L, and natural pH) at 28 ± 1°C, 220 rpm, and for 7–10 d. High-speed centrifugation (14,000 g, 10 min) was performed to separate the culture broth and mycelium, which were extracted by ethyl acetate (EtOAc) and methanol (MeOH)-assisted sonication, respectively. Then concentrated at 50°C under reduced pressure until constant weight and dissolved in dimethyl sulfoxide (DMSO) to obtain 20 mg/mL of EtOAc crude extract (extracellular metabolites) and MeOH crude extract (intracellular metabolites).

### 2.2.4 Re-screening of antimicrobial activity assay

The extracellular and intracellular metabolites were re-screened for antimicrobial activity using the disc diffusion method (Hu et al., 2017; Rjeibi et al., 2020). For antifungal assay: Briefly, a tested fungal plug (6 mm diam.) and a same-sized sterile filter paper disc were placed at the appropriate position of the PDA plate (90 mm diam.). The disc was impregnated with 10 µL of metabolite (20 mg/mL). DMSO was used as a negative control. All plates were incubated at 28 ± 1°C. The radial growth of the tested strains was measured after 2–7 d. Negative control plates as $R_1$ and experimental plates containing metabolites as $R_2$. The percentage inhibition (%) = $(R_1\text{-}R_2)/R_1 \times 100\%$ (Hajieghrari et al., 2008).

For antibacterial assay: The sterile disc (6 mm diam.) was placed at the center of the nutrient agar (NA) plate, which had been coated with tested bacteria, and then impregnated with 10 µL of extracellular or intracellular metabolite (20 mg/mL). Equal volumes of DMSO were used as a negative control. The diameters of the inhibition zone (d) were measured after culturing for 24–48 hours at 25°C ± 1°C for phytopathogenic bacteria and 35°C ± 1°C for other tested bacteria. Similarly, the MICs were determined. The assays were repeated three times.

## 2.3 Morphological observations

Morphological characteristics were observed on PDA. In this study, the final screened-out strain with antimicrobial potential was inoculated on PDA, cultured at 28°C for 5–7 days, and then placed at 4°C for preservation to promote sporulation. Macroscopic morphology was examined under a digital microscope (VHX-7000, Keyence). After sporulation, micromorphological features and dimensions of the spores were determined in 25% lactic acid under a Zeiss Axiolab 5 light microscope equipped with an Axiocam 208 camera.

## 2.4 Genome-sequencing, annotation, and analyses

Strain with the strongest antimicrobial activity from the re-screening was selected for whole-genome sequencing to deeply analyze its biosynthetic capacity. Genomic DNA was sequenced using a combination of second-generation Illumina sequencing technologies and third-generation PacBio sequencing technology at Guangzhou Genedenovo Biotechnology Co., Ltd. The endophytic strain was grown in a 1 L Erlenmeyer flask containing 500 mL of PDB at 28°C under 220 rpm for three days. The fermentation broth was centrifuged at 14,000 g for 10 min at 4°C, the supernatant was discarded, and the mycelium was collected and used for genomic DNA extraction. Genomic DNA was extracted using commercial kits, and DNA quality was assayed using Qubit (Thermo Fisher Scientific, Waltham, MA) and Nanodrop (Thermo Fisher Scientific, Waltham, MA). Qualified genomic DNA was fragmented with G-tubes (Covaris, Woburn, MA, USA) and end-repaired to prepare SMRTbell DNA template libraries with a fragment size of >10 Kb. Then, library quality was detected by Qubit® 2.0 Flurometer (Life Technologies, CA, USA), and average fragment size was estimated on a Bioanalyzer 2100 (Agilent, Santa Clara, CA). Subsequently, SMRT sequencing was performed on the Pacific Biosciences Sequel sequencer (PacBio, Menlo Park, CA) following standard protocols (MagBead Standard Seq v2 loading, 1 × 180 min movie) with the P4-C2 chemistry.

Continuous long reads attained from SMRT sequencing were corrected for random errors in the long seed reads (seed length threshold 6 Kb) by aligning shorter reads from the same library using MECAT. The resulting corrected, preassembled reads were used for *de novo* assembly using MECAT with an overlap-layout-consensus (OLC) strategy (Myers et al., 2000; Xiao et al., 2017). The Open reading frame (ORF) was predicted using the GeneMark-ES (Ter-Hovhannisyan et al., 2008). Repetitive elements were identified by RepeatMasker (Chen, 2004). Noncoding RNAs, such as rRNAs prediction, were carried out using RNAmmer (Lagesen et al., 2007), and tRNAs were identified by tRNA-scan-SE (Lowe and Eddy, 1997).

Functional annotation of predicted protein-coding genes against National Center for Biotechnology Information (NCBI) non-redundant Protein (Nr) database, Gene Ontology (GO), eukaryotic orthologous groups (KOG), Kyoto Encyclopedia of

Genes and Genomes (KEGG), and SwissProt databases were conducted by the BlastP method. Moreover, the assembled genome sequence was analyzed for secondary metabolite biosynthesis gene clusters (BGCs) using antiSMASH 6.1.1.

## 2.5 Non-targeted metabolomics analyses

The PDB fermentation conditions for endophytic fungi were identical to the genomic sequencing assay. The fermentation broth of the endophyte was separated by high-speed centrifugation (14,000 g, 10 min). We selected extracellular or intracellular metabolites with better inhibitory activity for non-target metabolomics analyses based on the results of the re-screening assay. The culture broth samples were thawed at 4°C, and 100 µL aliquots were mixed with 400 µL of cold methanol/acetonitrile/$H_2O$ (2:2:1, v/v/v). Following vortex mixing, low-temperature sonication for 30 min and resting for 10 min at -20°C. After that, the mixture was centrifuged for 20 min (14,000 g, 4°C). The supernatant was dried in a vacuum centrifuge. The samples were re-dissolved in 100 µL acetonitrile/water (1:1, v/v) for LC-MS analyses. Analyses were performed using a UHPLC (1290 Infinity LC, Agilent Technologies) coupled to a quadrupole time-of-flight (AB Sciex Triple TOF 6600). The chromatographic separation and The ESI source conditions were as previously described (Huang et al., 2019).

## 2.6 Comparative genomics analyses

Relevant genomic data released by NCBI were selected for comparative genomic profiling. Three *de novo* gene prediction programs, Augustus v.2.7, GeneMark+ES v.4.0, and SNAP v.2013-02-16, were used to predict the protein-coding regions if only genomic data were available in NCBI. The Maximum likelihood tree of genomes was performed using single-copy orthologous genes. Pathogen-host interaction (PHI), carbohydrate-active enzymes (CAZymes), and BGCs were annotated using PHI-base v. 4.13, dbCAN2 v. 11, and antiSMASH v. 6.1.1, respectively. Genes encoding BGCs were aligned using MAFFT v. 7, the substitution model was determined by ModelFinder, and ML tree inference was performed in IQ-TREE using 10,000 ultrafast bootstraps. Finally, the annotation results of each genome were compared and analyzed.

## 2.7 Data analysis

Data were analyzed by ANOVA, followed by comparisons of means using the LSD test in Data Processing System (DPS v9.50) ($P < 0.05$).

# 3 Results

## 3.1 Identification of endophytic fungi

All culturable endophytic strains were sequenced and used for multigene phylogenetic analyses. 54 strains belonging to Dothideomycetes, Eurotiomycetes, Pezizomycetes, Leotiomycetes, and Agaricomycetes were successfully isolated and identified from tissue segments of *R. roxburghii*. The largest number of endophytes was found in root tissues (20 isolates), followed by the stem (14 isolates), leaf (9 isolates), fruit (6 isolates), seed (4 isolates), and flower (one isolate). Of these isolates, 51 strains were identified at the species level, covering 28 confirmed species. The remaining genus-level isolates, including two unidentified species, may belong to new taxa. The phylogenetic relationship was constructed with combined ITS, LSU, TUB, and RPB2, as illustrated in Figure 1A.

## 3.2 Preliminary screening results

Preliminary screening results of the 54 isolates for antimicrobial activity *in vitro* were represented in Figure 1B and Supplementary Table 1. In this study, the antibacterial activity of endophytic fungi was superior to the antifungal activity. The antimicrobial activity may be strain-specific owing to significant differences observed among strains of the same endophytic species, such as *Alternaria tenuissima*, *Emmia latemarginata*, and *Neofusicoccum* sp. Most endophytes exhibited broad-spectrum activities, whereas another small group did not display any antimicrobial activity, e.g., *Macrophomina phaseolina* and *Paraphoma vinacea*. Concretely, endophytes showed stronger antibacterial activity against *Ba. Subtilis*, and hardly any activity were observed against *Py. oryzae*. Although generally described as pathogens, some species as endophytic fungi also demonstrated potential antimicrobial activity, e.g., *Al. tenuissima* HGUP191067. In general, 15 endophytes were selected for subsequent experiments based on molecular identification and the strength of inhibition activity (Figure 1).

## 3.3 Re-screening of antimicrobial activity

### 3.3.1 Re-screening of antifungal activity

As observed from the trends of Figure 2 and Supplementary Table 1. These extracellular and intracellular metabolites, which were prepared from the preliminary screened-out strains, showed broad-spectrum antifungal activity against at least one phytopathogenic fungus. However, most metabolites exhibited less than 20% inhibition against the six tested fungi. Fortunately, endophytic strains *Epicoccum latusicollum* HGUP191049, *Neofusicoccum* sp. HGUP191080, and *Setophoma terrestris* HGUP190028 displayed potential as antifungal agents since their metabolites displayed over 30% inhibition rate against at least one of

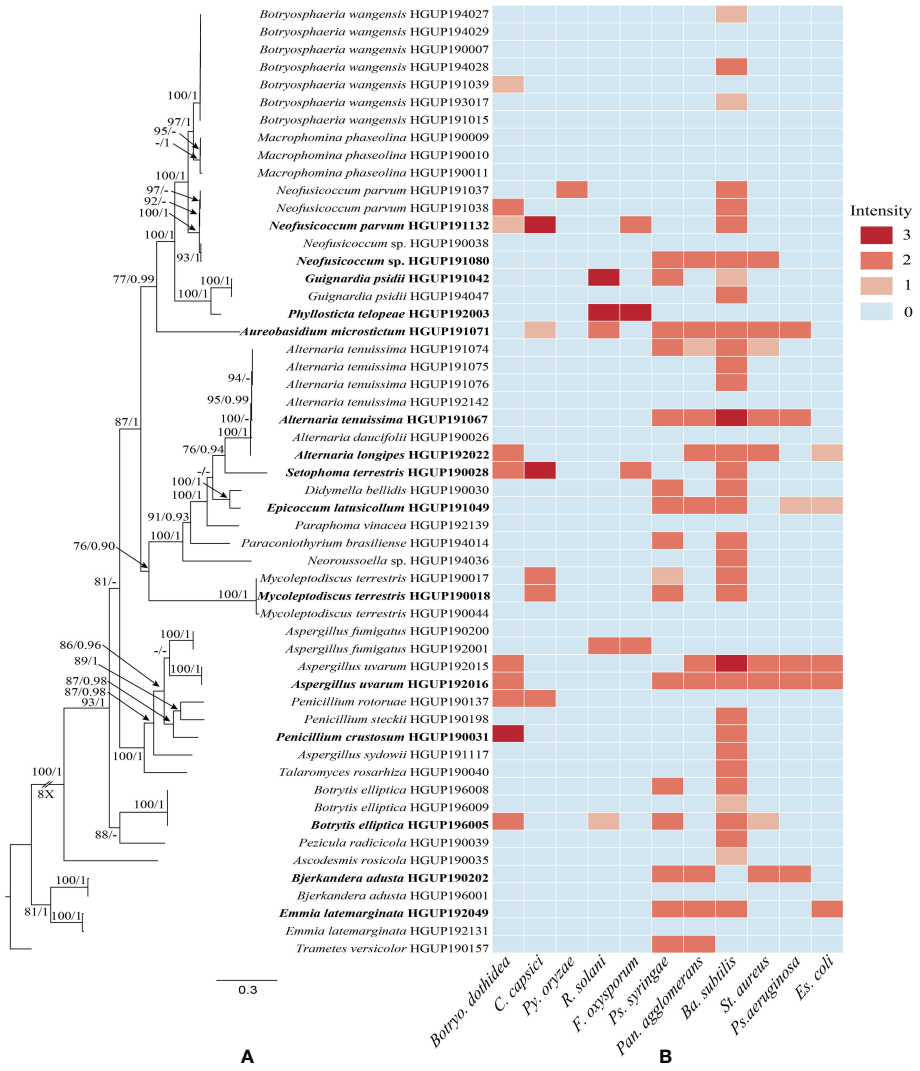



FIGURE 1
Phylogenetic relationships and preliminary screening for antimicrobial activity of endophytic fungi. **(A)** Phylogram generated from maximum likelihood (ML) analyses, based on combined ITS, LSU, TUB, and RPB2 sequence data. Bootstrap support values for ML greater than 75% and Bayesian posterior probabilities greater than 0.90 are given near nodes, respectively. Bold indicates strains that have been preliminarily screened out. **(B)** Heatmap of antimicrobial activity spectra against the tested strains.

the tested fungi. Of these, the extracellular metabolites of HGUP191049 and HGUP190028 were highly effective against *L. theobromae* with inhibition rates was 58.5 ± 3.4% and 51.4 ± 3.4%, respectively. Meanwhile, the inhibition rate of HGUP191049 also reaches 58.0 ± 2.2% against *Botryo. dothidea* and 45.3 ± 1.3% against *C. capsici*. Accordingly, *Ep. latusicollum* HGUP191049 holds good promise for developing antifungal agents.

### 3.3.2 Re-screening of antibacterial activity

As can be derived from Figure 3 and Supplementary Table 1, most secondary metabolites showed sub-moderate inhibition intensity (d < 10 mm). However, the extracellular metabolite of *Ep. Latusicollum* HGUP191049 was shown to have potent antibacterial activity against both tested Gram-positive (*St. aureus* and *Ba. subtilis*) and Gram-negative (*Ps. syringae* pv. *actinidiae*, *Es. coli*, and *Ps. aeruginosa*) bacteria, since inhibition zone diameters ranging from 15.3 ± 1.5 mm to 20.3 ± 2.5 mm. So the strain HGUP191049 was considered to be well antagonistic.

### 3.3.3 Determination of the MIC

Two endophytic isolates *Ep. latusicollum* HGUP191049 and *Se. terrestris* HGUP190028 had a better antimicrobial effect in the re-screening assay based on a broad spectrum and intensity. To evaluate the antimicrobial potential of the extracellular metabolites of the two isolates, in which MIC values were determined. As shown in Table 1, *Ep. latusicollum* HGUP191049

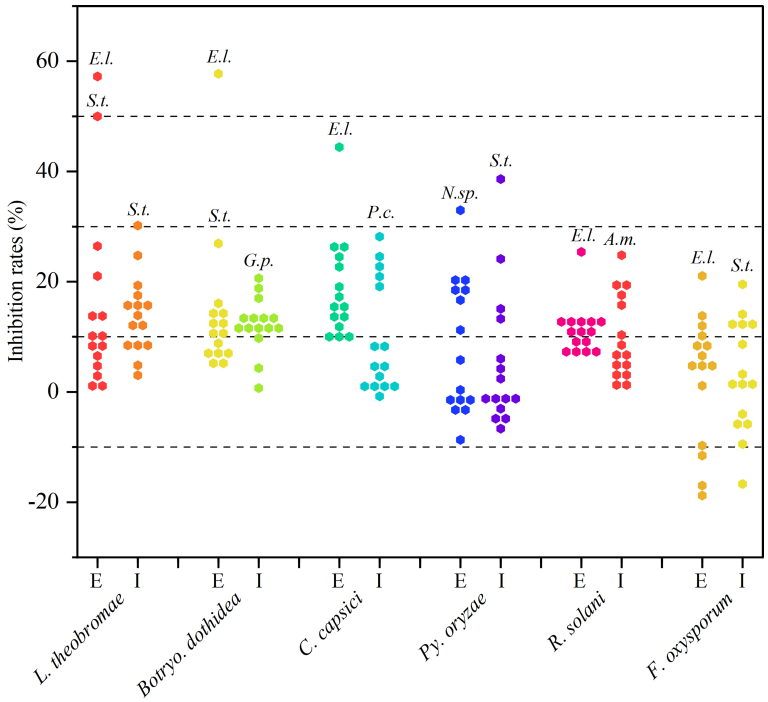



FIGURE 2
Results of re-screening for antifungal activity. E and I indicate extracellular and intracellular metabolites, respectively. *A.m.*, *E.l.*, *G.p.*, *N.sp.*, *P.c.*, and *S.t.*, represent *Aureobasidium microstictum* HGUP191071, *Epicoccum latusicollum* HGUP191049, *Guignardia psidii* HGUP191042, *Neofusicoccum* sp. HGUP191080, *Penicillium crustosum* HGUP190031, and *Setophoma terrestris* HGUP190028, respectively.

presented MIC values of 1.25 mg/mL, 2.50 mg/mL, and 1.25 mg/mL against *L. theobromae*, *Botryo. Dothidea*, and *R. solani*, respectively, whereas MIC values ranged from 0.31 mg/mL to 5.00 mg/mL against six tested bacteria. The endophytic strain HGUP191049 had more potential for antimicrobial properties than the strain HGUP190028.

## 3.4 Taxonomy of *Epicoccum latusicollum*

Sexual morph not observed. Asexual morph (Figure 4): Conidiomata pycnidial, aggregated, superficial, black, globose to subglobose or pyriform, glabrous, up to 140 µm diam., without distinct ostioles. Pycnidial wall pseudoparenchymatous, composed of oblong to isodiametric cells, 3–5 cell layers, 13–18 µm thick. Conidiogenous cells phialidic, smooth, hyaline, ampulliform to doliiform, 4.5–9.5 × 4–5 µm. Chlamydospores intercalary or terminal, pale brown, smooth, single or in chains, globose to oval. Conidia ellipsoidal to oblong, aseptate, hyaline, smooth, thin-walled, guttulate, 3–5.5 × 1.5–2.5 µm.

Culture characteristics: Colonies on PDA, 50–55 mm diam. after seven days of cultivation in the dark at 28°C, raised, margin regular, velvety, with abundant aerial mycelium, dense, white, pale yellow near the center; reverse: white to pale yellow, sienna pigment produced near the center.

Material examined: China, Guizhou Province, Guiyang City, from healthy stems of *R. roxburghii* (Rosaceae), 22 April 2020, H. Zhang (HGUP191049); living cultures were deposited in the Culture Collection at the Department of Plant Pathology, College of Agriculture, Guizhou University, China, No. GUCC 191049.1 and China General Microbiological Culture Collection Center, No. CGMCC 40110.

Notes: The screened strain HGUP191049 and the type of *Ep. latusicollum* are phylogenetically similar as they cluster together with well support (Supplementary Figure 1). Our collection resembles the type CGMCC 3.18346 in having a pycnidial wall, conidiogenous cells, and conidia. However, our collection slightly differs from the type in having aggregated conidiomata rather than solitary conidiomata (Chen et al., 2017). Therefore, the examined morphology overlaps and is phylogenetically identical to *Ep. latusicollum*. We report our collection as a new host record of *Ep. latusicollum* from the stem of *R. roxburghii*.

## 3.5 Genome sequencing and annotation

Genome sequencing of *Ep. latusicollum* HGUP191049 was conducted using a combination of single molecule real-time (SMRT) and Illumina sequencing technologies. The obtained genome of HGUP191049 was assembled into 22 scaffolds, about

10.3389/fpls.2022.1060478

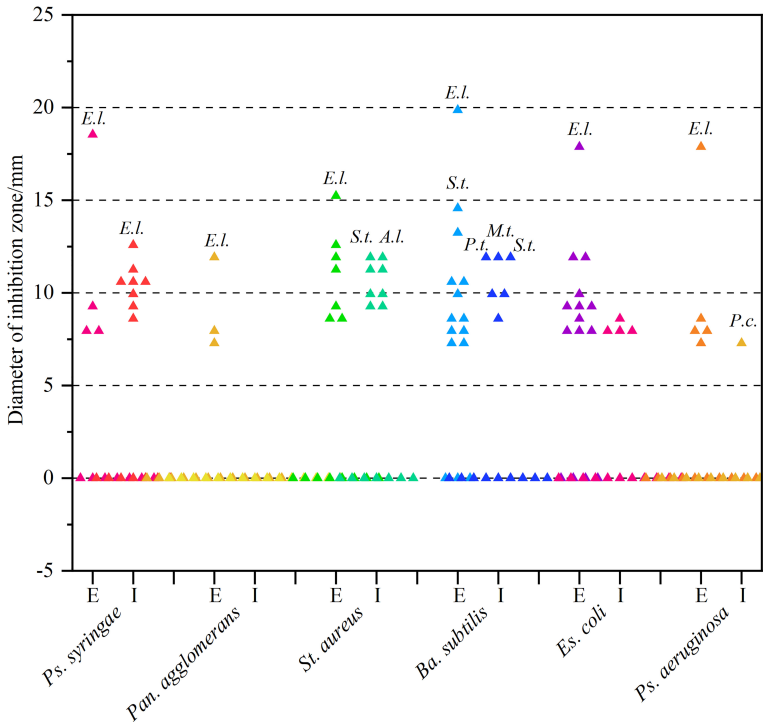



FIGURE 3
Results of re-screening for antibacterial activity. E and I indicate extracellular and intracellular metabolites, respectively. *A.l.*, *E.l.*, *M.t.*, *P.c.*, *P.t.*, and *S.t.* stand for *Alternaria longipes* HGUP192022, *Epicoccum latusicollum* HGUP191049, *Mycoleptodiscus terrestris* HGUP190018, *Penicillium crustosum* HGUP190031, *Phyllosticta telopeae* HGUP192003, and *Setophoma terrestris* HGUP190028, respectively.

33.24 megabase pairs (Mbp), and 10,500 genes (Figure 5; Table 2). The estimated genome size of HGUP191049 is broadly congruent with other estimates of genome size in *Epicoccum*, 33–35 Mbp (Supplementary Figure 2) (Fokin et al., 2017; Oliveira et al., 2017; Guo et al., 2021). The N50 and N90 length of the scaffolds were 1,859,063 bp and 1,112,482 bp, respectively. The GC content was 52.06% for the genome and 54.82%% for the coding sequences. In total, 10,310 protein-encoding genes were predicted from the genome assembly. Among them, 10,197, 9,523, 5,981, and 4,530 genes have functional annotations in the Nr, KEGG, SwissProt, and KOG databases, respectively. In this study, 325 genes (3.10%) were associated with secondary metabolite biosynthesis, transport, or catabolism in the KOG database (Supplementary Table 3).

The BGCs were analyzed using antiSMASH, and a total of 24 putative natural product BGCs of HGUP191049 were yielded, including three NRPSs, nine T1PKSs, four terpene synthases, one NRPS-T1PKS, one Indole-T1PKS, and six NRPS-like gene clusters (Supplementary Table 4). Of the 24 annotated BGCs, eight BGCs were found to share similarities in gene content with previously identified, while the remaining showed no significant similarities with currently known. These unknown BGCs could potentially shed light on the search for novel compounds. The antiSMASH and BLAST bioinformatics analyses identified three

TABLE 1 The minimum inhibitory concentration (MIC) of extracellular metabolites of two endophytic isolates against 12 tested strains.

| Species | Strain no. | MIC concentration (mg/mL) | | | | | |
|---|---|---|---|---|---|---|---|
| | | *L. theobromae* | *Botryo. dothidea* | *C. capsici* | *Py. oryzae* | *R. solani* | *F. oxysporum* |
| *Setophoma terrestris* | HGUP190028 | 5.00 | 5.00 | 10.00 | 20.00 | – | 10.00 |
| *Epicoccum latusicollum* | HGUP191049 | 1.25 | 2.50 | 10.00 | – | 1.25 | 10.00 |
| **Species** | **Strain no.** | ***Ps. syringae*** | ***Pan. agglomerans*** | ***St. aureus*** | ***Ba. subtilis*** | ***Es. coli*** | ***Ps. aeruginosa*** |
| *Setophoma terrestris* | HGUP190028 | – | – | 2.50 | 0.31 | 10.00 | – |
| *Epicoccum latusicollum* | HGUP191049 | 0.31 | 1.25 | 5.00 | 0.62 | 2.50 | 2.50 |

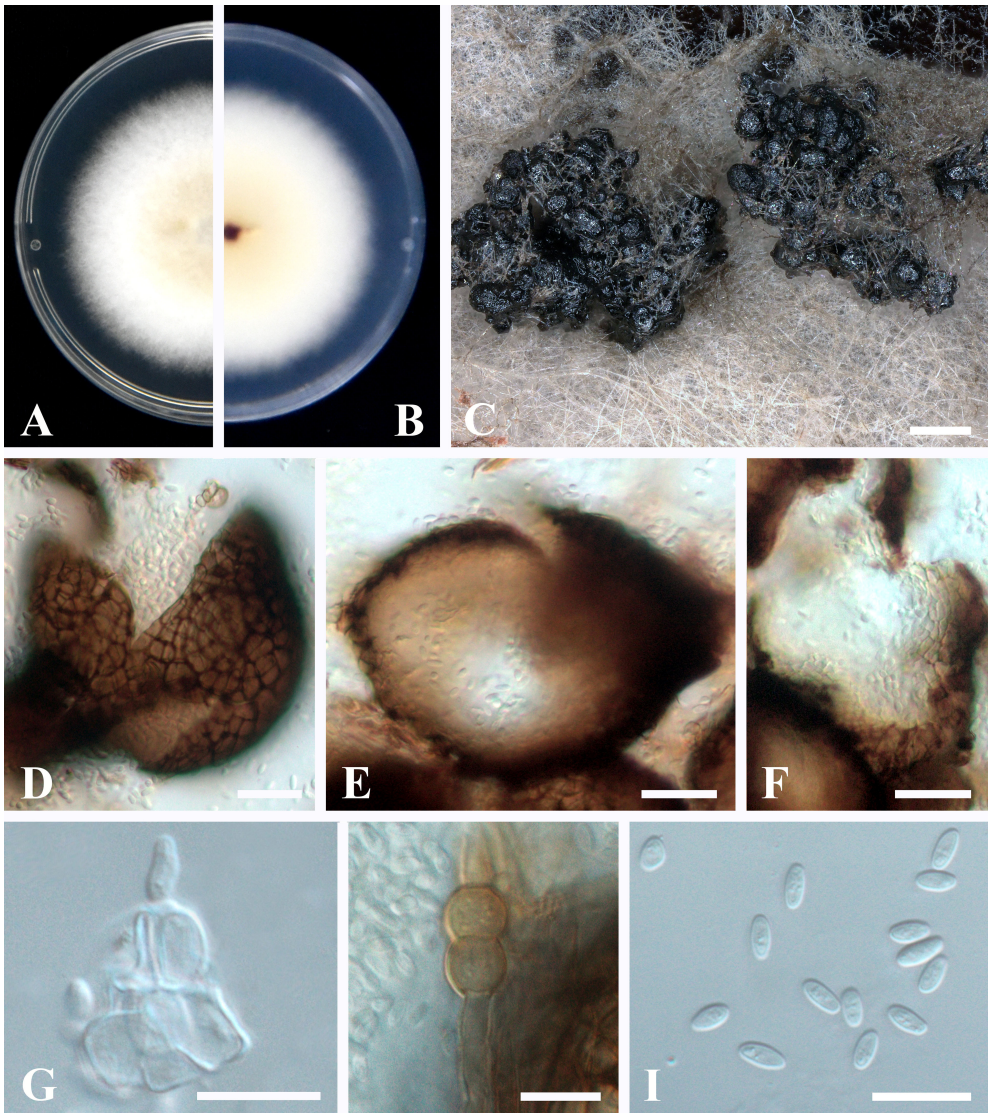



FIGURE 4
*Epicoccum latusicollum* (HGUP191049). **(A, B)**. Colony on PDA (front and reverse). **(C)** Pycnidia forming on PDA. **(D)** Pycnidia. **(E)** Section of pycnidium. **(F)** Section of pycnidial wall. **(G)** Conidiogenous cells. **(H)** Chlamydospores. **(I)** Conidia. Scale bars: **C** = 200 μm; **D**–**F** = 20 μm; **G**–**I** = 10 μm.

complete BGCs encoding dimethylcoprogen, (-)-mellein, and melanin. Other annotated potential products were squalestatin S1 (40% similarity), phomasetin (40%), oxyjavanicin (25%), patulin (20%), and azanigerone A (26%), respectively. Among these putative natural products, oxyjavanicin (Supplementary Figure 3), squalestatin S1 (Supplementary Figure 4), and patulin (Figure 6) have been reported to exhibit antimicrobial activity (Nicolaou et al., 1994; Paytubi et al., 2017; Kato et al., 2020). In this study, the putative patulin BGC is cluster 16 (T1PKS), sharing only 20% similarity to BGC0000120. We detected this compound in the secondary metabolites of *Ep. latusicollum* HGUP191049 (Figure 6).

## 3.6 Non-target metabolomics

According to the non-target metabolomics profiling, we detected 999 metabolites for positive mode, 523 for negative mode, and 18286 (92.3%) unknown metabolites (positive and negative modes), which indicated that *Ep. Latusicollum* HGUP191049 might produce a large number of new compounds. It was determined by conducting a literature search whether known metabolites had antimicrobial activity. The results revealed about 120 compounds with antimicrobial activity, 7.9% of the known compounds (Supplementary Table 5). Specifically, some antimicrobial compounds with different structures were

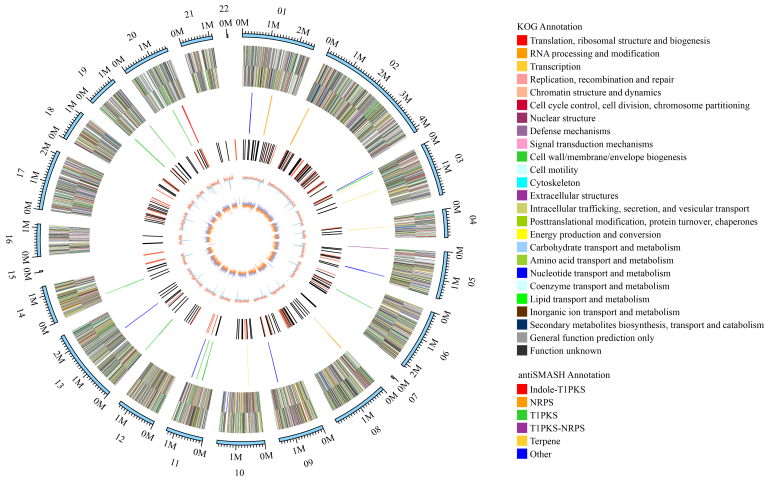



FIGURE 5
Circular map of genomic features of *Epicoccum latusicollum* HGUP191049. The peripheral circles represent the scaffolds (Mb scale), consisting of 22 scaffolds. From outer to inner circles (second to the sixth circle) are KOG annotation (forward and reverse strands), different colors indicate different functional classification; antiSMASH annotation (forward and reverse strands), different colors indicate different types of biosynthetic gene clusters (BGCs); ncRNA (black indicates tRNA, red indicates rRNA); GC content (red indicates greater than the mean, blue indicates less than the mean); GC skew (used to measure the relative content of G and C, GC skew = (G-C)/(G+C); purple indicates greater than 0, orange indicates less than 0).

illustrated in Figure 7, including polyketide (e.g., kendomycin), alkaloids (e.g., berberine), terpenoids (e.g., geniposidic acid), flavonoids (e.g., nevadensin), steroids (e.g., fluticasone propionate), naphthoquinone compounds (e.g., atovaquone), anthraquinones (e.g., hypericin), phenolic compounds (e.g., mangostine), coumarin compounds (e.g., 6-methylcoumarin), fatty acid compounds (e.g., phenyllactic acid), carbamates (e.g., geldanamycin), amides (e.g., benzamide), heterocyclic compounds (e.g., kojic acid), antibiotic compounds (e.g., norfloxacin), and other antimicrobial compounds (e.g., (+)-trans-chrysanthemic acid, (S)-(-)-citronellic acid, and azadirachtin A). Within this, flavonoids are one of the most abundant groups of antimicrobial secondary metabolites. The highly structural diversity demonstrated that *Ep. latusicollum* HGUP191049 is a talented producer of antimicrobial compounds.

## 3.7 Comparative genomics analyses results

### 3.7.1 Prediction and comparative analyses of pathogenicity-related genes

To identify and compare potential protein-coding genes related to pathogenicity and virulence in the genomes, whole genome blast analyses were performed against the pathogen-host interaction (PHI) gene database v. 4.13 at $E<1*10^{-20}$ and identity≥70% (Prasad et al., 2015; Urban et al., 2017; Xu et al., 2020). Screening of PHI annotated phenotypes showed that most genes belonged to "reduced virulence", "unaffected pathogenicity", and "loss of pathogenicity". In contrast, few genes were associated with the "effector (plant avirulence determinant)" (one gene), "enhanced antagonism" (one gene), and "chemistry target

TABLE 2 Genome features of *Epicoccum latusicollum* HGUP191049.

| Genome features | Value | Genome features | Value |
|---|---|---|---|
| Size of assembled genome (Mbp) | 33.24 | Protein-coding genes (≥ 60 aa) | 10,304 |
| GC content of assembled genome (%) | 52.06 | Min protein length (aa) | 52 |
| Number of scaffolds | 22 | Max protein length (aa) | 9,186 |
| N50 Length (bp) | 1,859,063 | tRNA genes | 206 |
| N90 Length (bp) | 1,112,482 | rRNA genes | 88 |
| Maximum length (bp) | 4,138,377 | Depth | 295X |
| Minimum length (bp) | 38,673 | Genes assigned to KOG categories | 4,530 |
| Average gene length (bp) | 1422.07 | Total length of contigs | 33242988 |
| All protein-coding genes | 10,310 | Putative biosynthetic gene clusters for secondary metabolites | 24 |

10.3389/fpls.2022.1060478

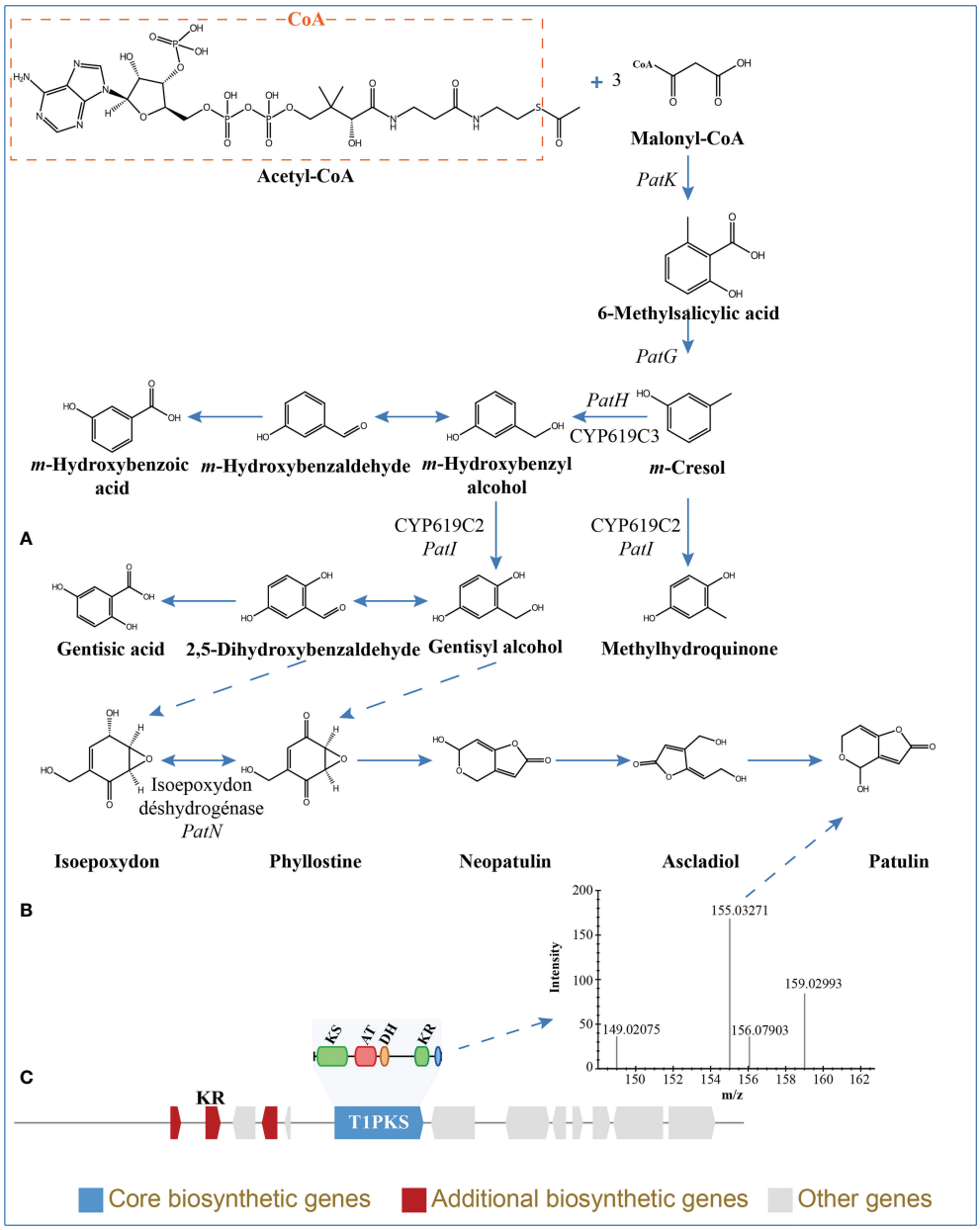



FIGURE 6
Patulin in *Epicoccum latusicollum* HGUP191049. **(A)** Biosynthetic pathways of patulin (Puel et al., 2010; Nielsen et al., 2017). **(B)** Non-targeted metabolic profiling spectrum. **(C)** Schematic representation of the putative BGC of patulin (cluster 16). KS, ketosynthase; AT, acyl transferase; DH, dehydratase; KR, ketoreductase; T1PKS, type I polyketide synthases.

sensitivity to chemical" (none) phenotypes (Figure 8). Genes of the "increased pathogenicity (Hypervirulence)" type are key pathogenic ones. As illustrated in Figure 8 and Supplementary Figure 2, strains with different nutrient modes of the same *Epicoccum* species may possess the same gene numbers of "increased pathogenicity (hypervirulence)" type, such as *Ep. latusicollum* (HGUP191049 and T41), *Ep. nigrum* (cf0051 and ICMP 19927), and *Ep. sorghinum* (BS2-1 and USPMTOX48), having 8, 7, 8 genes of this type for them, respectively. Of this phenotypic gene, the seven genomes in this study shared seven identical genes of this type, whereas *Ep. latusicollum* (HGUP191049 and T41) and *Ep. sorghinum* (BS2-1 and USPMTOX48) had one more of this type gene than the other genomes, namely PHI:5494, which may be a vital contributor to the difference in pathogenicity between species. In addition, an endophyte of the same species may contain more "loss of

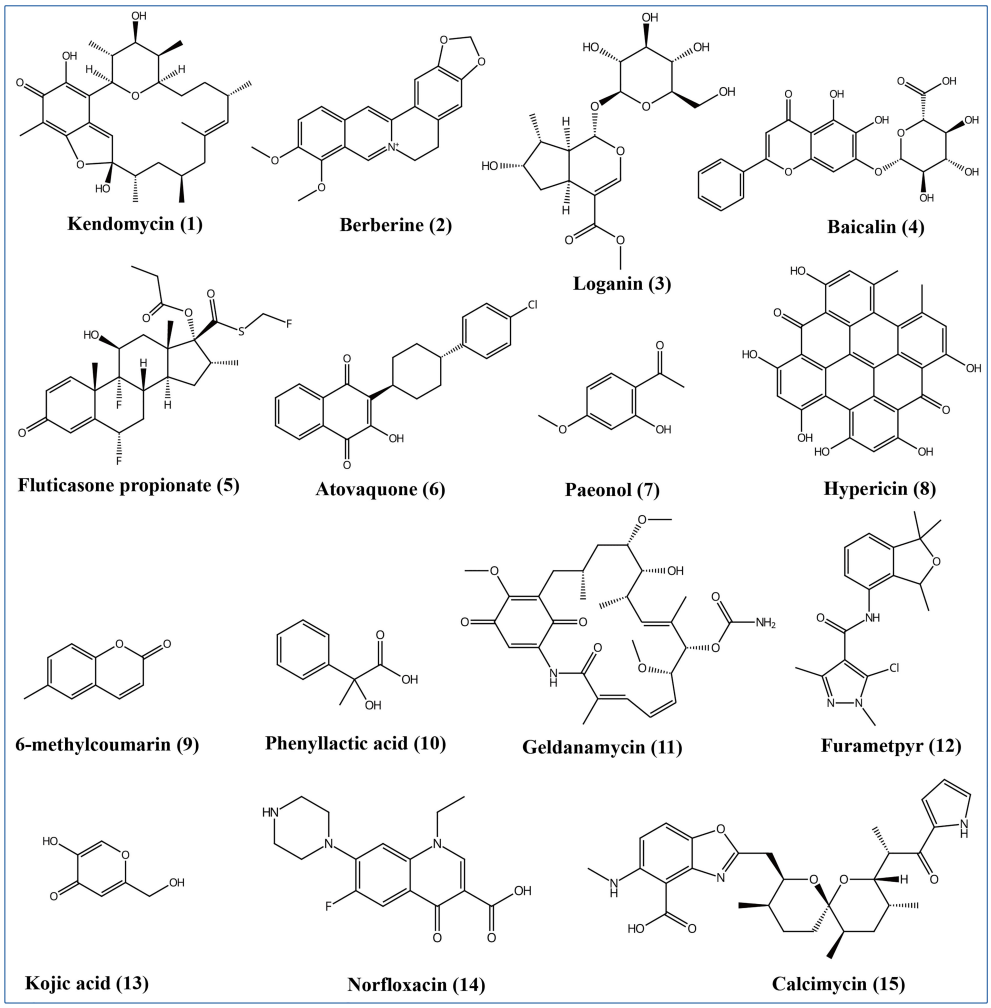



FIGURE 7
Some different structural types of antimicrobial secondary metabolites from *Epicoccum latusicollum* HGUP191049.

pathogenicity phenotypic genes" than a pathogen, e.g. endophytic strain HGUP191049 has two more genes of this type than pathogenic strain T41, and endophytic cf0051 has four more genes than pathogenic ICMP 19927. Moreover, for *Ep. latusicollum*, endophytic HGUP191049 had six "loss of pathogenicity" phenotypic genes (PHI: 2145, PHI: 4095, PHI: 10527, PHI: 9899, and PHI: 8875) different from the pathogenic T41 (PHI: 8734, PHI: 5232, and PHI: 9357), which may be one of the factors contributing to their differences in pathogenicity within species.

### 3.7.2 Prediction and comparative analyses of carbohydrate-active enzyme genes

Carbohydrate-active enzymes (CAZymes) are essential for fungal biological activity. CAZymes are responsible for degrading host plant cells and establishing colonization for plant pathogenic and endophytic fungi. As biocontrol fungi, CAZymes can be used to destroy the cell walls of pathogens and nematodes (Yang et al., 2019). The CAZymes involved in the degradation of plant cell walls were further classified into the degradation of cellulose, hemicellulose, and pectin, and those involved in the degradation of fungal cell walls were grouped into the degradation of chitin and β-1,3-glucan (Zhao et al., 2013; Kubicek et al., 2014; Yang et al., 2019).

As can be derived from Figure 9, the main CAZyme gene families that differ significantly between *Epicoccum* and *Didymella* are GH10, GH28, GH43, and PL1. In this study, 41.5% (17/41) of the families are identical among and within species in *Epicoccum*, such as GH6, all of which are 3 in number. Other families differ in the number of characteristics by 1–2, with a few 3, as in GH43. However, it is significantly different for GH18, which belongs to a family associated with chitin degradation, with numbers ranging from 9 to 15, which may be an important factor influencing the difference in the

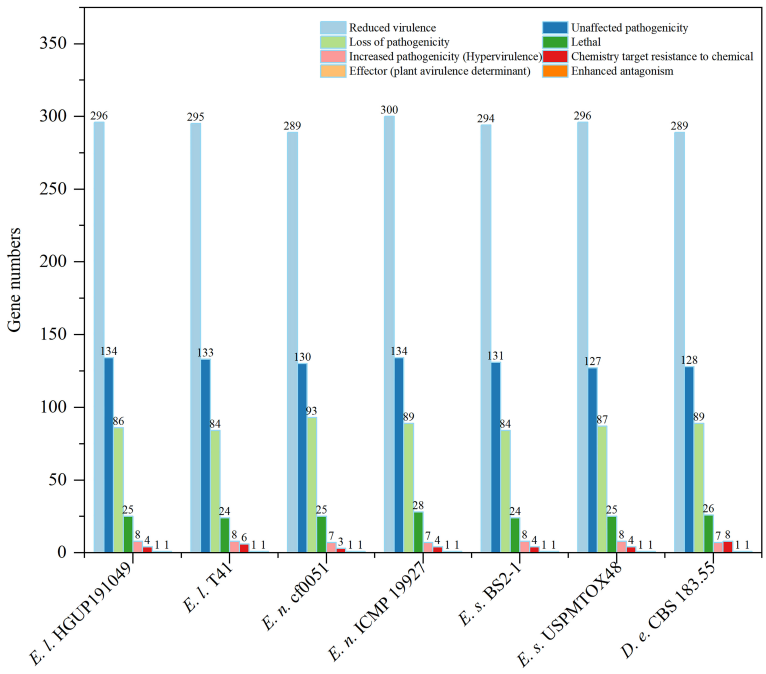



FIGURE 8
Comparisons and annotations of pathogen-host interactions (PHI). *D.e*, *Didymella exigua*; *E.l*, *Epicoccum latusicollum*; *E.n*, *Epicoccum nigrum*; *E.s*, *Epicoccum sorghinum*.

antifungal potential of *Epicoccum* spp. Of *Ep. latusicollum*, the biological activity of strain HGUP191049 distinguished from T41 in having different amounts of GH3 and GH45, GH43, GH78 and PL3, and GH18, for the degradation of cellulose, hemicellulose, pectin, and chitin, respectively.

### 3.7.3 Prediction and comparative analyses of BGCs

In this study, there were 177 BGCs from six *Epicoccum* genomes, of which PKS accounted for 35.0%, NRPS for 18.6%, terpene for 15.8%, hybrid PKS/NRPS for 6.8%, indole for 2.3%, and other unknown BGCs (NRPS-like) for 21.5% (Figure 10), which suggests that *Epicoccum* is a promising source of terpenes besides the traditional PKS- and NRPS-encoded compounds.

Phylogenetic relationships of BGCs from six *Epicoccum* strains and evolutionarily adjacent species *D. exigua* were analyzed to investigate differences among secondary metabolites of *Epicoccum* spp. (Figure 10). The result showed that BGCs could be grouped into 26 clades. The same types of BGCs with high identity may encode the same secondary metabolites, while the corresponding BGCs of a compound may be in different evolutionary branches. Notably, *Epicoccum* species have BGCs encoding the same compounds. The same branch of *Ep. latusicollum* (Contig0006.1 and JACCMO010000004.1) and *Ep. sorghinum* (VXJJ01000017.1 and MIEO01000350.1) (Clade 18), which all encode oxyjavanicin, where *Ep. nigrum*, the BGC JAASLF010000044.1, which encodes this compound, belongs to Clade 8. Similarly, squalestatin S1 is also encoded by BGCs from six different *Epicoccum* genomes. Consequently, it is presumed that oxyjavanicin and squalestatin S1, both of which have been reported as antimicrobial agents, are secondary metabolites shared by *Epicoccum* spp.

## 4 Discussion

Medicinal plants have long been used as a source of medicine. Approximately 8,000 medicinal plants have been developed into drugs and biocides, contributing more than 7,000 compounds to the pharmaceutical industry (Kaul et al., 2012). However, the overuse of medicinal plants in traditional folk medicine practices has led to environmental degradation and loss of biodiversity. Developing bioactive compounds based on endophytic fungi can reduce deforestation and the extinction of important and valuable medicinal plants (Uzma et al., 2019). Medicinal plants are a valuable source for exploring biologically active endophytes (Kaul et al., 2012). In this study, *R. roxburghii* is an economically important source of medicine and food. Its fruit is rich in vitamin C (up to 2 000 mg/100 g), superoxide dismutase (SOD), and flavonoids (Xu et al., 2019; Hou et al., 2020). The root, leaf, and fruit of *R. roxburghii* have been used as traditional medicinal materials to treat several diseases, such as dyspepsia, enteritis, and scurvy. In addition, some components

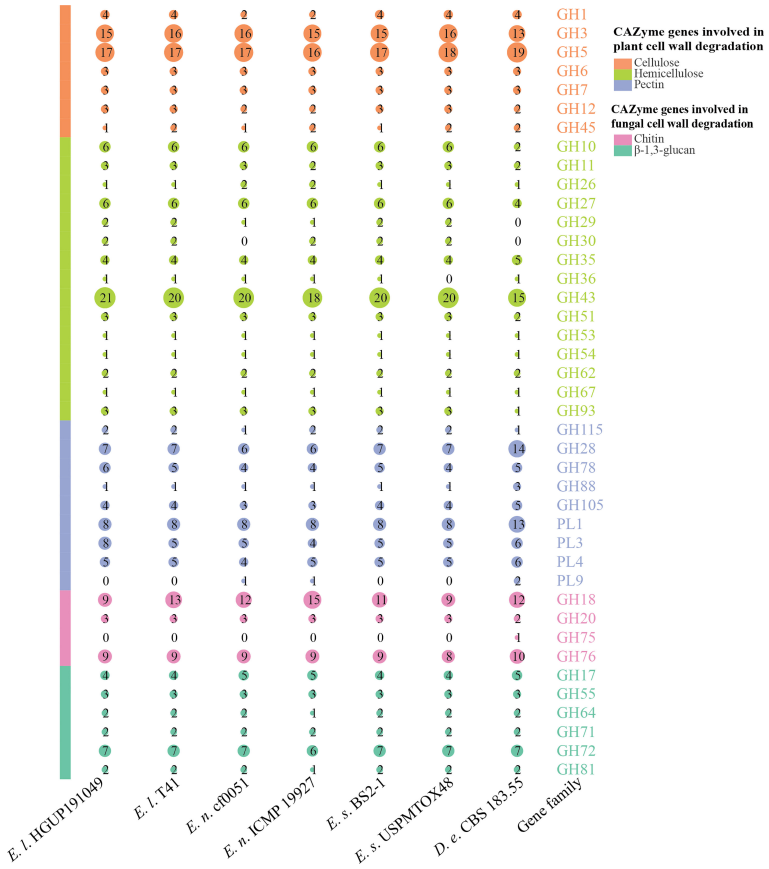



FIGURE 9
Comparisons and annotations of carbohydrate-active enzyme genes. *D.e, Didymella exigua*; *E.l, Epicoccum latusicollum*; *E.n, Epicoccum nigrum*; *E.s, Epicoccum sorghinum*. Different circle sizes indicate the number of different gene families.

extracted from *R. roxburghii* have been demonstrated to possess biological activities, including hypoglycemic, hypolipidemic, immune-enhancing, and antitumor effects (Zhang et al., 2021b). More importantly, *R. roxburghii* is also a plant source of antimicrobial compounds (Ma et al., 2020; Wang et al., 2021). Therefore, we selected *R. roxburghii* as a candidate for screening endophytic fungi with antimicrobial activity.

Species-level identification of fungi is a critical step to ensure reproducibility and is essential for both basic scientific research (ecology, taxonomy) and applied scientific research (genomics, bioprospecting). However, only 14% of fungal secondary metabolites studies have combined morphological and molecular data for identification (Raja et al., 2017). The results of these investigations suggest that the identification of fungi in most such studies is unreliable, as a single gene (mainly ITS) may fail to distinguish closely related members of certain genera phylogenetically. More than a quarter of GenBank fungal ITS sequences have not been adequately confirmed taxonomically (Zhang et al., 2021b). For accurate species identification, molecular data (preferably polygenic) should be combined with morphological studies (Woudenberg et al., 2017). In this study, we obtained the antimicrobial active strain *Ep. latusicollum* HGUP191049, whose taxonomic status was confirmed by morphology and multigene phylogenetic analyses.

In this study, the strains isolated from *R. roxburghii* with antimicrobial activity were screened out by multigene phylogenetic analyses (ITS, LSU, RPB2, TUB, TEF, and ACT), the plate confrontation method, and the disc diffusion method, namely *Ep. latusicollum* HGUP191049, *Neofusicoccum* sp. HGUP191080, and *Se. terrestris* HGUP190028. *Epicoccum latusicollum* has been reported to be capable of causing several plant diseases, including leaf spots on tobacco and *Elaeagnus pungens* (Guo et al., 2021; Qi et al., 2021), stalk rot on maize (*Zea mays* L.) (Xu et al., 2022), and root rot on *Nicotiana tabacum* (Gai et al., 2020). In this work, this species is first reported as an endophytic fungus with antimicrobial activity and is a new host record from *R. roxburghii*. Another strain with antimicrobial activity, *Neofusicoccum* sp. HGUP191080 may phylogenetically represent a new species and requires further identification by morphology. *Neofusicoccum* species, which are endophytes or

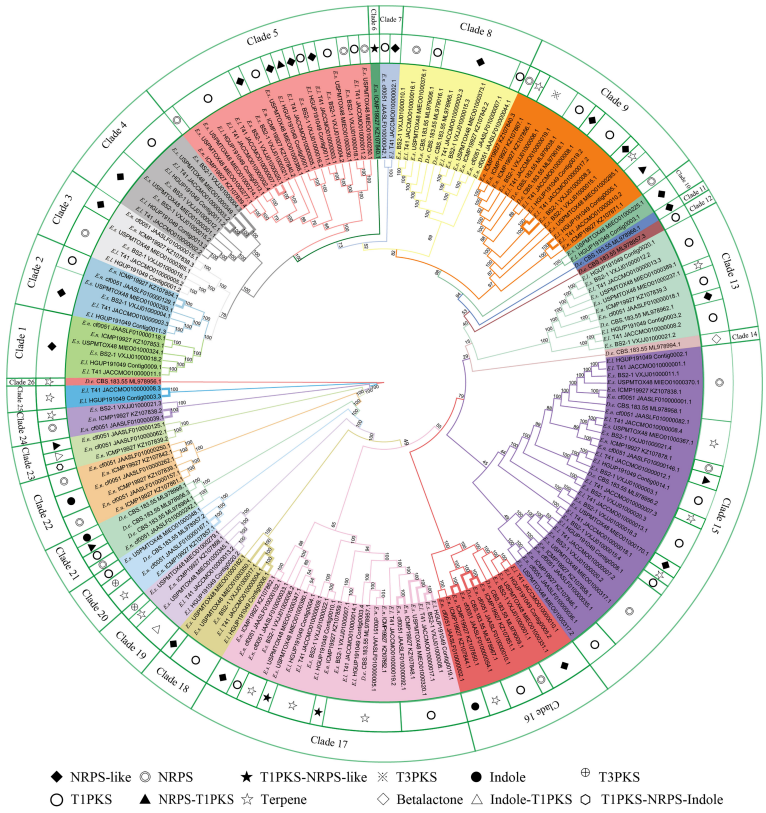



FIGURE 10
Phylogenetic analyses of biosynthetic gene clusters (BGCs). Bootstrap support values for maximum likelihood are given near nodes. *D.e, Didymella exigua*; *E.l, Epicoccum latusicollum*; *E.n, Epicoccum nigrum*; *E.s, Epicoccum sorghinum*. The species name is followed by the strain number, and the final number indicates the gene of each BGC. **Bolded** adjacent branches indicate coding for the same compound.

pathogens of plants, produce structurally different metabolites that show interesting biological activities such as antibacterial, cytotoxic, and phytotoxic (Salvatore et al., 2021). Finally, *Se. terrestris* caused pink root rot in various plants, such as squash, canola, and winter squash (Ikeda et al., 2012; Yang et al., 2017; Rivedal et al., 2018). However, as an endophyte isolated from *Dysoxylum binectariferum*, *Se. terrestris* is known to produce blennolides with anticancer and antimicrobial activity (Arora et al., 2018). Thus, some species commonly reported as pathogens may have potential biological activity as endophytic fungi. Further MIC assays demonstrated that the present study's antimicrobial strength and spectrum of *Ep. latusicollum* HGUP191049 were superior to other strains.

The development of genomics, transcriptomics, proteomics, metabolomics, high-throughput technologies, and computational resources has significantly broadened the understanding of the key pathways affecting the synthesis of fungal secondary metabolites (Palazzotto and Weber, 2018). In this study, genomics, non-target metabolomics, and comparative genomics were performed further to investigate the biosynthetic capacity of *Ep. latusicollum* HGUP191049. Genes required for secondary metabolite synthesis are typically arranged in a multigene biosynthetic gene cluster (Yang et al., 2019). With this high-quality genome sequence and annotation, we predicted a total of 24 BGCs, which may encode eight known compounds. Of these compounds, squalestatin S1, oxyjavanicin, and patulin were reported to have antimicrobial activity (Nicolaou et al., 1994; Paytubi et al., 2017; Kato et al., 2020). Genetic modification of BGCs and/or introduction of a particular mutation provides opportunities to obtain derivatives of the original metabolites (Ichikawa et al., 2012). Genome mining of gene clusters encoding biosynthetic pathways of fungal secondary metabolites has become a critical approach for new compound discovery (Weber et al., 2015). The sequencing and annotation of the *Ep. latusicollum* HGUP191049 genome is the foundation for the identification of antimicrobial compound BGCs, the activation of silencing gene clusters, and the identification and regulation of biosynthetic pathways. In this study, non-targeted metabolic analyses further revealed the biosynthetic capacity and potential antimicrobial compounds of *Ep. latusicollum* HGUP191049 by determining all detectable metabolites. Moreover, patulin, a compound encoded by gene cluster 16, has also been detected.

Comparative genomics aims to use an ensemble of related genomes to improve the understanding of each genome in the set (Haubold & Wiehe, 2004). *Epicoccum* is a genus in which endophytic, saprophytic, and pathogenic modes of nutrition coexist, such as *Ep. nigrum* is a primary saprophyte involved in the retting of flax (Brown, 1984), an endophytic fungus isolated from the leaves of *Lysidice rhodostegia* (Wang et al., 2010), even a pathogen that causes leaf spot disease on *Lablab purpureus* (Mahadevakumar et al., 2014). So, *Epicoccum* species may reshape their lifestyles among endophytic, saprophytic, and pathogenic to adapt to changing environmental conditions (Kuo et al., 2014). The fungus may secrete numerous proteins that facilitate colonization during interaction with the plant (Yin et al., 2015). Nine high-level phenotypic terms are defined in PHI-base to compare the pathogen-host interactions between organisms across the tree of life (Urban al., 2017). Using comparative genomic approaches, we show that PHI:5494, one of the "increased pathogenicity (hypervirulence)" type genes, may be an important factor in the difference in pathogenicity between *Epicoccum* species. Moreover, endophytic HGUP191049 had six "loss of pathogenicity" phenotypic genes different from the pathogenic T41, which may account for the lifestyle differences in *Ep. latusicollum*.

Currently, CAZyme gene families are defined and classified into six main categories in the CAZy database: glycosyltransferases (GTs), glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), carbohydrate-binding modules (CBMs), and enzymes of auxiliary activities (AAs) (Zhang et al., 2018). Of these families, GH18 is related to a family of chitin degradation in amounts ranging from 9 to 15 by comparative analyses, which may be an important factor contributing to the differences in the antifungal potential of *Epicoccum* spp.

The secondary metabolites of fungi constitute a rich source of natural products with antimicrobial activity. Genes encoding biosynthetic pathways of secondary metabolites are usually located on chromosomes forming BGCs (Yao et al., 2021). Results from comparative analyses show that *Epicoccum* is a promising source of terpenes. Terpenes exhibit antimicrobial activity owing to their highly lipophilic nature, which may interfere with the integrity and function of cell membranes (Sohrabi et al., 2015). In addition, oxyjavanicin and squalestatin S1 reported as antimicrobial agents (Nicolaou et al., 1994; Kato et al., 2020), are putative secondary metabolites shared by *Epicoccum* spp. In this study, a broad-spectrum antimicrobial potential strain was screened out from the endophytic fungi of *R. roxburghii* and analyzed for biosynthetic capacity.

## 5 Conclusion

We isolated 54 endophytic fungi from *R. roxburghii* and analyzed their multigene phylogenetic relationships. *In vitro* antimicrobial experiments revealed that the endophytic strain with broad-spectrum antimicrobial potential, *Ep. latusicollum* HGUP191049, was screened out. Multi-omics analyses suggested that *Epicoccum* spp. is an ideal source of antimicrobial compounds. In conclusion, plants with specific medicinal value are promising sources for isolating endophytes with corresponding particular functions.

## Data availability statement

The datasets presented in this study can be found in online repositories. The accession numbers of the sequences deposited in GenBank are: ITS: MZ541933–MZ541986; LSU: MZ540051–MZ540080; RPB2: MZ546146–MZ546149 and OP321271-OP321292; TUB: MZ546150–MZ546167 and OP312077–OP312084; TEF1: MZ546168–MZ546171; ACT: MZ546172. The Ep. latusicollum HGUP191049 whole genome sequence data have been submitted to the GenBank database under accession no. JANURY000000000.

## Author contributions

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## Funding

This work is supported by the following projects: the Guizhou Provincial Science and Technology Projects (No. [2021]221 and [2020]1Y043), the Guizhou Province Science and Technology Innovation Talent Team Project (Qian Ke He Pingtai Rencai–CXTD [2021]004), and the Natural Science Foundation of China (No. 32060009).

## Acknowledgments

The authors thank Prof. Yan-Feng Han and Dr. Li Luo for their crucial help in sampling. We also thank Prof. Zhong Li, Associate Prof. Xin Xie, Associate Prof. Zhi-Bo Zhao, Associate Prof. Hai-Xia Ding, Dr. Xian-Feng Hu, and Ying Shen for providing the tested strains.

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1060478/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Phylogenetic analysis of *Epicoccum latusicollum* HGUP191049, phylogenetic tree generated from a maximum likelihood analysis based on the combined LSU, ITS, RPB2, and TUB sequence data, the tree is rooted with *Didymella exigua* CBS 183.55 and *D. rumicicola* CBS 683.79, **T** type or ex-type.

SUPPLEMENTARY FIGURE 2
Maximum likelihood phylogram and genome statistics of *Epicoccum* species analysed in this study. The genome accession numbers of *Ep. latusicollum* (HGUP191049 and T41), *Ep. nigrum* (cf0051 and ICMP 19927), *Ep. sorghinum* (BS2-1 and USPMTOX48), and *Didymella exigua* (CBS 183.55) are JANURY000000000, JACCMO000000000, JAASLF000000000, NCTX00000000, VXJJ00000000, MIEO00000000, and VOSY00000000, respectively.

SUPPLEMENTARY FIGURE 3
Putative oxyjavanicin in *Epicoccum latusicollum* HGUP191049. a. Oxyjavanicin structure b. Schematic representation of the putative BGC of oxyjavanicin (cluster 11). **TD**: thioesterase domain, **KS**: ketosynthase, **AT**: acyl transferase, **PT**: product template, **T1PKS**: type I polyketide synthases.

SUPPLEMENTARY FIGURE 4
Putative squalestatin S1 in *Epicoccum latusicollum* HGUP191049. **(A)** Squalestatin S1 structure **(B)** Schematic representation of the putative BGC of squalestatin S1 (cluster 4).

## References

Agrawal, S., Samanta, S., and Deshmukh, S. K. (2022). The anti-diabetic potential of endophytic fungi: Future prospects as therapeutic agents. *Biotechnol. Appl. Bioc.* 69, 1159–1165. doi: 10.1002/bab.2192

Alhadrami, H. A., Sayed, A. M., El-Gendy, A. O., Shamikh, Y. I., Gaber, Y., Bakeer, W., et al. (2021). A metabolomic approach to target antimalarial metabolites in the *Artemisia annua* fungal endophytes. *Sci. Rep.* 11, 2770. doi: 10.1038/s41598-021-82201-8

Arora, D., Chashoo, G., Singamaneni, V., Sharma, N., Gupta, P., and Jaglan, S. (2018). *Bacillus amyloliquefaciens* induces production of a novel blennolide K in coculture of *Setophoma terrestris*. *J. Appl. Microbiol.* 124, 730–739. doi: 10.1111/jam.13683

Brown, A. E. (1984). *Epicoccum nigrum*, a primary saprophyte involved in the retting of flax. *Trans. Br. Mycol. Soc.* 83, 29–35. doi: 10.1016/S0007-1536(84)80242-9

Chen, N. (2004). Using repeat masker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* 5, 4–10. doi: 10.1002/0471250953.bi0410s05

Cheng, T., Kolařík, M., Quijada, L., and Stadler, M. (2022). A re-assessment of *Taxomyces andreanae*, the alleged taxol-producing fungus, using comparative genomics. *IMA Fungus* 13, 17. doi: 10.1186/s43008-022-00103-4

Chen, Q., Hou, L. W., Duan, W. J., Crous, P. W., and Cai, L. (2017). Didymellaceae revisited. *Stud. Mycol.* 87, 77–103. doi: 10.1016/j.simyco.2017.06.002

Deshmukh, S. K., Verekar, S. A., and Bhave, S. V. (2015). Endophytic fungi: a reservoir of antibacterials. *Front. Microbiol.* 5, 715. doi: 10.3389/fmicb.2014.00715

Fernando, K., Reddy, P., Hettiarachchige, I. K., Spangenberg, G. C., Rochfort, S. J., and Guthridge, K. M. (2020). Novel antifungal activity of *Lolium*-associated *Epichloë* endophytes. *Microorganisms* 8, 955. doi: 10.3390/microorganisms8060955

Fokin, M., Fleetwood, D., Weir, B. S., and Villas-Boas, S. (2017). Genome sequence of the saprophytic ascomycete *Epicoccum nigrum* strain ICMP 19927, isolated from new Zealand. *Genome Announ.* 5, e00557–e00517. doi: 10.1128/genomeA.00557-17

Freeman, E. M. (1904). The seed-fungus of *Lolium temulentum*, philosophical transactions of the royal society of London. *Ser. B Containing Papers Biol. Character* 196, 1–27. doi: 10.1098/rstb.1904.0001

Gai, X., Jiang, N., Lu, C., Xia, Z., and Qin, X. (2020). First report of *Epicoccum latusicollum* causing root rot on. *Nicotiana tabacum China. J. Plant Pathol.* 102, 1291. doi: 10.1007/s42161-020-00548-9

Gao, L., Ma, J., Liu, Y., Huang, Y., Mohamad, O. A. A., Jiang, H., et al. (2021). Diversity and biocontrol potential of cultivable endophytic bacteria associated with halophytes from the West aral Sea basin. *Microorganisms* 9, 1448. doi: 10.3390/microorganisms9071448

Gashgari, R., Gherbawy, Y., Ameen, F., and Alsharari, S. (2016). Molecular characterization and analysis of antimicrobial activity of endophytic fungi from medicinal plants in Saudi Arabia. *Jundishapur J. Microbiol.* 9, e26157. doi: 10.5812/jjm.26157

Glass, N. L., and Donaldson, G. C. (1995). Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Appl. Environ. Microb.* 61, 1323–1330. doi: 10.1128/aem.61.4.1323-1330.1995

Guo, Z., Yu, Z., Wang, H., Xie, H., and Liu, T. (2021). Leaf spot caused by *Epicoccum latusicollum* on tobacco in China. *Plant Dis.* 105, 501. doi: 10.1094/PDIS-07-20-1443-PDN

Gupta, S., Chaturvedi, P., Kulkarni, M. G., and Van Staden, J. (2020). A critical review on exploiting the pharmaceutical potential of plant endophytic fungi. *Biotechnol. Adv.* 39, 107462. doi: 10.1016/j.biotechadv.2019.107462

Hajieghrari, B., Torabi-Giglou, M., Mohammadi, M. R., and Davari, M. (2008). Biological potantial of some Iranian *Trichoderma* isolates in the control of soil borne plant pathogenic fungi. *Afr. J. Biotechnol.* 7, 967–972.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids symposium Ser.* 41, 95–98.

Haubold, B., and Wiehe, T. (2004). Comparative genomics: methods and applications. *Naturwissenschaften* 9, 405–421. doi: 10.1007/s00114-004-0542-8

Hou, Z., Yang, H., Zhao, Y., Xu, L., Zhao, L., Wang, Y., et al. (2020). Chemical characterization and comparison of two chestnut rose cultivars from different regions. *Food Chem.* 323, 126806. doi: 10.1016/j.foodchem.2020.126806

Huang, Y., Shen, L., Jiang, J., Xu, Q., Luo, Z., Luo, Q., et al. (2019). Metabolomic profiles of bovine mammary epithelial cells stimulated by lipopolysaccharide. *Sci. Rep.* 9, 19131. doi: 10.1038/s41598-019-55556-2

Hu, C., Zhou, S. W., Chen, F., Zheng, X. H., Shen, H. F., Lin, B. R., et al. (2017). Two unusual benzamido nine-membered dilactones from marine-derived *Streptomyces antibioticus* H12-15. *Molecules* 22, 557. doi: 10.3390/molecules22040557

Ichikawa, N., Sasagawa, M., Yamamoto, M., Komaki, H., Yoshida, Y., Yamazaki, S., et al. (2012). DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* 41, D408–D414. doi: 10.1093/nar/gks1177

Ikeda, K., Kuwabara, K., Urushibara, T., Soyai, P., Miki, S., and Shibata, S. (2012). Pink root rot of squash caused by *Setophoma terrestris* in Japan. *J. Gen. Plant Pathol.* 78, 372–375. doi: 10.1007/s10327-012-0398-2

Jia, M., Chen, L., Xin, H. L., Zheng, C. J., Rahman, K., Han, T., et al. (2016). A friendly relationship between endophytic fungi and medicinal plants: a systematic review. *Front. Microbiol.* 7, 906. doi: 10.3389/fmicb.2016.00906

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

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

Kato, S., Motoyama, T., Uramoto, M., Nogawa, T., Kamakura, T., and Osada, H. (2020). Induction of secondary metabolite production by hygromycin b and identification of the 1233A biosynthetic gene cluster with a self-resistance gene. *J. Antibiotics* 73, 475–479. doi: 10.1038/s41429-020-0295-4

Kaul, S., Gupta, S., Ahmed, M., and Dhar, M. K. (2012). Endophytic fungi from medicinal plants: a treasure hunt for bioactive metabolites. *Phytochem. Rev.* 11, 487–505. doi: 10.1007/s11101-012-9260-6

Kubicek, C. P., Starr, T. L., and Glass, N. L. (2014). Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annu. Rev. Phytopathol.* 52, 427–451. doi: 10.1146/annurev-phyto-102313-045831

Kumar, A., Patil, D., Rajamohanan, P. R., and Ahmad, A. (2013). Isolation, purification and characterization of vinblastine and vincristine from endophytic fungus. *Fusarium oxysporum isolated Catharanthus roseus. PloS One* 8, e71805. doi: 10.1371/journal.pone

Kuo, H. C., Hui, S., Choi, J., Asiegbu, F. O., Valkonen, J., and Lee, Y. H. (2014). Secret lifestyles of *Neurospora crassa*. *Sci. Rep.* 4, 5135. doi: 10.1038/srep05135

Kusari, S., Hertweck, C., and Spiteller, M. (2012). Chemical ecology of endophytic fungi: origins of secondary metabolites. *Chem. Biol.* 19, 792–798. doi: 10.1016/j.chembiol.2012.06.004

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Liu, Y. J., Whelen, S., and Hall, B. D. (1999). Phylogenetic relationships among ascomycetes: evidence from an RNA polymerse II subunit. *Mol. Biol. Evol.* 16, 1799–1808. doi: 10.1093/oxfordjournals.molbev.a026092

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

Madsen, R., Lundstedt, T., and Trygg, J. (2010). Chemometrics in metabolomics a review in human disease diagnosis. *Anal. Chim. Acta* 659, 23–33. doi: 10.1016/j.aca.2009.11.042

Mahadevakumar, S., Jayaramaiah, K. M., and Janardhana, G. R. (2014). First report of leaf spot disease caused by *Epicoccum nigrum* on *Lablab purpureus* in India. *Plant Dis.* 98, 284. doi: 10.1094/PDIS-07-13-0798-PDN

Manganyi, M. C., and Ateba, C. N. (2020). Untapped potentials of endophytic fungi: A review of novel bioactive compounds with biological applications. *Microorganisms* 8, 1934. doi: 10.3390/microorganisms8121934

Ma, Y., Wang, Y., Zhang, H., Sun, W., Li, Z., Zhang, F., et al. (2020). Antimicrobial mechanism of strictinin isomers extracted from the root of *Rosa roxburghii* tratt (Ci Li gen). *J. Ethnopharmacol.* 250, 112498. doi: 10.1016/j.jep.2019.112498

Minh, B. Q., Nguyen, M. A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., et al. (2000). A whole-genome assembly of drosophila. *Science* 287, 2196–2204. doi: 10.1126/science.287.5461.2196

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nicolaou, K. C., Yue, E. W., Naniwa, Y., De Riccardis, F., Nadin, A., Leresche, J. E., et al. (1994). Zaragozic acid a/squalestatin S1: Synthetic and retrosynthetic studies. *Angewandte Chemie Int. Edition English* 33, 2184–2187.

Nielsen, J. C., Grijseels, S., Prigent, S., Ji, B., Dainat, J., Nielsen, K. F., et al. (2017). Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species. *Nat. Microbiol.* 2, 17044. doi: 10.1038/nmicrobiol.2017.44

Oliveira, R. C., Davenport, K. W., Hovde, B., Silva, D., Chain, P. S., Correa, B., et al. (2017). Draft genome sequence of sorghum grain mold fungus *Epicoccum sorghinum*, a producer of tenuazonic acid. *Genome announc.* 5, e01495–e01416. doi: 10.1128/genomeA.01495-16

Palazzotto, E., and Weber, T. (2018). Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Curr. Opin. Microbiol.* 45, 109–116. doi: 10.1016/j.mib.2018.03.004

Pal, P. P., Shaik, A. B., and Begum, A. S. (2020). Prospective leads from endophytic fungi for anti-inflammatory drug discovery. *Planta Med.* 86, 941–959. doi: 10.1055/a-1140-8388

Paytubi, S., de la Cruz, M., Tormo, J. R., Martín, J., González, I., Gonzalez-Menendez, V., et al. (2017). A high-throughput screening platform of microbial natural products for the discovery of molecules with antibiofilm properties against *Salmonella*. *Front. Microbiol.* 8, 326. doi: 10.3389/fmicb.2017.00326

Prasad, P., Varshney, D., and Adholeya, A. (2015). Whole genome annotation and comparative genomic analyses of bio-control fungus *Purpureocillium lilacinum*. *BMC Genomics* 16, 1004. doi: 10.1186/s12864-015-2229-2

Puel, O., Galtier, P., and Oswald, I. P. (2010). Biosynthesis and toxicological effects of patulin. *Toxins* 2, 613–631. doi: 10.3390/toxins2040613

Qi, X. L., He, J., Li, D. W., and Huang, L. (2021). First report of leaf spot on *Elaeagnus pungens* caused by *Epicoccum latusicollum* in China. *For. Pathol.* 51, e12716. doi: 10.1111/efp.12716

Rahaman, M. S., Siraj, M. A., Sultana, S., Seidel, V., and Islam, M. A. (2020). Molecular phylogenetics and biological potential of fungal endophytes from plants of the sundarbans mangrove. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.570855

Raja, H. A., Miller, A. N., Pearce, C. J., and Oberlies, N. H. (2017). Fungal identification using molecular tools: A primer for the natural products research community. *J. Nat. Prod.* 80, 756–770. doi: 10.1021/acs.jnatprod.6b01085

Rambaut, A. (2014) *FigTree v1.4: Tree figure drawing tool*. Available at: http://treebio.ed.ac.uk/software/figtree/.

Rivedal, H. M., Stone, A. G., and Johnson, K. B. (2018). First report of *Setophoma terrestris* causing pink root rot of winter squash (*Cucurbita maxima*) in Oregon. *Plant Dis.* 102, 2661. doi: 10.1094/PDIS-06-18-0921-PDN

Rjeibi, I., Zaabi, R., and Jouida, W. (2020). Characterization of polysaccharides extracted from pulps and seeds of *Crataegus azarolus* l. var. *aronia*: preliminary structure, antioxidant, antibacterial, α-amylase, and acetylcholinesterase inhibition properties. *Oxid. Med. Cell. Longev.* 2020, 1903056. doi: 10.1155/2020/1903056

Rojas, E. C., Jensen, B., Jørgensen, H. J., Latz, M. A., Esteban, P., Ding, Y., et al. (2020). Selection of fungal endophytes with biocontrol potential against *Fusarium* head blight in wheat. *Biol. Control* 144, 104222. doi: 10.1016/j.biocontrol.2020.104222

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

Sagita, R., Quax, W. J., and Haslinger, K. (2021). Current state and future directions of genetics and genomics of endophytic fungi for bioprospecting efforts. *Front. Bioeng. Biotech.* 9, 649906. doi: 10.3389/fbioe.2021.649906

Salvatore, M. M., Alves, A., and Andolfi, A. (2021). Secondary metabolites produced by *Neofusicoccum* species associated with plants: a review. *Agriculture* 11, 149. doi: 10.3390/agriculture11020149

Singh, A. K., Sharma, R. K., Sharma, V., Singh, T., Kumar, R., and Kumari, D. (2017). Isolation, morphological identification and *in vitro* antibacterial activity of endophytic bacteria isolated from *Azadirachta indica* (neem) leaves. *Vet. World* 10, 510–516. doi: 10.14202/vetworld.2017.510-516

Sohrabi, R., Huh, J. H., Badieyan, S., Rakotondraibe, L. H., Kliebenstein, D. J., Sobrado, P., et al. (2015). In planta variation of volatile biosynthesis: an alternative biosynthetic route to the formation of the pathogen-induced volatile homoterpene DMNT *via* triterpene degradation in *Arabidopsis* roots. *Plant Cell* 27, 874–890. doi: 10.1105/tpc.114.132209

Stierle, A., Strobel, G. , and Stierle, D. (1993). Taxol and taxane production by *Taxomyces andreanae*, an endophytic fungus of pacific yew. *Science* 260, 214–216. doi: 10.1126/science8097061

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O., and Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* 18, 1979–1990. doi: 10.1101/gr.081612.108

Urban, M., Cuzick, A., Rutherford, K., Irvine, A., Pedro, H., Pant, R., et al. (2017). PHI-base: A new interface and further additions for the multi-species pathogen–host interactions database. *Nucleic Acids Res.* 45, D604–D610. doi: 10.1093/nar/gkw1089

Uzma, F., Mohan, C. D., Siddaiah, C. N., and Chowdappa, S. (2019). "Endophytic fungi: Promising source of novel bioactive compounds," in *Advances in endophytic fungal research* (Cham: Springer), 243–265.

Vilgalys, R., and Hester, M. (1990). Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *J. Bacteriol.* 172, 4238–4246. doi: 10.1128/jb.172.8.4238-4246.1990

Walker, M. C., Eslami, S. M., Hetrick, K. J., Ackenhusen, S. E., Mitchell, D. A., and van der Donk, W. A. (2020). Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. *BMC Genomics* 21, 387. doi: 10.1186/s12864-020-06785-7

Wang, J. M., Ding, G. Z., Fang, L., Dai, J. G., Yu, S. S., Wang, Y. H., et al. (2010). Thiodiketopiperazines produced by the endophytic fungus *Epicoccum nigrum*. *J. Nat. Prod.* 73, 1240–1249. doi: 10.1021/np1000895

Wang, R., He, R., Li, Z., Lin, X., and Wang, L. (2021). HPLC-Q-Orbitrap-MS/MS phenolic profiles and biological activities of extracts from roxburgh rose (*Rosa roxburghii* tratt.) leaves. *Arab. J. Chem.* 14, 103257. doi: 10.1016/j.arabjc.2021.103257

Wang, L., Ren, L., Li, C., Gao, C., Liu, X., Wang, M., et al. (2019). Effects of endophytic fungi diversity in different coniferous species on the colonization of *Sirex noctilio* (Hymenoptera: Siricidae). *Sci. Rep.* 9, 5077. doi: 10.1038/s41598-019-41419-3

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH 3.0–a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 43, W237–W243. doi: 10.1093/nar/gkv437

White, T. J., Bruns., T., and Lee, S. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics," in *MA Innis, DH gelfand, JJ sninsky, TJ white, eds. PCR protocols: a guide to methods and applications*, vol. Vol. 18. (New York: Academic Press Inc), 315–322.

Woudenberg, J. H. C., Hanse, B., Van Leeuwen, G. C. M., Groenewald, J. Z., and Crous, P. W. (2017). Stemphylium revisited. *Stud. Mycol.* 87, 43–76. doi: 10.1016/j.simyco.2017.06.001

Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., et al. (2017). MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads. *Nat. Methods* 14, 1072–1074. doi: 10.1038/nmeth.4432

Xu, R., Liu, X., Peng, B., Liu, P., Li, Z., Dai, Y., et al. (2020). Genomic features of *Cladobotryum dendroides*, which causes cobweb disease in edible mushrooms, and identification of genes related to pathogenicity and mycoparasitism. *Pathogens* 9, 232. doi: 10.3390/pathogens9030232

Xu, J., Vidyarthi, S. K., Bai, W., and Pan, Z. (2019). Nutritional constituents, health benefits and processing of *Rosa roxburghii*: a review. *J. Funct. Foods* 60, 103456. doi: 10.1016/j.jff.2019.103456

Xu, X., Zhang, L., Yang, X., Li, J., Wang, X. , Zhao, J., et al. (2022). First report of maize stalk rot caused by *Epicoccum latusicollum* on maize (*Zea mays* l.) in China. *Plant Dis.* 106, 2255. doi: 10.1094/PDIS-11-21-2392-PDN

Yang, Y., Liu, X., Cai, J., Chen, Y., Li, B., Guo, Z., et al. (2019). Genomic characteristics and comparative genomics analysis of the endophytic fungus *Sarocladium brachiariae*. *BMC Genomics* 20, 782. doi: 10.1186/s12864-019-6095-1

Yang, Y., Zuzak, K., Harding, M., Neilson, E., Feindel, D., and Feng, J. (2017). First report of pink root rot caused by *Setophoma* (Pyrenochaeta) *terrestris* on canola. *Can. J. Plant Pathol.* 39, 354–360. doi: 10.1080/07060661.2017.1355849

Yao, G., Chen, X., Zheng, H., Liao, D., Yu, Z., Wang, Z., et al. (2021). Genomic and chemical investigation of bioactive secondary metabolites From a marine-derived fungus Penicillium steckii. *Front. Microbiol.* 12, 600991. doi: 10.3389/fmicb.2021.600991

Ye, Z., Pan, Y., Zhang, Y., Cui, H., Jin, G., McHardy, A. C., et al. (2017). Comparative whole-genome analysis reveals artificial selection effects on *Ustilago esculenta* genome. *DNA Res.* 24, 635–648. doi: 10.1093/dnares/dsx031

Yin, Z. Y., Liu, H. Q., Li, Z. P., Ke, X. W., Dou, D. L., Gao, X. N., et al. (2015). Genome sequence of *Valsa* canker pathogens uncovers a potential adaptation of colonization of woody bark. *New Phytol.* 208, 1202–1216. doi: 10.1111/nph.13544

Yuan, X. L., Cao, M., Liu, X. M., Du, Y. M., Shen, G. M., Zhang, Z. F., et al. (2018). Composition and genetic diversity of the *Nicotiana tabacum* microbiome in different topographic areas and growth periods. *Int. J. Mol. Sci.* 19, 3421. doi: 10.3390/ijms19113421

Zhang, Z. B., Du, S. Y., Ji, B., Ji, C. J., Xiao, Y. W., Yan, R. M., et al. (2021a). New helvolic acid derivatives with antibacterial activities from *Sarocladium oryzae* DX-THL3, an endophytic fungus from dongxiang wild rice (*Oryza rufipogon* griff.). *Molecules* 26, 1828. doi: 10.3390/molecules26071828

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., et al. (2020). PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20, 348–355. doi: 10.1111/1755-0998.13096

Zhang, H., Wei, T. P., Li, L. Z., Luo, M. Y., Jia, W. Y., Zeng, Y., et al. (2021b). Multigene phylogeny, diversity and antimicrobial potential of endophytic sordariomycetes from *Rosa roxburghii*. *Front. Microbiol.* 12, 755919. doi: 10.3389/fmicb.2021.755919

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418

Zhang, P., Yuan, X. L., Du, Y. M., Zhang, H. B., Shen, G. M., Zhang, Z. F., et al. (2019). Angularly prenylated indole alkaloids with antimicrobial and insecticidal activities from an endophytic fungus *Fusarium sambucinum* TE-6L. *J. Agric. Food Chem.* 67, 11994–12001. doi: 10.1021/acs.jafc.9b05827

Zhao, Z., Liu, H., Wang, C., and Xu, J. R. (2013). Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics* 15, 6. doi: 10.1186/1471-2164-15-6

Zhao, Y., Zhang, C., Folly, Y. M. E., Chang, J., Wang, Y., Zhou, L., et al. (2019). Morphological and transcriptomic analysis of the inhibitory effects of *Lactobacillus plantarum* on *Aspergillus flavus* growth and aflatoxin production. *Toxins* 11, 636. doi: 10.3390/toxins11110636

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fpls.2023.1093358



OPEN ACCESS

EDITED BY
Milen I. Georgiev,
Bulgarian Academy of Sciences, Bulgaria

REVIEWED BY
Changle Ma,
Shandong Normal University, China
Andrey Stoyanov Marchev,
Bulgarian Academy of Sciences, Bulgaria

*CORRESPONDENCE
Basil J. Nikolau
dimmas@iastate.edu

†PRESENT ADDRESSES
Geng Ding,
Department of Genetics Development and Cell Biology, Iowa State University, Ames, IA, United States
Yosia Mugume,
US Department of Energy Plant Research Lab, Michigan State University, East Lansing, MI, United States
Maria Emilia Dueñas,
Laboratory for Biomedical Mass Spectrometry, Newcastle University, Newcastle, United Kingdom
Meiling Liu,
Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, United States

‡These authors have contributed equally to this work

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

RECEIVED 08 November 2022
ACCEPTED 30 January 2023
PUBLISHED 16 February 2023

CITATION
Ding G, Mugume Y, Dueñas ME, Lee YJ, Liu M, Nettleton DS, Zhao X, Li L, Bassham DC and Nikolau BJ (2023) Biological insights from multi-omics analysis strategies: Complex pleotropic effects associated with autophagy. *Front. Plant Sci.* 14:1093358. doi: 10.3389/fpls.2023.1093358



# Biological insights from multi-omics analysis strategies: Complex pleotropic effects associated with autophagy

Geng Ding[1†‡], Yosia Mugume[2†‡], Maria Emilia Dueñas[3†], Young Jin Lee[3], Meiling Liu[4†], Daniel S. Nettleton[4], Xuefeng Zhao[5], Ling Li[6], Diane C. Bassham[2] and Basil J. Nikolau[1*]

[1]Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, United States, [2]Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA, United States, [3]Department of Chemistry, Iowa State University, Ames, IA, United States, [4]Department of Statistics, Iowa State University, Ames, IA, United States, [5]Research Information Technology, College of Liberal Arts & Sciences, Iowa State University, Ames, IA, United States, [6]Department of Biological Sciences, Mississippi State University, Mississippi State, MS, United States

Research strategies that combine molecular data from multiple levels of genome expression (i.e., multi-omics data), often referred to as a systems biology strategy, has been advocated as a route to discovering gene functions. In this study we conducted an evaluation of this strategy by combining lipidomics, metabolite mass-spectral imaging and transcriptomics data from leaves and roots in response to mutations in two *AuTophaGy-related* (*ATG*) genes of *Arabidopsis*. Autophagy is an essential cellular process that degrades and recycles macromolecules and organelles, and this process is blocked in the *atg7* and *atg9* mutants that were the focus of this study. Specifically, we quantified abundances of ~100 lipids and imaged the cellular locations of ~15 lipid molecular species and the relative abundance of ~26,000 transcripts from leaf and root tissues of WT, *atg7* and *atg9* mutant plants, grown either in normal (nitrogen-replete) and autophagy-inducing conditions (nitrogen-deficient). The multi-omics data enabled detailed molecular depiction of the effect of each mutation, and a comprehensive physiological model to explain the consequence of these genetic and environmental changes in autophagy is greatly facilitated by the *a priori* knowledge of the exact biochemical function of the ATG7 and ATG9 proteins.

KEYWORDS

**autophagy, multi-omics, lipidomics, RNA-seq, arabidopsis**

## Introduction

Beginning with the investments that supported the completion of the Human Genome Project (www.genome.gov/human-genome-project), many analytical advances have enabled the rapid expansion in the number of genomes that are now fully sequenced. Thus, at the time of writing (July, 2022) NCBI had cataloged over 70,000 annotated genomes (https://

www.ncbi.nlm.nih.gov/genome/browse#!/overview), and this number is rapidly expanding. For example, the Earth BioGenome Project (www.earthbiogenome.org) envisions sequencing the genomes of all known eukaryotic organisms; approximately 1.8 million species.

With the prospect of having access to such an enormous data resource, namely the genetic blueprint of a large portion of the earth's eukaryotic biosphere, one can ask the question, are we, the scientific endeavor, prepared to take advantage of this opportunity to better understand the complexity of biology. Although there are many ways that one can address this question, at the core is the ability to identify the structure and function of individual genes as they integrate in a system that maintains biological veracity and viability. In this context, since 2000 the plant biology research community has had access to high quality genome sequence for *Arabidopsis thaliana* (L.) Heynh (Arabidopsis Genome, 2000; M. et al., 2001), and this has empowered many advances in the plant sciences. Indeed, with research support from the National Science Foundation, the Arabidopsis 2010 projects (www.arabidopsis.org/portals/masc/projects.jsp) led to increased definitions of gene functions, as captured in The Arabidopsis Information Resource database (www.arabidopsis.org) (S. Y. et al., 2003; Reiser et al., 2017), and this progress has been cataloged in the annual reports from the Multinational Arabidopsis Steering Committee (Parry et al., 2020). Analogous resources are now being assembled for other plant species, with obvious practical expansion to the major crops (e.g., maize, rice, wheat etc.). Yet despite these advances, approximately 30% of the annotated genes in most sequenced genomes are still annotated as either "hypothetical genes" or "genes of unknown function", and even for those genes that are annotated with some functionality, these are often with ambiguous annotations (e.g., Lobb et al., 2020). In addition, more recent genomic characterizations have identified the potential of "orphan genes" to determine and regulate novel traits (Arendsee et al., 2014).

Overcoming this challenge with new sophisticated instrumentation has empowered the measurement of numerous molecular attributes of biological systems in a high throughput manner, which has enabled the global expression profile of the genome at the level of the transcriptome, proteome and metabolome. Collecting these comprehensive datasets are often justified as resources that will provide insights to gene-function. Yet, despite this enhanced ability to generate multi-platform data, it's not obvious how these data inputs have enhanced the ability to discover new gene-functions. In this study we evaluated this question by applying integrated transcriptomics and metabolomics strategies on two *Arabidopsis* mutants in genes whose biological function is known by the fact that these mutations affect the process of autophagy, and *in vitro* characterizations have provided accurate insights on the biochemical functions of each gene-product.

Autophagy is a degradation pathway that engulfs organelles and macromolecules into vesicles, called autophagosomes and thereby transfers them to the vacuole, where they are degraded, and breakdown products are recycled (Yu et al., 2018). Genetic dissection of autophagy has identified a collection of mutants that affect this process, and these are collectively known as *AuTophaGy-related* (*ATG*) genes. These were initially characterized in the yeast, *Saccharomyces cerevisiae* (Tsukada and Ohsumi, 1993; Farre and Subramani, 2016; Majeed et al., 2022), and homologs have also been characterized from a wide range of eukaryotes (Yin et al., 2016), including plants (Soto-Burgos et al., 2018). For the purpose of this study, we selected *Arabidopsis* mutants in two of these genes, *ATG9* (AT2G31260) and *ATG7* (AT5G45900), whose biochemical functions have been identified by extensive characterizations, particularly in yeast and humans. Moreover, the specific function of ATG9 in plants is based on structural conservation of orthologs from *Arabidopsis* (Lai et al., 2020), yeast (Matoba et al., 2020) and humans (Guardia et al., 2020), which speaks to the phylogenetic conservation of the autophagic process. Thus, ATG9 is a lipid scramblase (Matoba et al., 2020) and ATG7 is an E1-like enzyme that catalyzes the covalent conjugation of the ATG12 and ATG5 proteins (Mizushima et al., 1998) and covalent attachment of phosphatidylethanolamine to the ATG8 protein (Ichimura et al., 2004).

Specifically, the data generated herein were used to evaluate new insights concerning these two autophagy genes (i.e., *ATG7* and *ATG9*). But more broadly for understanding processes in plant system, the aim of the study was to assess how accurately multi-omics data (i.e., integrated transcriptomics and metabolomics data) can be used to predict gene functionalities without *a priori* knowledge concerning the specific functions of genes.

# Results

## Experimental platform to explore relationships between gene expression, lipids and autophagy

The experimental platform schematically illustrated in Supplemental Figure 1, was designed to compare three *Arabidopsis* genotypes, WT and two mutants that are blocked in autophagy (i.e., *atg7* and *atg9*), all in the Col-0 ecotype background. These genetic stocks were grown in normal conditions (i.e., (+N)-condition) and in conditions that were designed to induce autophagy, by depriving nitrogen from the growth medium (i.e., (-N)-condition). Three days after this environmental induction of autophagy, seedlings were harvested in triplicate for lipidomics and transcriptomics analyses. The harvested seedlings were separated into the root and aerial tissues, which were separately analyzed in order to evaluate if autophagy is distinct between these two organs. Initially a small segment of the roots was stained with monodansylcadaverine and examined by fluorescence microscopy to visualize autophagosomes (Contento et al., 2005), which confirmed that the autophagy process was induced in the WT plants by the (-N)-treatment, whereas this process was inhibited in the *atg7* and *atg9* mutants (Hanaoka et al., 2002; Thompson et al., 2005). Regardless of the seedling genotype, all seedlings showed an obvious yellowing in the autophagy-inducing condition, which began to appear approximately 48-hours after the (-N) treatment.

## Changes in patterns of gene expression

Sequencing of the RNA and statistical analysis of the read-counts from each of the resulting tissue samples identified the differentially expressed genes (DEGs) whose expression was affected by the genetic and/or environmental alteration in the autophagic state of the seedlings. The statistical significance of these comparisons was evaluated at two levels of significance, $q<0.05$ and $q<0.1$ (Table 1;

Supplemental Table 1). The entire list of DEGs identified by these comparisons are provided in Supplemental Table 2. Because the evaluations using a statistical significance level of q<0.1 generated more comprehensive lists of DEG that ultimately provided higher confidence in the GO enrichment analysis, all subsequent evaluations of DEGs were conducted using this latter statistical significance level.

These comparisons indicate that the environmental condition that stimulated autophagy (i.e., (-N)-treatment) induced major changes in gene expression among all genotypes and organs. The *atg7* mutant showed the largest number of DEGs in the (-N)/(+N) comparison (~8,300 DEGs in leaves and ~5,200 DEGs in roots), as compared to the numbers observed in the WT genotype (~6,700 DEGs in leaves and ~3,400 DEGs in roots). In contrast, the *atg9* mutant showed the lowest number of DEGs in this comparison (~3,200 DEGs in leaves and ~3,700 in DEGs in roots). In all these comparisons, the number of up-regulated DEGs were 10-40% greater than the number of down-regulated DEGs.

The conclusions based on the (-N)/(+N) induced changes in the transcriptomes are reinforced when we compared the number of DEGs observed in the *atg7* vs. WT and *atg9* vs. WT comparisons in the (-N) condition. Namely, the *atg7* mutation has the larger effect in altering the transcriptome (i.e., ~5000 and ~1200 DEGs in leaves and roots, respectively), than the *atg9* mutation (i.e., 350 and<100 DEGs in leaves and roots, respectively). These rather large differences in the altered transcriptomes between the two mutants and WT, which are induced by the (-N) treatment, contrast to the situation we found when these comparisons were made in the (+N) condition. Specifically, in the (+N) condition there were only 24 and ~250 DEGs in leaves and roots, respectively, between *atg7* and WT plants, and for the *atg9* mutant these respective numbers were 34 and 58 DEGs. In most of these comparisons (with the exception of the *atg9* mutant), the leaf tissue is considerably more responsive than the root tissue to changes in the (-N)-treatment and/or genotype.

There is very small overlap among the lists of DEGs that respond to the (-N)-treatment between the *atg7* mutant and the *atg9* mutant. Namely, of the ~5000 DEGs that are detected in the *atg7* (-N) vs WT (-N) leaf comparison, only 312 also occur in the *atg9* (-N) vs WT (-N) leaf comparison (Figure 1A). The same conclusion is drawn from the comparison of root transcriptomes; namely, of the ~1200 DEGs in the *atg7* (-N) vs WT (-N) comparison, only 43 also occurred in the *atg9* (-N) vs. WT (-N) root comparison (Figure 1A). However, when we considered the smaller number of DEGs that are detected in the *atg9* (-N) vs WT (-N) comparison, a large majority of these (between 50% and 90%) are also DEGs in the *atg7* (-N) vs WT (-N) comparisons of both roots and leaves (Figure 2A). It should be noted that a small number of these latter DEGs (<10) switched from being induced or suppressed in one mutant to being suppressed or induced, respectively, when one compares their expression in the other mutant.

Functional insights from these DEG lists were gained by conducting GO enrichment analysis *via* the sequential use of the annotation pipelines, DAVID (Huang et al., 2009; d. W. Huang and Lempicki, 2009) and REVIGO (Supek et al., 2011). These analyses identified 177 GO enrichment terms that are >5-fold enriched among the down-regulated and upregulated DEG lists. These terms were further categorized as being either common between the two tissues (i.e., roots and leaves) or they are uniquely associated with either the root or leaf tissue. Supplemental Figure 2 identifies the 15 GO terms that are common to both leaves and roots, and these include "secondary metabolic process" and "autophagy" terms in *atg7* vs. WT comparison, and the "lipid binding" term in *atg9* vs. WT comparison. Figures 1B, 2B show the GO terms that differed among the two tissues (i.e., leaves and roots). These GO terms were further categorized into five higher level functional annotations, i.e., metabolism, DNA replication/cell cycle, signaling, photosynthesis or lipid metabolism, and these higher order functionalities are color coded in Figures 1B, 2B. These analyses indicate that the *atg7* and *atg9* mutations affect different down-stream processes, and in each mutant the processes that are altered in leaves are different from those that are altered in roots. Specifically, in the *atg7* mutant the upregulated DEGs in leaves are enriched with GO terms associated with metabolism and signaling functionalities, whereas the down-regulated DEGs in this organ are associated with DNA replication/cell cycle functionalities (Figure 1B). The GO terms that are categorized in the signaling functionality identify processes that are specifically associated with autophagy (e.g., pre-autophagosomal structure and autophagosome) and with a combination of stresses that are possibly mediated by changes in the oxidative status of the root tissue (e.g., hypoxia and anoxia, glutathione and ascorbate metabolism, and

TABLE 1 Number of differentially expressed genes in leaves and roots of WT, and *atg7* and *atg9* mutants grown in normal (+N) or autophagy-inducing (-N) conditions.

| | Leaf | | | | Root | | | |
|---|---|---|---|---|---|---|---|---|
| | Up-regulated | | Down-regulated | | Up-regulated | | Down-regulated | |
| Comparison | Genome-wide[a] | Lipid genes[b] | Genome-wide | Lipid genes | Genome-wide | Lipid genes | Genome-wide | Lipid genes |
| 1 WT(-N)/WT(+N) | 3768 | 122 | 2994 | 117 | 1850 | 112 | 1574 | 35 |
| 2 atg7(-N)/atg(+N) | 4355 | 160 | 3974 | 116 | 3074 | 176 | 2157 | 38 |
| 3 atg9(-N)/atg9(+N) | 1881 | 82 | 1332 | 45 | 2186 | 125 | 1508 | 28 |
| 4 atg7/WT, (+N) | 18 | 1 | 6 | 0 | 95 | 2 | 152 | 11 |
| 5 atg7//WT, (-N) | 2301 | 98 | 2741 | 49 | 812 | 48 | 397 | 13 |
| 6 atg9/WT, (+N) | 11 | 0 | 23 | 1 | 19 | 1 | 38 | 2 |
| 7 atg9/WT, (-N) | 86 | 4 | 264 | 6 | 38 | 0 | 45 | 4 |

[a]Among all annotated genes in the Arabidopsis genome that show statistically significant differential expression (At a statistical significance level of <0.1).
[b]Among the lipid metabolism associated genes identified in the AraLip database (http://aralip.plantbiology.msu.edu) that show statistically significant differential expression (At a statistical significance level of <0.1).

10.3389/fpls.2023.1093358

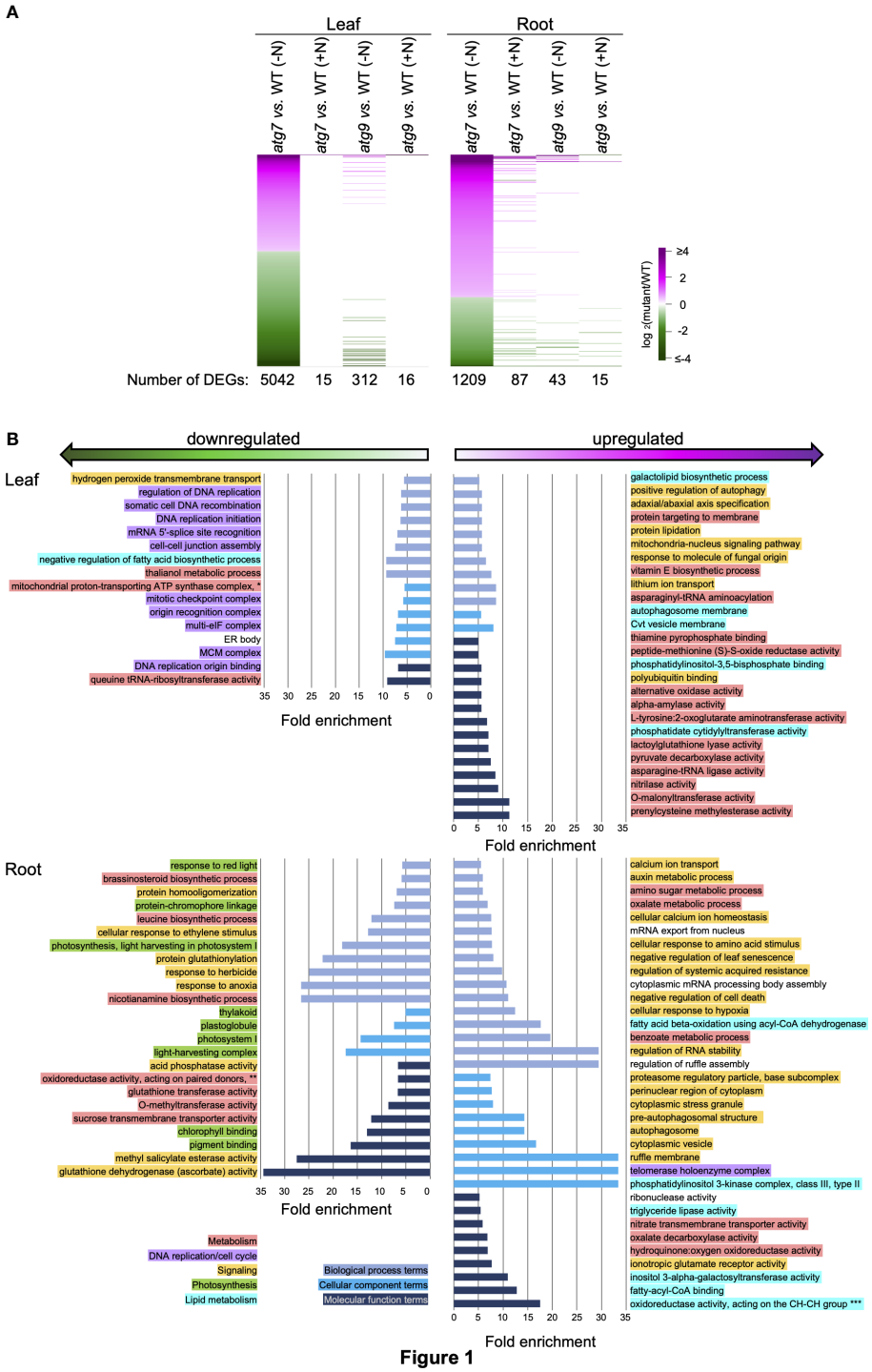



**Figure 1**

FIGURE 1
Biological functionalities identified by DEGs induced by the *atg7* mutation in leaves and roots. **(A)**. Heat map visualization of the relative expression of 5042 and 1209 DEGs identified by comparing transcript abundances in leaves and roots of the *atg7* mutant versus WT grown in the (-N) condition. The expression of each of these DEGs is compared to their relative expression levels in the other indicated comparisons. The color scale indicates the relative expression level of each DEG in each comparison. **(B)**. Using the list of upregulated and downregulated DEGs for leaves and roots, GO terms that show >5-fold enrichment were generated by the sequential use of DAVID and REVIGO. The GO terms that are unique to the altered transcriptome of the leaves and roots are presented. GO enrichment terms that are shared between the leaves and roots are presented in Supplemental Figure 2. The graphs plot the fold-enrichment for each GO enrichment terms associated with Biological Processes (■), Cellular Component (■) and Molecular Function (■). Each GO enrichment term is categorized with higher level functionalities associated with metabolism (■), DNA replication (■), signaling (■), photosynthesis (■) and lipid metabolism (■). GO terms: *mitochondrial proton-transporting ATP synthase complex, coupling factor F(o) **oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors ***oxidoreductase activity, acting on the CH-CH group of donors, with a flavin as acceptor.

10.3389/fpls.2023.1093358

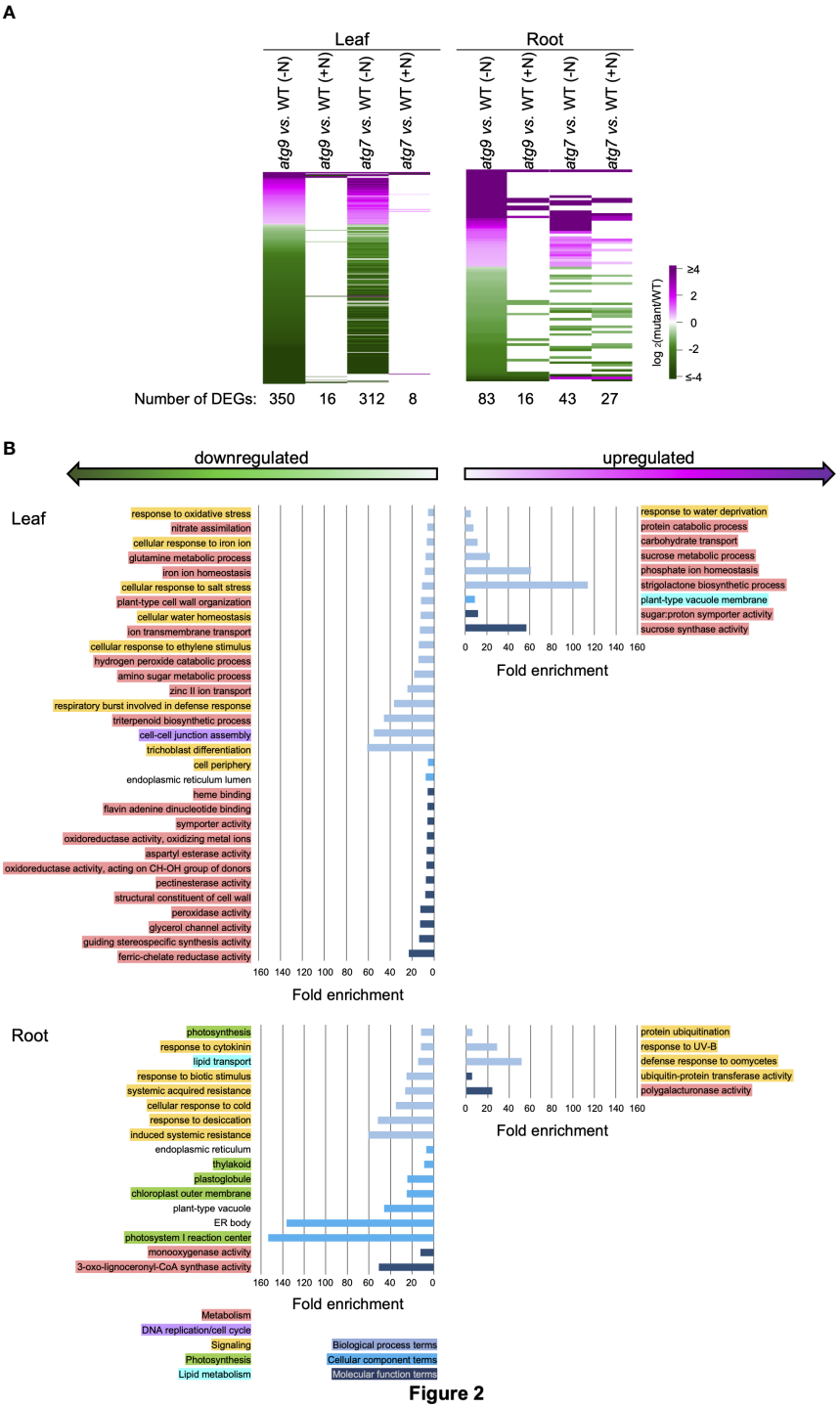



**Figure 2**

FIGURE 2
Biological functionalities identified by DEGs induced by the *atg9* mutation in leaves and roots. **(A)**. Heat map visualization of the relative expression of 350 and 83 DEGs identified by comparing transcript abundances in leaves and roots of the *atg9* mutant versus WT grown in (-N) condition. The expression of each of these DEGs is compared to their relative expression levels in the other indicated comparisons. The color scale indicates the relative expression level of each DEG in each comparison. **(B)**. Using the list of upregulated and downregulated DEGs for leaves and roots, GO terms that show >5-fold enrichment were generated by the sequential use of DAVID and REVIGO. The GO terms that are unique to the altered transcriptome of the leaves and roots are presented. GO enrichment terms that are shared between the leaves and roots are presented in Supplemental Figure 2. The graphs plot the fold-enrichment for each GO enrichment terms associated with Biological Processes (■), Cellular Component (■) and Molecular Function (■). Each GO enrichment term is categorized with higher level functionalities associated with metabolism (■), DNA replication (■), signaling (■), photosynthesis (■) and lipid metabolism (■).

responses associated with cell death, senescence, stress granules, and herbicides). Similar analyses of the GO terms identified by the DEGs in the *atg9* mutant indicate that metabolism functionalities are affected in the leaves, whereas in the roots the predominant functionality that is affected is signaling (Figure 2B).

## Changes in the transcriptome in response to autophagy

By combining the comparisons of the transcriptomes of the *atg7* and *atg9* mutants to those of WT plants that were grown in either the normal, (+N)-condition or the autophagy-inducing, (-N)-condition, we were able to distinguish transcriptomic changes that occurred because of the change in the N-status of the plants from those that are associated with the induction of autophagy. This identification depended on the fact that the *atg7* or *atg9* mutations block autophagy, and therefore we reasoned that the changes in the transcriptome that occurred in response to the (-N) or (+N) treatment of either the *atg7* or *atg9* mutants that were different from those observed for the WT plants are associated with the process of autophagy. This strategy is schematically illustrated in the expression ratio plots shown in Supplemental Figure 3A. Specifically, the plot illustrates the expression ratio of a hypothetical gene in the (-N)/(+N) comparison in WT plants (green datapoints), and in autophagy mutant plants (orange datapoints). Genes that show an expression ratio where the green and orange datapoints are indistinguishable from each other, but show altered expression due to the (-N) treatment, were identified as responding to the (-N)-treatment of the plants but not altered by autophagy (Supplemental Table 4C). All other combinations of gene expression ratios, where the green and orange datapoints are statistically distinguishable from each other ($q<0.1$), are changing their expression levels in response to the autophagic state of the plants. These autophagy-responding genes can be further classified into three categories: a) those genes that showed enhanced differential expression in the autophagy mutant compared to the WT; b) those genes that showed reduced differential expression in the autophagy mutant compared to the WT; and c) those genes that showed opposite differential expression in the autophagy mutant than in the WT.

Using this integrated comparison strategy, we identified 863 DEGs in leaves and 239 DEGs in roots whose expression is responding to the *atg7*-defect in autophagy. In parallel we identified 14 DEGs in leaves and 8 DEGs in roots whose expression is responding to the *atg9*-defect in autophagy (Figure 3; Supplemental Figure 3B). Thus, by integrating these comparisons we were able to distinguish the smaller subset of autophagy-responding DEGs from the larger number of DEGs that are responding to the change in the N-status of the plants (Table 1). The functionalities of the genes that are responding to the (-N)-treatment (Supplemental Figures 4–6) or to the autophagy defects (Figure 3) were explored by conducting GO enrichment analysis.

Consistent with prior studies, which indicate that the *atg7* mutant has a more severe phenotypic defect than the *atg9* mutant (Shin et al., 2014), we found a larger number of DEGs that are responding to the *atg7*-defect in autophagy than the number responding to the *atg9*-defect in autophagy. The Venn diagrams shown in Supplemental Figure 3B indicate that there is very little overlap among the autophagy-responsive genes in the *atg7* and *atg9* mutations, and similarly, there is a very small overlap in these autophagy-responsive genes between leaf and root tissues. These findings suggest that in the absence of either *ATG7* or *ATG9* gene-functions there are distinct down-stream effects, which may also be reflective of the differential expression pattern of these genes between leaves and roots.

Because of the small number of genes whose expression is affected by autophagy due to the *atg9* mutation (22 DEGs), it was difficult to further assess functionalities that are associated with this genetic defect. However, such analyses were pursued with the gene list identified from the *atg7* mutant (Figure 3A). Only a small fraction (~5%) of these autophagy-responsive DEGs occur in both leaves and roots, and the majority (~80%) are unique to leaves. Using the list of leaf and root DEGs as inputs we conducted GO term enrichment analysis to identify a broad overview of the types of processes that are affected by autophagy. These analyses identified 92 GO terms that show statistically significant enrichment ($p< 0.1$) (Supplemental Figure 7), and Figure 3 focuses on those GO terms that show more than 5-fold enrichment. Comparison of the GO terms recovered with the list of leaf and root DEGs indicates that different processes are being affected by autophagy in the two organs. This difference is most pronounced with the GO Biological Process terms, which are populated by signaling processes in roots, whereas in leaves they are populated by metabolism associated processes. This enrichment of metabolism associated terms is also apparent in the GO Molecular Function category.

## Changes in the lipidome

Lipidomics data were gathered in parallel to the described transcriptomics data. These analyses identified 156 specific lipid species, which included LysoPC, LysoPE, LysoPG, PC, PA, PG, PE, PI, PS, MGDG and DGDG. Upon statistical analyses, we obtained rigorous data on the abundance changes for 97 lipid molecules. PCA analyses of the leaf lipidomics data indicates that the major driver that separates the lipidomes of these samples is associated with the nitrogen treatment of the seedlings (Figure 4A, Supplemental Figure 8). Namely, PC1 that accounts for ~54% of the variance in the data, separates samples based upon whether they were grown on (+N) or (-N) media. Separation of the sample by genotype is captured by PC2, and accounts for only 7% of the variance in the data.

Figure 4B illustrate that there are only a limited number of lipids species whose abundances are significantly affected by the *atg7* mutation (3 and 14 specific lipids in the (+N) and (-N) conditions, respectively), and these appear to be enriched with lyso-PC, PC, PI and MGDG lipids. As with leaf tissue, the effect of the *atg7* mutation on the root lipidome was more evident in the (-N) treatment, with the abundances of 13 specific lipids being altered in the mutant (Figure 4B), whereas in the (+N) condition, no lipid species where affected by the *atg7* mutation (Figure 4B). Comparison of the lipidomes of roots and leaves indicate that the *atg7* mutation induced different changes in the two organs; only ~15% of the altered lipids are shared between leaves and roots. In contrast to

10.3389/fpls.2023.1093358

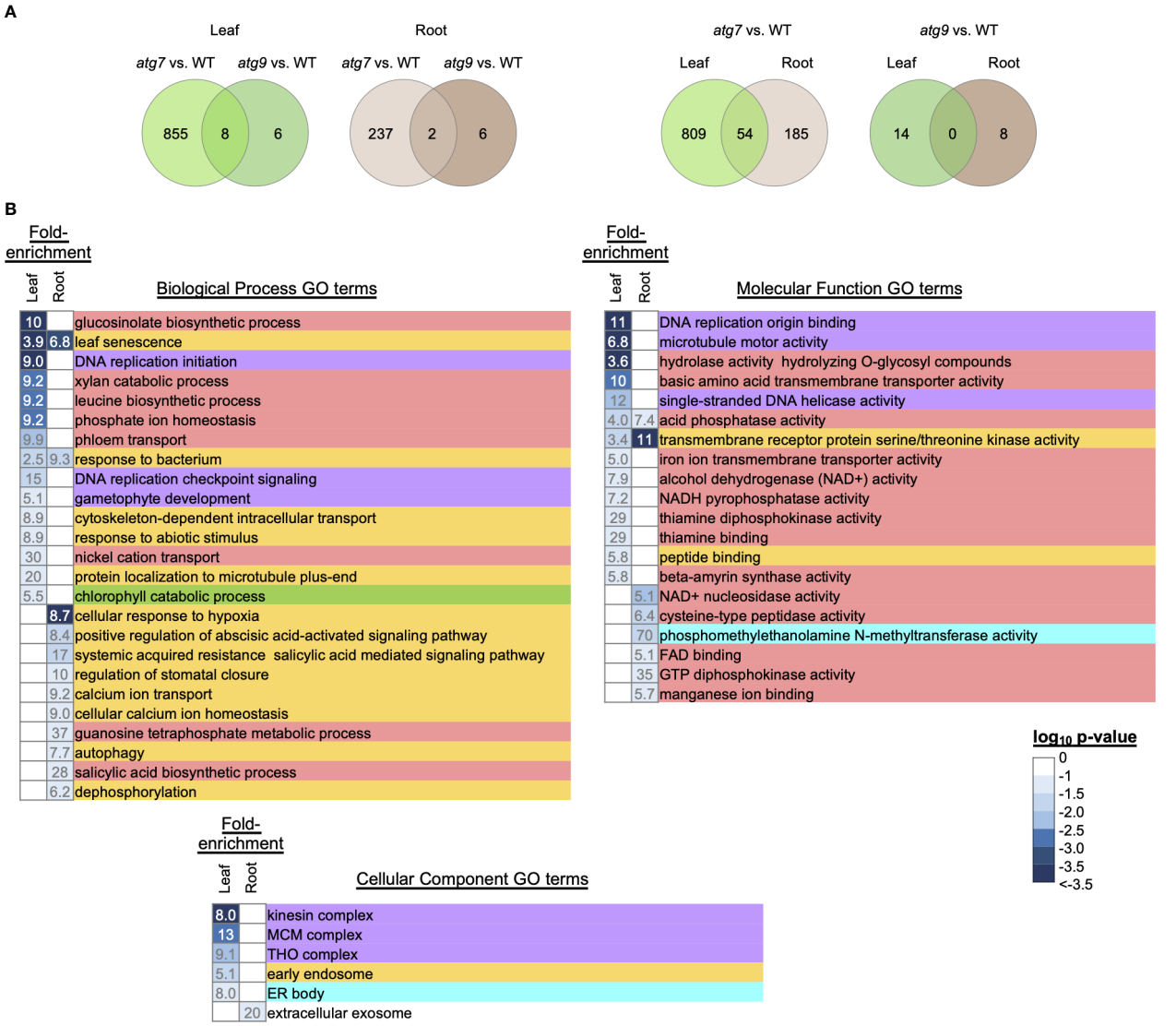



FIGURE 3
GO enrichment analysis of genes that are differentially expressed upon alterations in autophagy. DEGs that are responding to the alteration in autophagy were identified by the strategy described in Supplemental Figure 3. Venn diagrams identify the overlap of the autophagy responsive DEGs that were due to the *atg7* or *atg9* mutations in either leaves or roots **(A)**. The leaf and root GO terms showing >5-fold enrichment were identified by using the list of autophagy responsive DEGs that were responding to the *atg7* mutation **(B)**. The intensity of the blue color shading of the fold-enrichment data is proportional to the $\log_{10}$ p-value for each GO term as determined by DAVID. Each GO enrichment term is color coded to indicate higher level functional category associated with metabolism (■), DNA replication (■), signaling (■), photosynthesis (■) and lipid metabolism (■).

the effect of the *atg7* mutation on the leaf and root lipidomes, the *atg9* mutation did not affect a detectable change in the lipidomes of these organs, irrespective of whether the plants were grown in the nitrogen replete or nitrogen deficient condition (Supplemental Figure 8).

## Spatial distribution of specific lipids

We used MALDI-MSI to query the spatial distribution of specific lipids in the organs from WT and the autophagy mutants. Initial studies evaluated both roots and leaves of these plants grown either in the (+N) or (-N) conditions. Unfortunately, examination of root cross-sections by MSI proved technically difficult, probably due to the low abundance of lipids in this organ, and only a limited set of data could be collected (Supplemental Figure 10). Therefore, a complete set of MS images were collected from the leaves of WT, and *atg7* and *atg9* mutant plants that were grown either in the (+N) or (-N) conditions. To ensure reproducibility of the data, all images were generated from at least three different biological samples, and Supplemental Figure 9 shows images of an exemplary set of triplicate data of the lipid, LysoPG(18:2). Similar reproducibility was obtained with all the lipid images, and Figure 5 shows representative images for 13 distinct phospholipids. These lipid distribution images were selected to ensure we determined the spatial distribution of a wide range of phospholipids, and warranted by a robust imaging signal, which ensured reliable reproducibility in the collected data. These data indicate that of the lipids that were imaged, there is no apparent differential distribution of each

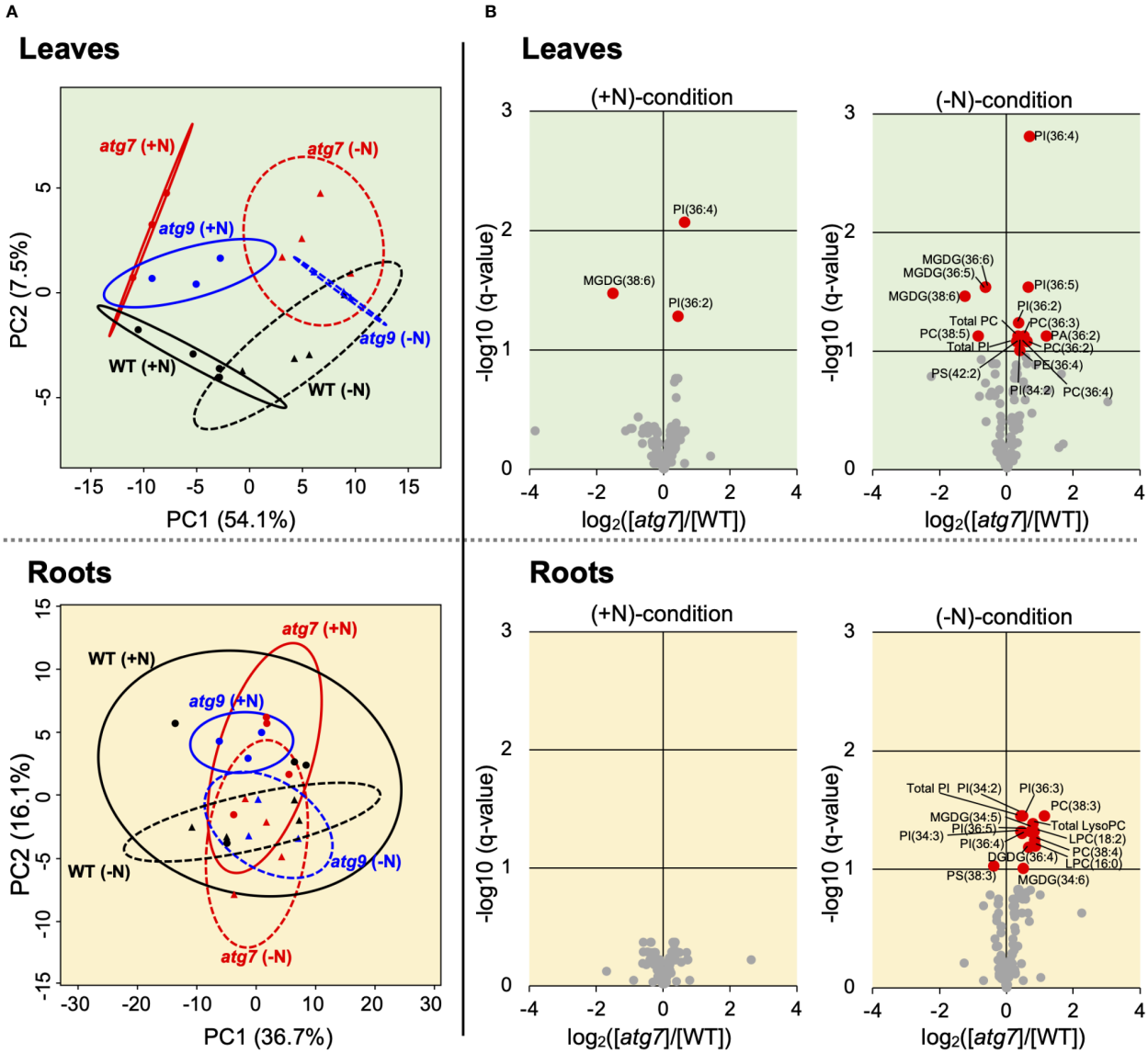



FIGURE 4
Changes in the leaf and root lipidomes induced by the *atg7* or *atg9* mutations and modified by (-N) treatment. PCA analysis of lipidomics data obtained from leaves and roots of WT (n=4), *atg7* (n=4) and *atg9* (n=3) mutant plants grown in either (+N) or (-N) conditions. The 95% confidence ellipses of each treatment are indicated **(A)**. Volcano plot visualization of the differences in the leaf and root lipidomes between WT and *atg7*-mutant grown in (+N) or (-N) conditions. q-values were obtained from p-values, which were based on the Wald test (n = 4). Red-colored datapoints represent statistically significant difference in accumulation of specific lipids **(B)**. Parallel volcano plots of the *atg9* mutant data are presented in Supplemental Figure 8.

phospholipid molecule among the cell populations of these leaves, irrespective of the autophagic status of the seedling determined by the combination of genotype and nitrogen treatment.

## Changes in the expression of lipid metabolism associated genes

Gene loci encoding functions associated with lipid metabolism in *Arabidopsis* are categorized in the AraLip database (http://aralip.plantbiology.msu.edu/) (Li-Beisson et al., 2013). AraLip identifies 775 *Arabidopsis* genes that have either been demonstrated to be involved in lipid metabolism or are hypothesized to be associated with such processes. We examined the transcriptome data to specifically identify those lipid genes whose expression were affected by either the *atg7* or *atg9* mutation in the context of the nitrogen treatment. These analyses identified 197 lipid genes (i.e., the altered lipid transcriptome) whose expression was significantly affected by either the *atg* mutations or the nitrogen treatment (Figure 6; Supplemental Table 3). The altered lipid-related transcriptome is larger in the *atg7* mutant than in the *atg9* mutant, and in both mutants the size of this altered transcriptome is increased significantly by the (-N) treatment (Figure 6); specifically, there are 184 lipid-related DEGs in the *atg7* mutant, and only 13 such DEGs in the *atg9* mutant. This parallels the changes we observed in the lipidome of the *atg7* and *atg9* mutant seedlings, with hardly any detectable changes in the lipids of the latter mutant (Figure 4; Supplemental Figure 8).

10.3389/fpls.2023.1093358

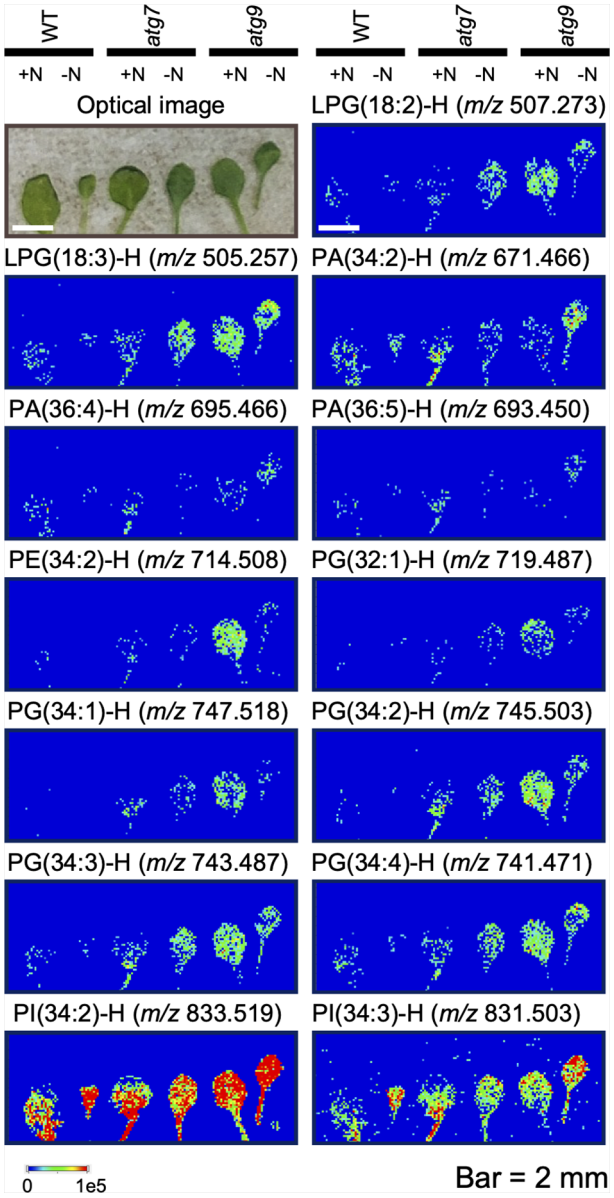



FIGURE 5
Mass spectrometric imaging of the spatial distribution of specific phospholipids in WT, *atg7* and *atg9* mutant leaves. One of the first pair of true leaves from plants grown in either (+N) or (-N) treatment of the indicated genotypes were used in these imaging experiments. Leaves were longitudinally fractured to reveal the leaf interior, and this exposed surface was imaged optically or by mass-spectrometry. All mass-spectrometric images detected the deprotonated ion ($[M-H]^-$) of the indicated phospholipid molecular species in negative mode using 1,5-diaminonaphthalene (DAN) as the matrix. The colored scale bar shows the relative intensities within a sample. The ion signals were normalized to the total ion count and the maximum values of $1x10^5$ was used for all the images.

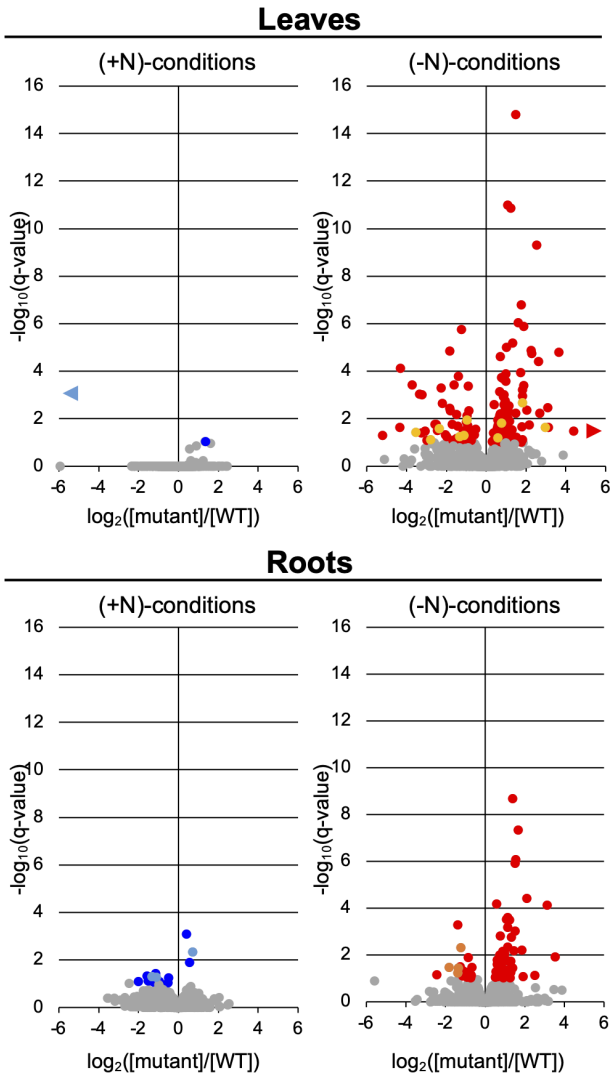



FIGURE 6
Changes in expression of lipid metabolism genes. Volcano plot visualization of lipid metabolism DEGs, comparing relative abundance of individual transcripts in the *atg7*- (●, ●) or *atg9*-mutant (●, ●), in leaves and roots from plants grown either in (+N)-or (-N)-conditions. Grey data-points (●) represent transcripts whose relative abundance is statistically indistinguishable in these comparisons (q > 0.1). Data points that are out of the range of the x-axis are identified by horizontal isosceles triangles (▶), which point to the direction of the data-point, either to the left or right of the zero x-axis value. Lipid metabolism genes were identified in AraLip (Li-Beisson et al., 2013).

Moreover, the size of the altered lipid-related transcriptome is larger in the leaves than in the roots [i.e., more DEGs in leaves (146 genes) than roots (75 genes)] (Figure 6). In the case of the *atg7* mutant grown in the (-N) condition, where the altered lipid-related transcriptome is sufficiently large to provide statistical rigor, the number of the DEGs that are upregulated is between 2- and 4-fold larger than the number of down regulated genes, in both leaves and roots.

Insights into the lipid-related gene functionalities that were affected by these treatments were gleaned by identifying the lipid pathways pinpointed by the DEGs (Supplemental Table 3). These annotations identified 15 lipid metabolic pathways, which can be classified as either a) biosynthetic (i.e., fatty acid biosynthesis, elongation and metabolism, chloroplastic lipid metabolism, cuticle and wax esters deposition and triacylglycerol biosynthesis); b) catabolic (i.e., fatty acid degradation, triacylglycerol metabolism and lipases), c) membrane lipid reorganization (i.e., lipid transport and binding proteins, and phospholipid metabolism) and d) signaling lipids (i.e., phosphoinositide, oxylipin and sphingolipid metabolism). Because the affected lipid-related transcriptome is larger in the *atg7* mutant than the *atg9* mutant, we were

able to surmise that in the leaves and roots of the former mutant, the down-regulated genes in the (-N) treatment are enriched with lipid biosynthetic pathways and pathways that reorganize membrane lipids In contrast, the up-regulated genes are enriched in processes associated with lipid catabolism and the metabolism of signaling lipids.

## Discussion

The premise for this study was to assess the ability of multi-omics data, specifically integrated lipidomics and transcriptomics data, to enable the deduction of gene function without specific *a priori* knowledge concerning the biochemical function of the gene product(s). We focused the study on two genes, *ATG7* (At5g45900) and *ATG9* (AT2G31260), which are known to be components of the autophagic system of plants because they are homologs of the yeast genes involved in this process, and mutations in these two *Arabidopsis* genes block the process of autophagy. In yeast approximately 40 *ATG* genes have been characterized, and the molecular model of the process of autophagy is primarily based on the characterization of these genes and gene products (Farre and Subramani, 2016; Majeed et al., 2022). Multi-omics studies of *Arabidopsis atg* mutants have been interpreted in the context of this autophagy model, and this includes for example the characterizations of *atg4, atg5*, *atg9, atg7 and atg18* mutants (Masclaux-Daubresse et al., 2014; Avin-Wittenberg et al., 2015; Havé et al., 2019).

The *Arabidopsis atg7* mutants are completely blocked in autophagy and are hypersensitive to shortage in nutrients (Doelling et al., 2002), and carbon- and N-deficiency (Thompson et al., 2005; Chung et al., 2010). Prior detailed biochemical characterizations of *ATG7* have established that it encodes an enzyme that is similar to the ATP-dependent, E1 ubiquitin-activating enzyme, which catalyzes the covalent conjugation of ATG5 and ATG12, a process required for autophagy (Mizushima et al., 1998). The resulting ATG5-ATG12 conjugate connects to the dimeric ATG16 protein, and the complex is incorporated into the phagophore, allowing for lipidation of ATG8 with a phosphatidylethanolamine molecule (Ichimura et al., 2000; Hanada et al., 2007; Chung et al., 2010). These biochemical processes result in the recruitment of ATG8 and cargo macromolecules into autophagosomes, and sealing of the new organelle (Li and Vierstra, 2012; Bao et al., 2017; Martens and Fracchiolla, 2020).

The *ATG9* gene encodes the only integral membrane protein of the core autophagy machinery, and it functions in the development and expansion of the autophagosome (Lang et al., 2000). Mutations in the *ATG9* gene results in the failure to release autophagosomes from the ER (Zhuang et al., 2017), which is probably due to the inability to incorporate membrane lipids into the developing organelle. This supposition is based on *in vitro* characterization of ATG9 that identified it as a scramblase (Maeda et al., 2020; Orii et al., 2021), which together with ATG2 (Maeda et al., 2019) incorporates preexisting ER-derived phospholipids and newly synthesized lipids into the nascent and growing autophagosome (Schütter et al., 2020). This mechanistic understanding of ATG9 function is based on structural characterization of ATG9-orthologs from *Arabidopsis* (Lai et al., 2020), yeast (Matoba et al., 2020) and humans (Guardia et al., 2020), which is indicative of the phylogenetic conservation of the autophagic process across a wide range of species (Zhang et al., 2021).

The molecular characterization of mutants by a multi-omics systems biology strategy, as conducted in this study, cannot provide analogous mechanistic insights on gene-function. Rather, the lipidomics and transcriptomics data that have been presented enable more detailed molecular depiction of the effect of each mutation, namely the molecular phenotype of the mutations. Moreover, the biochemical mechanistic understanding of the function of the *ATG7* and *ATG9* genes products provides the basis for proposing a more comprehensive physiological model to explain the consequence of the disruption in each of these gene functions. Hence, understanding the physiological function of genes requires the need for detailed biochemical characterizations of gene products, and the multi-omics data associated with the characterization of strains that are missing those genetic functions. Additional physiological insights could also be gained when these gene products are overexpressed, although such genetic manipulations do not always impact changes in the molecular phenotype. Previous omics-type studies support the idea that it is difficult to generalize about function from genome-scale data due to indirect effects that may differ among different genotypes and species. For example, in maize, a mutant defective in autophagy presents more molecular phenotypes in nitrogen-replete conditions than in nitrogen-deficient conditions (McLoughlin et al., 2018), whereas our analysis and a previous study using adult *Arabidopsis* mutant plants (Masclaux-Daubresse et al., 2014), found more molecular phenotypes upon nitrogen deficiency. Therefore, although there appear to be different responses in lipid profiles associated with the genetic and environmental disruption of autophagy, these differences require a mechanism of lipid recycling, which has previously been reported in both maize and *Arabidopsis* autophagy mutants (McLoughlin et al., 2018; Havé et al., 2019).

Specifically, the lipidomics and transcriptomics data generated within this study provide insights on the molecular response of *Arabidopsis*, downstream of the genetic block associated with the *atg7* and *atg9* mutations. The initial observation from these studies was the finding that the *atg7* mutation has a greater impact than the *atg9* mutation on the assessed transcriptome and lipidome. Moreover, this difference in the magnitude of the molecular response to each of these mutations was reflected in both organs that were assessed, i.e., leaves and roots. Indeed, the leaves showed a larger molecular response to these mutants than the roots. Although, the magnitude difference in the morphological phenotypes of the *atg7* and *atg9* mutants had previously been noted (Shin et al., 2014; Kang et al., 2018), the mechanism to explain this difference is not clear. This is particularly intriguing considering that the *atg9* mutation is thought to directly affect lipid-related processes, as it is required for incorporation of lipids into autophagosomes and therefore for autophagosome expansion. Despite this direct connection, the *atg9* mutant showed very little change in lipid profile, and also in lipid-related transcriptome, when compared with the *atg7* mutant. As a linear pathway cascade, it might be expected that a genetic block anywhere in the autophagy pathway should generate similar downstream responses. Yet, not only is the magnitude of the response different between the two mutants, but as visualized by GO enrichment analysis, the *atg7* and *atg9* mutations affected different processes, and these differences are also sustained as related to different organs (i.e., leaves and roots), but not the

different tissues of the leaf. Specifically, as revealed by the MSI experiments, the cellular distribution of the lipids that were imaged in the leaves of *Arabidopsis* is not differentially affected by either of the two mutants that were studied.

There are a number of potential explanations for the difference in the downstream responses between the *atg7* and *atg9* mutants. A relatively trivial explanation may be associated with the nature of the mutant alleles that were used in these experiments. For example, the weaker phenotype of the *atg9* mutant maybe due to the fact that the *atg9-4* (SALK_145980) allele (Floyd et al., 2015) used in this study is not a null allele and thus any remaining *ATG9* activity in the mutant is sufficient for the expression of a weak phenotype. Molecular characterizations have identified *atg9-4* as a sequence-indexed T-DNA insertion allele (signal.salk.edu/cgi-bin/tdnaexpress) (Alonso et al., 2003), with the insertion located in the 7$^{th}$ exon of the *ATG9* locus. Although it is possible that this allele can express a truncated ATG9 protein that would have some residual activity, we postulate that this is unlikely because characterization of the *atg9-3* allele (SALK_128991), which is disrupted by a T-DNA insertion at an upstream exon position, also shows a similar "weak" phenotype (see phenotype data concerning the *atg9-3* allele deposited at TAIR (Reiser et al., 2017). Another possibility is that the *ATG9* functionality may be redundant in the pathway, possibly as a consequence of another *ATG9*-like paralog in the *Arabidopsis* genome; however BLAST analysis of the genome indicates that such a paralog is not identifiable. There is a possibility however that the *ATG9*-function maybe provided by protein(s) that does not share sequence homology with the ATG9 protein. For example, flippases and floppases catalyze reactions, which collectively are equivalent to the scramblase reaction that ATG9 is thought to catalyze (i.e., lipid scramblase). An alternative explanation maybe associated with the observations that different ATG proteins, and therefore different branches of the autophagy pathway, maybe responding to different environmental signals that induce autophagy (Huang et al., 2019; Aroca et al., 2021). Indeed this possibility may also provide the explanation to account for the differences in the molecular phenotypes that are apparent between roots and leaves of the *atg7* and *atg9* mutants. We explored this possibility by evaluating both our own RNA-Seq data and the publicly available *Arabidopsis* transcriptome data (Sullivan et al., 2019). In the absence of any mutations, the *ATG9* transcript accumulates at ~50% higher levels than the *ATG7* transcript, irrespective of the organ. Consistent with their role in autophagy, the abundance of both the *atg7* and *atg9* transcripts is induced by ~2-fold in leaves, but such an induction does not occur in roots by treatments that induce autophagy. This lack of induction in the expression of these genes in roots, when autophagy is induced, may therefore be the basis for the relatively weak phenotype that is observed in this organ. Therefore, the reason that the autophagy response in the *atg9* mutant is weaker than the *atg7* mutant requires additional more mechanistic studies than those conducted herein.

Collectively, the observations reported in this study indicate that downstream of each of these two autophagy genes, the process of autophagy is differently expressed in different organs, and these differences manifest organ-specific phenotypes. Moreover, the multi-omics research strategy applied herein describes these phenotypes at the molecular level, rather than the more traditional whole organismal level. Yet, the explanation of the mechanism(s) that generate these molecular phenotypes requires *a priori* knowledge that can only be generated by the specific biochemical characterization of the gene products.

# Materials and methods

## Biological materials and growth conditions

Seed stocks of wild-type *Arabidopsis* (ecotype Col-0) and *atg7-2* (Chung et al., 2010) and *atg9-4* (SALK_145980) (Floyd et al., 2015) mutant lines were obtained from the Arabidopsis Biological Resource Center (Columbus, OH). Seeds were sterilized with 70% ethanol, followed by a 10-min incubation in 0.1% (v/v) Tween 20 (Thermo Fisher Scientific, Waltham, MA) and 50% (v/v) bleach solution. The seeds were then washed with sterile water, at least three times. Subsequently, the suspended seeds were vernalized by incubating at 4°C in darkness for 2 days.

After vernalization, the seeds were suspended in sterile 0.1% (w/v) agarose (VWR, Radnor, PA), and sown on 3.5 cm x 4 cm autoclaved stainless steel growth mesh (14 Mesh T304 Woven Stainless 0.017" wire diameter, TWP Inc. Berkeley, California), which was laid on ½ strength Murashige and Skoog (MS) solid medium composed of 2.15 g/L Murashige and Skoog Basal Salt Mixture (MilliporeSigma, Burlington, MA), 0.05% (v/v) Murashige and Skoog Vitamin Solution (MilliporeSigma), 1% (w/v) sucrose (Thermo Fisher Scientific), 6g/L Phytoblend Agar (Caisson Labs, Smithfield, UT), and 2mM MES (MilliporeSigma) at pH 5.7. Each 10 cm x 10 cm square Petri dish, containing 4 growth meshes, was placed in a growth room maintained at 22 °C under continuous illumination ($50 \pm 10$ µE $m^{-2}$ $s^{-1}$) for a period of 5 days.

Subsequently, the growth mesh, carrying the germinated seedlings, was sterilely moved onto a sterile 7.5 cm x 8.5 cm stainless steel platform mesh (10 Mesh Woven Stainless 0.025" wire diameter, TWP Inc.), which was in a 11.4 cm × 8.6 cm × 6.4 cm Phytatray dish (MilliporeSigma) that contained sterile liquid medium, composed of ½ strength MS liquid media, which contained 10 mM $NH_4NO_3$ and 9.4 mM $KNO_3$ (i.e., the (+N) condition). The volume of the medium was adjusted so that the growth mesh that carried the seeds was in contact with the surface of the medium, and thus as seedlings grew the root system extended into the liquid medium. After 1 day incubation in the +N liquid medium, half the growth meshes from each Phytatray dish were moved into a Phytatray dish that contained nitrogen-deficient liquid medium (-N medium, which contains no nitrogen salts). This medium was composed of 5% (v/v) Murashige and Skoog Basal Salt Micronutrient Solution (MilliporeSigma), 0.05% (v/v) Murashige and Skoog Vitamin Solution, 1.5 mM $CaCl_2$, 0.75 mM $MgSO_4$, 0.625 mM $KH_2PO_4$, 2.5 mM KCl, 2 mM MES and 1% (w/v) sucrose. After an additional 3-day incubation, the seedlings from both the +N and -N media were harvested by cutting the hypocotyls that were extending below the growth mesh, and leaf and root tissues were collected separately.

## Autophagy status verification by fluorescence microscopy

For MDC staining, *Arabidopsis* seedlings were stained with 50 mM monodansylcadaverine (MDC) (MilliporeSigma) in phosphate-buffered

saline, pH 7.4 for 10 minutes, followed by 3 brief washes with phosphate-buffered saline, as previously described by (Contento et al., 2005). MDC-stained seedlings roots were imaged in the late elongation zone and neighboring cells in the differentiation zone, while root tips and older differentiation region cells were excluded. MDC fluorescence was visualized using a Zeiss Axio Imager.A2 upright microscope (Jean, Germany) using a 40X objective and a UV lamp and 4',6-diamidino- 2-phenylindole-specific filter (DAPI). The number of motile MDC-stained puncta in all cells visible in the focal plane was quantified and expressed per frame for at least 10 images per sample for 3 biological replicates.

## RNA extraction

Excess moisture was removed from the collected fresh tissue by blotting the tissue with dry paper towels, then the tissue was transferred to 15 mL centrifuge tubes (Thermo Fisher Scientific), flash frozen with liquid $N_2$, and stored at -80°C. About 1 g of harvested plant tissue was pulverized under liquid $N_2$ in an RNase-free mortar and pestle. RNA was extracted with TRIzol Reagents (Thermo Fisher Scientific), followed by treatment with DNase I (QIAGEN, Hilden, Germany) and clean-up with RNeasy Mini Kit (QIAGEN). The quality of RNA was analyzed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The RNA integrity number (RIN) of all samples was ≥7.8.

## RNA-Seq analysis

cDNA library construction and sequencing were conducted using a BGISEQ-500 system with 100-cycle paired-ends at BGI Americas (https://www.bgi.com/us/). Data files have been deposited in the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under accession number PRJNA769614. The annotation file in the gff3 format and the genome DNA sequence of TAIR10 Release-39 were downloaded from http://ftp.ensemblgenomes.org/pub/plants/release-39/gff3/arabidopsis_thaliana/ and http://ftp.ensemblgenomes.org/pub/plants/release-39/fasta/arabidopsis_thaliana/dna/respectively. The splicing sites are extracted from the annotation file. The reads are aligned using HISAT2 (hierarchical indexing for spliced alignment of transcripts) (V2.1.0) (Kim et al., 2015) with known splice sites, mapped reads are counted by HTSeq-count (V0.9.1) (Anders et al., 2015) for raw counts and by cufflinks (V2.2.1) (Trapnell et al., 2010), for normalized counts.

## Lipid extraction

Lipids were extracted using a modification of a standard protocol (Shiva et al., 2018). The collected plant tissues (leaf or root) were transferred to a 50 mL Teflon-lined screw-capped glass tube (Thermo Fisher Scientific) containing 3 mL preheated isopropanol (Thermo Fisher Scientific) containing 0.01% (v/v) butylated hydroxytoluene (BHT) (MilliporeSigma) and 1 µM 1,2-didecanoyl-sn-glycero-3-phosphocholine (MilliporeSigma) as an internal standard. The tubes were incubated at 75°C for 15 min to quench the action of any lipases.Following the addition of 1.5 mL chloroform and 0.6 mL water the mixture was vigorously shaken at room temperature for 1 h. The clear liquid extract was transferred to another 50 mL tube using glass Pasteur pipettes, and the remnant tissue was further extracted for 30-minutes with another 4 mL chloroform/methanol (2:1) that contained 0.01% BHT. The clear liquid from this second extraction was removed and combined with the initial liquid extract. This chloroform/methanol (2:1) extraction was repeated three times, the last extraction being incubated overnight. The residue tissue remaining after lipid extraction was dried at 105°C, and the dry weight of each sample determined; each leaf tissue sample weighed approximately 20 mg, and each root tissue sampled weighed approximately 10 mg.

All extract aliquots from each biological sample were combined into a single screw capped tube and stored at -80°C under a nitrogen gas atmosphere. The solvent from each extract was removed by evaporation with the aid of a stream of $N_2$ gas, and the lipid residue was dissolved in 1 mL chloroform and transferred to 2.0 mL clear glass vial with Teflon-lined screw cap (Thermo Fisher Scientific). The solvent was again evaporated with $N_2$ gas, and the vials were shipped overnight on dry ice to the Kansas Lipidomics Research Center (https://www.k-state.edu/lipid/) for lipidomics analysis.

## Lipidomics analysis by ESI triple-quadrupole MS

Mass spectrometric analysis of the lipidome was conducted at the Kansas Lipidomics Research Center using a Xevo TQ-S mass spectrometer (Waters Co., Milford, MA). Individual lipids were identified by direct infusion in positive ion mode with precursor and neutral loss scans (Xiao et al., 2010; Peters et al., 2010; Li et al., 2014), using the scans shown in Supplemental Table 4. Response factor corrections were applied to the MGDG and DGDG analyses to correct for differences in the response of the mass spectrometer to unsaturated galactolipid species as compared to the saturated internal standards. Phospholipid data did not require such response factor corrections, as the biological phospholipids and the internal standard have similar response factors.

The metabolomics data from this study are available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench (https://www.metabolomicsworkbench.org) (Sud et al., 2016), where it has been assigned Study ID ST002252. The data can be accessed directly *via* its Project DOI: http://dx.doi.org/10.21228/M8KQ6P.

## Statistical analysis

Lipidomics data were gathered from 44 individual samples that represented three or four replicates per +N/-N treatment for each genotype that was evaluated. Collected data were normalized using the upper-quartile normalization method (Bullard et al., 2010). Transcriptomics data were collected from 41 tissue samples that were gathered in parallel to those used for lipidomics analysis. Due to potential contamination of three root samples by leaves– indicated by assessing the abundance of transcripts encoding for 15 genes coding for photosynthetic functions that would be expected to be expressed in leaves and not in

roots [i.e., RUBISCO small subunit genes (At1g67090 and At5g38410), photosystem I subunit genes (e.g., AT1G30380 and AT1G52230)] - the root replicate samples (one of each genotype) were removed from further analysis. The remaining 38 samples that represented between three and four replicates of root and leaf tissues of the three genotypes, grown in either (+N) or (-N) conditions were further analyzed. Initially, transcripts which showed undetectable expression levels across all the samples were removed, which provided abundance data for 26093 transcripts in the root samples, and 26356 transcripts in the leaf samples. The transcriptomics data was analyzed by using the DESeq function in R package DESeq2 (Love et al., 2014). Data was normalized by using the median of ratio method (Anders and Huber, 2010) in DESeq2 package.

Three statistical tests were applied to the transcriptomics datasets:

Test 1) comparison between N-treatments for a given genotype, e.g.,

$$log_2\left[GeneExpression_{Genotype1(-N)}\right] - log_2\left[GeneExpression_{Genotype1(+N)}\right];$$

Test 2) comparison between genotypes for a given N-condition (+ or -), e.g.,

$$log_2\left[Gene\ Expression_{Genotype1(+N)}\right] - log_2\left[Gene\ Expression_{Genotype2(+N)}\right];$$

Test 3) N-treatment effect tests among the 3 genotypes, e.g.,

$$\left(log_2\left[Gene\ Expression_{Genotype1(-N)}\right] - log_2\left[Gene\ Expression_{Genotype1(+N)}\right]\right) - \left(log_2\left[Gene\ Expression_{Genotype2(-N)}\right] - log_2\left[Gene\ Expression_{Genotype2(+N)}\right]\right)$$

Pearson's correlation was calculated between logcpm, which was defined by $\log_2[10^6 \bullet (r_{i,j} + \delta)/R_j]$, where $r_{i,j}$ is the expression data for transcript $i$, $\delta$ is the half of the smallest non-zero measurement/count, $R_j$ is the upper quartile for sample $j$. The log fold change (LFC) and q-values (Nettleton et al., 2006) which were calculated from p-values were used for interpretation of the results.

Principal component analysis (PCA) was conducted using the pcomp function of the factoextra package in R. Volcano plots were generated by plotting the negative logarithm of the p value on the y-axis and the LFC on the x-axis in Excel. Heat maps of gene expression levels were created with Perseus (https://maxquant.net/perseus/) (Tyanova et al., 2016).

## Gene ontology analysis

Gene lists were subjected to analysis *via* DAVID (https://david.ncifcrf.gov/home.jsp) (Huang et al., 2009; d. W. Huang and Lempicki, 2009). Function annotation chart was generated by DAVID for each gene list, and terms for Biological Process, Cellular Component and Molecular Function were further reduced using REVIGO (http://revigo.irb.hr/), choosing "Small (0.5)" as the option for reduction (Supek et al., 2011). The fold-enrichment and p-value from DAVID is matched to the reduced lists and plotted.

## MALDI mass spectrometric imaging

Leaves were prepared for mass spectrometric imaging as previously described (Klein et al., 2015). The first pair of true leaves were collected and the abaxial side was adhered to Scotch box sealing tape (3M, Maplewood, MN). After 1-2 h drying in a vacuum, the packing tape was folded so that the adaxial surface of the leaves adhered to the tape, and the tape-encased leaf-tape "sandwich" was passed through a rolling mill to ensure adherence of the tape to the surfaces of the leaf. The two adhering tapes were separated, which resulted in fracturing of the leaf through the mesophyll cell interior, resulting in two separated pieces of packing tape, each retaining the abaxial and adaxial halves of the leaf. Each of these pieces of packing tape, with the exposed leaf surface facing upward, were adhered to a microscope slide using double-sided tape. After acquiring optical images, the samples were subjected to matrix deposition by sublimating (Hankin et al., 2007) 2,5-dihydroxybenzoic acid (DHB; 98%; MilliporeSigma) or 1,5-diaminonaphthalene (DAN; 97%; MilliporeSigma) at 140°C for 4 min at a pressure of~50 mtorr.

Mass spectrometry imaging data were collected using a MALDI-linear ion trap-Orbitrap mass spectrometer (MALDI-LTQ-Orbitrap Discovery; Thermo Fisher Scientific). The instrument was modified to incorporate an external 355 nm frequency tripled Nd : YAG laser (UVFQ; Elforlight, Daventry, United Kingdom) and a f=60 mm focus lens (Feenstra et al., 2017). TunePlus and XCalibur (Thermo Fisher Scientific) were used to define imaging parameters and to acquire data, respectively. Leaves were acquired using 100 µm raster step size. Mass spectra were acquired with 10 laser shots per spectrum mostly using an Orbitrap mass analyzer (resolution of 30,000 at *m/z* 400) for an *m/z* scan range of 200–1000. Data were obtained in positive and negative ion mode, using DHB and DAN, respectively.

MS images were generated using ImageQuest (Thermo Fisher Scientific) with a mass window of ±5 ppm, and without normalization. The same minimum and maximum plot values were used across all images and compounds. Peak assignments were based on accurate mass measurement, and comparison with online databases (METLIN, https://metlin.scripps.edu). Images were saved from a screenshot of the 2-D image window.

## 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/, PRJNA769614 http://dx.doi.org/10.21228/M8KQ6P,ST002252.

## Author contributions

BN, DB contributed to conception and design of the study. GD, YM, MD, LL gathered the data, organized the data. ML, DN, XZ performed the statistical analysis. BN, DB, YJL, DN provided mentorship and project administration. GD, YM, MD, wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was partially funded by a grant from the US Department of Energy, grant number DE-SC0014038, Hatch Project

No. IOW03802 from the United Sates Department of Agriculture's National Institute of Food and Agriculture, Iowa State University Foundation, and Iowa State University's Center for Metabolic Biology.

## Acknowledgment

The lipid analyses described in this work were performed at the Kansas Lipidomics Research Center Analytical Laboratory. Instrument acquisition and lipidomics method development were supported by the US National Science Foundation (including support from the Major Research Instrumentation program; current award DBI-1726527), K-IDeA Networks of Biomedical Research Excellence (INBRE) of National Institute of Health (P20GM103418), USDA National Institute of Food and Agriculture (Hatch/Multi-State project 1013013), and Kansas State University.

## Conflict of interest

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

## Publisher's note

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.2023.1093358/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Experimental workflow for tissue preparation. Schematic representation of the workflow used to prepare root and leaf tissue samples from seedlings grown in (+N) and (-N) conditions.

SUPPLEMENTARY FIGURE 2
GO enrichment analysis of genes that are differentially expressed in the *atg7* and *atg9* mutant. GO enrichment terms that are common to leaves and roots and identified from downregulated and upregulated DEGs in the comparison between *atg7* and WT, or *atg9* and WT, in (-N)-conditions.

SUPPLEMENTARY FIGURE 3
Identification of genes that show differential expression in response to the alteration in the autophagic state of the tissue or nitrogen status. **(A)** The graph illustrates the expression response of a hypothetical gene as the ratio of expression in (-N) versus (+N) conditions that induce autophagy in either WT plants (green datapoints) or mutant plants (either *atg7* or *atg9*) (orange datapoints). Genes that show expression patterns where the green and orange datapoints are statistically distinguishable from each other we define as genes that are responding to autophagy, whereas genes that show expression patterns where the green and orange datapoints are statistically indistinguishably separated are responding to the (-N)-treatment. The numbers of the DEGs associated with each expression pattern identified in leaves and roots of the *atg7* or *atg9* mutants are tabularized next to the illustrative graph. **(B)** Venn diagram visualization of the number of DEGs responding autophagy in the leaves and roots of the *atg7* or *atg9* mutants. **(C)** Venn diagram visualization of the number of DEGs responding to the (-N)-treatment and not to autophagy, in the leaves and roots of the *atg7* or *atg9* mutants.

SUPPLEMENTARY FIGURE 4
Biological process GO terms of the down- and up-regulated genes in (-N) condition that are associated with nitrogen status in the *atg7* vs. WT and *atg9* vs. WT comparisons. The statistical significance of the enrichment of each term, evaluated by $\log_{10}$ p-value, is indicated by the intensity of the blue color shading the fold-enrichment data.

SUPPLEMENTARY FIGURE 5
Cellular component GO terms of the down- and up-regulated genes in (-N) condition that are associated with nitrogen status in the *atg7* vs. WT and *atg9* vs. WT comparisons. The statistical significance of the enrichment of each term, evaluated by $\log_{10}$ p-value, is indicated by the intensity of the blue color shading the fold-enrichment data.

SUPPLEMENTARY FIGURE 6
Molecular function GO terms of the down- and up-regulated genes in (-N) condition that are associated with nitrogen status in the *atg7* vs. WT and *atg9* vs. WT comparisons. The statistical significance of the enrichment of each term, evaluated by $\log_{10}$ p-value, is indicated by the intensity of the blue color shading the fold-enrichment data.

SUPPLEMENTARY FIGURE 7
GO analysis of genes that are differentially expressed upon alterations in autophagy (all levels of enrichment).

SUPPLEMENTARY FIGURE 8
Volcano plot representation of the lipidome, comparing the relative abundance of individual lipids between WT and *atg9* leaf grown in (+N) condition **(A)**; between WT and *atg9* leaf grown in -N conditions **(B)**; between WT and *atg9* roots grown in (+N) condition **(C)**; and between WT and *atg9* roots grown in (-N).

SUPPLEMENTARY FIGURE 9
Evaluation of the reproducibility of mass spectrometric imaging data. The optical images and MS images of the spatial distribution of the deprotonated ion of lysoPG(18:2)-H, which was acquired in negative mode using DAN as the matrix. Images were acquired from three replicate leaves, each from a different plant, of the indicated genotype, that was grown in either in (+N)- or (-N)-conditions. Leaves were fractured longitudinally to expose the adaxial and abaxial surfaces of the leaf interior, which were imaged by mass spectrometry. The ion signals were normalized to the total ion count and the maximum values used was $1x10^5$.

## References

Alonso, J. M., Vries Blankestijn-de, H. C., Hanhart, J., and Koornneef, M.. (2003). Genome-wide insertional mutagenesis of arabidopsis thaliana. *Science* 301, 653–657. doi: 10.1126/science.1086391

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Nat. Precedings* 11 (10), R106. doi: 10.1186/gb-2010-11-10-r106

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the flowering plant arabidopsis thaliana. *Nature* 408, 796–815. doi: 10.1038/35048692

Arendsee, Z. W., Li, L., and Wurtele, E. S. (2014). Coming of age: orphan genes in plants. *Trends Plant Sci.* 19, 698–708. doi: 10.1016/j.tplants.2014.07.003

Aroca, A., Yruela, I., Gotor, C., and Bassham, D. C. (2021). Persulfidation of ATG18a regulates autophagy under ER stress in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 118. doi: 10.1073/pnas.2023604118

Avin-Wittenberg, T., Bajdzienko, K., Wittenberg, G., Alseekh, S., Tohge, T., Bock, R., et al. (2015). Global analysis of the role of autophagy in cellular metabolism and energy homeostasis in arabidopsis seedlings under carbon starvation. *Plant Cell* 27, 306–322. doi: 10.1105/tpc.114.134205

Bao, Y., Mugume, Y., and Bassham, D. C. (2017). Biochemical methods to monitor autophagic responses in plants. *Methods enzymology* 588, 497–513. doi: 10.1016/bs.mie.2016.09.090

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-seq experiments. *BMC Bioinf.* 11, 94. doi: 10.1186/1471-2105-11-94

Chung, T., Phillips, A. R., and Vierstra, R. D. (2010). ATG8 lipidation and ATG8-mediated autophagy in arabidopsis require ATG12 expressed from the differentially controlled ATG12A AND ATG12B loci. *Plant J.* 62, 483–493. doi: 10.1111/j.1365-313X.2010.04166.x

Contento, A. L., Xiong, Y., and Bassham, D. C. (2005). Visualization of autophagy in arabidopsis using the fluorescent dye monodansylcadaverine and a GFP-AtATG8e fusion protein. *Plant J.* 42, 598–608. doi: 10.1111/j.1365-313X.2005.02396.x

Doelling, J. H., Walker, J. M., Friedman, E. M., Thompson, A. R., and Vierstra, R. D. (2002). The APG8/12-activating enzyme APG7 is required for proper nutrient recycling and senescence in arabidopsis thaliana. *J. Biol. Chem.* 277, 33105–33114. doi: 10.1074/jbc.M204630200

d. W. Huang, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Farre, J. C., and Subramani, S. (2016). Mechanistic insights into selective autophagy pathways: lessons from yeast. *Nat. Rev. Mol. Cell Biol.* 17, 537–552. doi: 10.1038/nrm.2016.74

Feenstra, A. D., Dueñas, M. E., and Lee, Y. J. (2017). Five micron high resolution MALDI mass spectrometry imaging with simple, interchangeable, multi-resolution optical system. *J. Am. Soc. Mass Spectrom* 28, 434–442. doi: 10.1007/s13361-016-1577-8

Floyd, B. E., Morriss, S. C., MacIntosh, G. C., and Bassham, D. C. (2015). Evidence for autophagy-dependent pathways of rRNA turnover in arabidopsis. *Autophagy* 11, 2199–2212. doi: 10.1080/15548627.2015.1106664

Guardia, C. M., Tan, X. F., Lian, T., Rana, M. S., Zhou, W., et al. (2020). Structure of human ATG9A, the only transmembrane protein of the core autophagy machinery. *Cell Rep.* 31, 107837. doi: 10.1016/j.celrep.2020.107837

Hanada, T., Noda, N. N., Satomi, Y., Ichimura, Y., Fujioka, Y., Takao, T., et al. (2007). The Atg12-Atg5 conjugate has a novel E3-like activity for protein lipidation in autophagy. *J. Biol. Chem.* 282, 37298–37302. doi: 10.1074/jbc.C700195200

Hanaoka, H., Noda, T., Shirano, Y., Kato, T., Hayashi, H., Shibata, D., et al. (2002). Leaf senescence and starvation-induced chlorosis are accelerated by the disruption of an arabidopsis autophagy gene. *Plant Physiol.* 129, 1181–1193. doi: 10.1104/pp.011024

Hankin, J. A., Barkley, R. M., and Murphy, R. C. (2007). Sublimation as a method of matrix application for mass spectrometric imaging. *J. Am. Soc. Mass Spectrom* 18, 1646–1652. doi: 10.1016/j.jasms.2007.06.010

Havé, M., Luo, J., Tellier, F., Balliau, T., Cueff, G., Chardon, F., et al. (2019). Proteomic and lipidomic analyses of the arabidopsis atg5 autophagy mutant reveal major changes in endoplasmic reticulum and peroxisome metabolisms and in lipid composition. *New Phytol.* 223, 1461–1477. doi: 10.1111/nph.15913

Huang, X., Zheng, C., Liu, F., Yang, C., Zheng, P., Lu, X., et al. (2019). Genetic analyses of the arabidopsis ATG1 kinase complex reveal both kinase-dependent and independent autophagic routes during fixed-carbon starvation. *Plant Cell* 31, 2973–2995. doi: 10.1105/tpc.19.00066

Huang, d. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13. doi: 10.1093/nar/gkn923

Ichimura, Y., Kirisako, T., Takao, T., Satomi, Y., Shimonishi, Y., Ishihara, N., et al. (2000). A ubiquitin-like system mediates protein lipidation. *Nature* 408, 488–492. doi: 10.1038/35044114

Ichimura, Y., Imamura, Y., Emoto, K., Umeda, M., Noda, T., Ohsumi, Y., et al. (2004). *In vivo* and *in vitro* reconstitution of Atg8 conjugation essential for autophagy. *J. Biol. Chem.* 279, 40584–40592. doi: 10.1074/jbc.M405860200

Kang, S., Shin, K. D., Kim, J. H., and Chung, T. (2018). Autophagy-related (ATG) 11, ATG9 and the phosphatidylinositol 3-kinase control ATG2-mediated formation of autophagosomes in arabidopsis. *Plant Cell Rep.* 37, 653–664. doi: 10.1007/s00299-018-2258-9

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

Klein, A. T., Yagnik, G. B., Hohenstein, J. D., Ji, Z., Zi, J., Reichert, M. D., et al. (2015). Investigation of the chemical interface in the soybean-aphid and rice-bacteria interactions using MALDI-mass spectrometry imaging. *Anal. Chem.* 87, 5294–5301. doi: 10.1021/acs.analchem.5b00459

Lai, L. T. F., Yu, C., Wong, J. S. K., Lo, H.S., Benlekbir, S., Jiang, L., et al. (2020). Subnanometer resolution cryo-EM structure of *Arabidopsis thaliana* ATG9. *Autophagy* 16, 575–583. doi: 10.1080/15548627.2019.1639300

Lang, T., Reiche, S., Straub, M., Bredschneider, M., and Thumm, M. (2000). Autophagy and the cvt pathway both depend on AUT9. *J. Bacteriol* 182, 2125–2133. doi: 10.1128/JB.182.8.2125-2133.2000

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism. *Arabidopsis book / Am. Soc. Plant Biologists* 11, e0161. doi: 10.1199/tab.0161

Li, M., Baughman, E., Roth, M. R., Han, X., Welti, R., Wang, X., et al. (2014). Quantitative profiling and pattern analysis of triacylglycerol species in arabidopsis seeds by electrospray ionization mass spectrometry. *Plant J.* 77, 160–172. doi: 10.1111/tpj.12365

Li, F., and Vierstra, R. D. (2012). Autophagy: a multifaceted intracellular system for bulk and selective recycling. *Trends Plant Sci.* 17, 526–537. doi: 10.1016/j.tplants.2012.05.006

Lobb, B., Tremblay, B. J., Moreno-Hagelsieb, G., and Doxey, A. C. (2020). An assessment of genome annotation coverage across the bacterial tree of life. *Microb. Genom* 6. doi: 10.1099/mgen.0.000341

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

Maeda, S., Yamamoto, H., Kinch, L. N., Garza, C. M., Takahashi, S., Otomo, C., et al. (2020). Structure, lipid scrambling activity and role in autophagosome formation of ATG9A. *Nat. Struct. Mol. Biol.* 27 (12), 1194–1201. doi: 10.1038/s41594-020-00520-2

Maeda, S., Otomo, C., and Otomo, T. (2019). The autophagic membrane tether ATG2A transfers lipids between membranes. *Elife* 8. doi: 10.7554/eLife.45777.020

Majeed, S. T., Majeed, R., and Andrabi, K. I. (2022). Expanding the view of the molecular mechanisms of autophagy pathway. *J. Cell Physiol.* 237, 3257–3277. doi: 10.1002/jcp.30819

Martens, S., and Fracchiolla, D. (2020). Activation and targeting of ATG8 protein lipidation. *Cell Discovery* 6, 23. doi: 10.1038/s41421-020-0155-1

Masclaux-Daubresse, C., Clément, G., Anne, P., Routaboul, J. M., Guiboileau, A., Soulay, F., et al. (2014). Stitching together the multiple dimensions of autophagy using metabolomics and transcriptomics reveals impacts on metabolism, development, and plant responses to the environment in arabidopsis. *Plant Cell* 26, 1857–1877. doi: 10.1105/tpc.114.124677

Matoba, K., Kotani, T., Tsutsumi, A., Tsuji, T., Mori, T., Noshiro, D., et al. (2020). Atg9 is a lipid scramblase that mediates autophagosomal membrane expansion. *Nat. Struct. Mol. Biol.* 27, 1185–1193. doi: 10.1038/s41594-020-00518-w

McLoughlin, F., Augustine, R. C., Marshall, R. S., Li, F., Kirkpatrick, L. D., Otegui, M. S., et al. (2018). Maize multi-omics reveal roles for autophagic recycling in proteome remodelling and lipid turnover. *Nat. Plants* 4, 1056–1070. doi: 10.1038/s41477-018-0299-2

Bevan, M., Mayer, K., White, O., Eisen, J.A., Preuss, D., Bureau, T., et al. (2001). Sequence and analysis of the arabidopsis genome. *Curr. Opin. Plant Biol.* 4, 105–110. doi: 10.1016/S1369-5266(00)00144-8

Mizushima, N., Noda, T., Yoshimori, T., Tanaka, Y., Ishii, T. M., George, D., et al, et al. (1998). A protein conjugation system essential for autophagy. *Nature* 395, 395–398. doi: 10.1038/26506

Nettleton, D., Hwang, J. T. G., Caldo, R. A., and Wise, R. P. (2006). Estimating the number of true null hypotheses from a histogram of p values. *J. Agricultural Biological Environ. Stat* 11, 337. doi: 10.1198/108571106X129135

Orii, M., Tsuji, T., Ogasawara, Y., and Fujimoto, T. (2021). Transmembrane phospholipid translocation mediated by Atg9 is involved in autophagosome formation. *J. Cell Biol.* 220. doi: 10.1083/jcb.202009194

Parry, G., Provart, N. J., Brady, S. M., Uzilday, B., and Committee, T. M. A. S. (2020). Current status of the multinational arabidopsis community. *Plant Direct* 4, e00248. doi: 10.1002/pld3.248

Peters, C., Li, M., Narasimhan, R., Roth, M., Welti, R., Wang, X., et al. (2010). Nonspecific phospholipase c NPC4 promotes responses to abscisic acid and tolerance to hyperosmotic stress in arabidopsis. *Plant Cell* 22, 2642–2659. doi: 10.1105/tpc.109.071720

Reiser, L., Subramaniam, S., Li, D., and Huala, E. (2017). "Using the arabidopsis information resource (TAIR) to find information about arabidopsis genes," in *Current protocols in bioinformatics*, vol. 60 . Ed. A. D. Baxevanis, et al, 1 11 11–11 11 45.

Schütter, M., Giavalisco, P., Brodesser, S., and Graef, M. (2020). Local fatty acid channeling into phospholipid synthesis drives phagophore expansion during autophagy. *Cell* 180, 135–149.e114. doi: 10.1016/j.cell.2019.12.005

Shin, K. D., Lee, H. N., and Chung, T. (2014). A revised assay for monitoring autophagic flux in arabidopsis thaliana reveals involvement of AUTOPHAGY-RELATED9 in autophagy. *Mol. Cells* 37, 399–405. doi: 10.14348/molcells.2014.0042

Shiva, S., Enninful, R. M., Roth, R., Tamura, P., Jagadish, K., Welti, R., et al, et al. (2018). An efficient modified method for plant leaf lipid extraction results in improved recovery of phosphatidic acid. *Plant Methods* 14, 14. doi: 10.1186/s13007-018-0282-y

Soto-Burgos, J., Zhuang, X., Jiang, L., and Bassham, D. C. (2018). Dynamics of autophagosome formation. *Plant Physiol.* 176, 219–229. doi: 10.1104/pp.17.01236

Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., et al. (2016). Metabolomics workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 44 (D1), D463-70. doi: 10.1093/nar/gkv1042

Sullivan, A., Purohit, P. K., Freese, N. H., Pasha, A., Esteban, E., Waese, J., et al. (2019). An 'eFP-seq browser' for visualizing and exploring RNA sequencing data. *Plant J.* 100, 641–654. doi: 10.1111/tpj.14468

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One* 6, e21800. doi: 10.1371/journal.pone.0021800

Rhee, S. Y., Beavis, W. T., Berardini, Z., Chen, G., Dixon, D., Doyle, A., et al. (2003). The arabidopsis information resource (TAIR): a model organism database providing a

centralized, curated gateway to arabidopsis biology, research materials and community. *Nucleic Acids Res.* 31, 224–228. doi: 10.1093/nar/gkg076

Thompson, A. R., Doelling, J. H., Suttangkakul, A., and Vierstra, R. D. (2005). Autophagic nutrient recycling in arabidopsis directed by the ATG8 and ATG12 conjugation pathways. *Plant Physiol.* 138, 2097–2110. doi: 10.1104/pp.105.060673

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., , M., Baren van, J., et al, et al. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Tsukada, M., and Ohsumi, Y. (1993). Isolation and characterization of autophagy-defective mutants of saccharomyces cerevisiae. *FEBS Lett.* 333, 169–174. doi: 10.1016/0014-5793(93)80398-E

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A. M., Hein, Y., Geiger, T., et al. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 13, 731–740. doi: 10.1038/nmeth.3901

Xiao, S., Gao, W., Chen, Q. F., Chan, S. W., Zheng, S. X., Ma, J., et al. (2010). Overexpression of arabidopsis acyl-CoA binding protein ACBP3 promotes starvation-induced and age-dependent leaf senescence. *Plant Cell* 22, 1463–1482. doi: 10.1105/tpc.110.075333

Yin, Z., Pascual, C., and Klionsky, D. J. (2016). Autophagy: machinery and regulation. *Microbial Cell* 3, 588–596. doi: 10.15698/mic2016.12.546

Yu, L., Chen, Y., and Tooze, S. A. (2018). Autophagy pathway: Cellular and molecular mechanisms. *Autophagy* 14, 207–215. doi: 10.1080/15548627.2017.1378838

Zhang, S., Hama, Y., and Mizushima, N. (2021). The evolution of autophagy proteins - diversification in eukaryotes and potential ancestors in prokaryotes. *J. Cell Sci.* 134. doi: 10.1242/jcs.233742

Zhuang, X., Chung, K. P., Cui, Y., Lin, W., Gao, C., Kang, B. H., et al. (2017). ATG9 regulates autophagosome progression from the endoplasmic reticulum in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 114, E426–e435. doi: 10.1073/pnas.1616299114

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 March 2023
DOI 10.3389/fpls.2023.1155258



OPEN ACCESS

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

REVIEWED BY
Jie Qiu,
Shanghai Normal University, China
Haihui Fu,
Jiangxi Agricultural University, China
Longjiang Fan,
Zhejiang University, China

*CORRESPONDENCE
Lifeng Wang
ifwang@hunaas.cn
Lianyang Bai
lybai@hunaas.cn
Tao Lin
lintao35@cau.edu.cn

†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 31 January 2023
ACCEPTED 13 March 2023
PUBLISHED 24 March 2023

CITATION
Chen K, Su X, Yang H, Peng Y, Wu L, Zhao Z, Lin T, Bai L and Wang L (2023) Multi-omics analyses reveal the crosstalk between the circadian clock and the response to herbicide application in *Oryza sativa*. *Front. Plant Sci.* 14:1155258. doi: 10.3389/fpls.2023.1155258



# Multi-omics analyses reveal the crosstalk between the circadian clock and the response to herbicide application in *Oryza sativa*

Ke Chen[1,2,3†], Xiao Su[4†], Haona Yang[1,3], Yajun Peng[1,3], Lamei Wu[1,3], Zhenghong Zhao[1,2,3], Tao Lin[4*], Lianyang Bai[1,2,3*] and Lifeng Wang[1,2,3*]

[1]Longping Branch, College of Biology, Hunan University, Changsha, China, [2]Key Laboratory of Indica Rice Genetics and Breeding in the Middle and Lower Reaches of Yangtze River Valley, Ministry of Agriculture and Rural Affairs, Hunan Rice Research Institute, Hunan Academy of Agricultural Sciences, Changsha, China, [3]Huangpu Research Institute of Longping Agricultural Science and Technology, Guangzhou, China, [4]State Key Laboratory of Agrobiotechnology, Beijing Key Laboratory of Growth and Developmental Regulation for Protected Vegetable Crops, College of Horticulture, China Agricultural University, Beijing, China

Plants have evolved circadian clock systems that enable biological processes to occur in tandem with periodic changes in the environment. However, it is largely unknown whether crosstalk occurs between the circadian clock and the response to herbicide in rice. We identified 19 conserved rhythmic metabolites which were response to pesticide application and their metabolic abundance peaked mainly at ZT2 or ZT14-ZT18. We found a series of glyphosate, s-Metolachlor, fenclorim, metcamifen and GA3 response genes were expressed following stable circadian rhythms. In order to determine the patterns of their temporal expression, co-expression network analysis was done on 10,467 genes that were periodically expressed throughout a 24-hour period. Next, we identified 4,031 potential direct target genes of *OsCCA1* in using DAP-seq data for *OsCCA1*. Of these, 339, 22, 53, 53 and 63 genes showed a response to glyphosate, s-Metolachlor, fenclorim, metcamifen and GA3 application, respectively. And they were mainly phased from dusk to midnight. Interestingly, we identified significant *OsCCA1* binding peaks in the promoter regions of four herbicide resistance genes, including *OsCYP81A12*, *OsCYP81E22*, *OsCYP76C2*, and *OsCYP76C4*. Finally, we found that herbicide application could affects the expression of some of the central oscillator genes of the rice circadian clock. Here, we used multi-omics data to reveal the crosstalk between the circadian clock and herbicide response processes at the epigenomics, transcriptome, and metabolome levels in rice. This work will serve as a theoretical guide for identifying rhythmic herbicide targets, leading to the creation of new herbicides or the breeding of crops resistant to herbicides.

KEYWORDS

***Oryza sativa L.*, circadian clock, herbicides, RNA-seq, metabolome**

# 1 Introduction

Rice (*Oryza sativa L.*) is one of the most important crops and is a valuable food source for approximately 4 billion people worldwide (Sasaki and Burr, 2000). The use of pesticides has been increased exponentially in recent decades due to their excellent protective power (Maggi et al., 2019). Pesticides are biologically active xenobiotic compounds that have at least one specific target site and the inhibition of target site activity is the main mode of action of pesticides. Although the application of pesticides is based on an evaluation of their visible phytotoxicity to non-target crops, there are likely to be many non-visual and subtle effects on the crops at the physiological, biochemical, and molecular levels (Banerjee et al., 2001; Bhatnagar-Mathur et al., 2008; Cheng et al., 2012; Ding et al., 2014).

To adapt to the periodic changes in light, temperature, and other environmental factors resulting from the rotation of the Earth, organisms have evolved complex circadian clock systems that enable their physiological processes to occur in accordance with periodic changes in environmental factors (Edgar et al., 2012). The circadian clock not only facilitates the adjustment of circadian oscillations in response to the light/dark cycle and maintenance of regional adaptability in plants but also participates in regulating various biotic/abiotic stress responses. In *Arabidopsis thaliana*, CIRCADIAN CLOCK ASSOCIATED 1 (*CCA1*), the central circadian oscillator, confers heterosis for defense against bacteria in hybrids without growth vigor penalties, and significantly enhances the growth heterosis of hybrids infected with pathogens (Yang et al., 2021). Mutations in the central clock component LUX ARRHYTHMO (*LUX*) disrupt jasmonic acid- and salicylic acid-mediated defense signaling, which results in compromised disease resistance (Zhang et al., 2019). In addition, Arabidopsis GIGANTEA (*GI*) affects adaptability to salt stress, through the Salt Overly Sensitive 2 (*SOS2*)-mediated SOS-signaling pathway (Kim et al., 2013). In rice, the central circadian oscillator component *OsCCA1* regulates tolerance to salt, drought, and osmotic stresses, by mediating the ABA signaling pathway (Wei et al., 2022). *OsPRR37* endows rice with tolerance to salt stress *via* the inhibition of transcription of high-affinity K+ transporters 2;1 (*OsHKT2;1*), thereby reducing the level of accumulated Na+ and reactive oxygen species (ROS) (Wei et al., 2021). Previous studies have shown that the application of herbicides at different times of the day results in different levels of efficiency in field, with the time of day being the second-most important factor (Martinson et al., 2002; Miller et al., 2003). Recently, it has also been shown that overexpression of the *Arabidopsis thaliana* central circadian oscillator *CCA1* and *TOC1* would affect the rhythmic effectiveness of glyphosate (Belbin et al., 2019). And several studies have been conducted to reveal the response of plants to pesticides and the circadian expression of plant endogenous genes and metabolites at the transcript and metabolite levels by sequencing. However, it remains largely unknown whether there is crosstalk between the circadian clock and the response to herbicide application at the transcriptional, metabolic, and other levels.

In this study, we first found 19 conserved rhythmic metabolites in response to pesticides application. To understand the molecular mechanisms of the crosstalk, we identified conserved and extremely conserved circadian genes in rice and analyzed the temporal expression patterns of these genes. In addition, we identified a series of conserved circadian genes could response to herbicides such as glyphosate, s-metolachlor and herbicide safeners such as fenclorim, metcamifen and GA3, and analyzed their expression patterns and functions. Further, we performed a co-expression network analysis to identify genes that were specifically highly expressed at each time. Next, we used the DAP-seq data of the rice central circadian oscillator *OsCCA1* to identify herbicide-responsive genes and herbicide resistance genes that may be directly regulated by *OsCCA1*. We found that some known herbicide non-target-site resistance genes might be directly regulated by *OsCCA1*. Further, we identified the effects of herbicide application on central circadian oscillators in rice. In this study, we used multi-omics data at the epigenomics, transcriptome, and metabolome levels to uncover the interaction between the circadian clock and the response to herbicide treatment in rice (Figure 1). This work will serve as a theoretical guide for identifying rhythmic herbicide targets, leading to the creation of new herbicides or the breeding of crops resistant to herbicides.

# 2 Results

## 2.1 A series of pesticide-responsive metabolites are rhythmic

Previous study found the application of herbicide butachlor, insecticide chlorpyrifos, and fungicide tricyclazole affect the accumulation of many metabolites in rice (Liu and Zhu, 2020). And we found the application of the herbicide s-metolachlor and herbicide safener, also a plant hormone, gibberellin 3 (GA3) also showed an effect to rice (Table S1). In addition, a study found the accumulation of many metabolites in rice with stable circadian rhythms (known as circadian rhythmic metabolites - CRMs) has been observed (Zhou et al., 2022). However, whether these metabolites with circadian rhythms are induced by pesticides is still known, and much is currently unknown about the crosstalk between circadian rhythms and herbicide response in rice. Therefore, we used the metabolite data from the two studies mentioned above (Table S2) as well as metabolomic data measured by our team and identified 19 CRMs responded to pesticides application. We found that amino acids were the most frequent rhythmic metabolite clade to respond to pesticides, followed by carbohydrates and organic acids (Figure 2A).

We then analyzed the phase distribution of these 19 CRMs and found that these metabolites reached peak abundance both during the day and night. Daily-phased metabolites reached peak metabolic abundance mainly at dawn (Zeitgeber time 2 (ZT2, zeitgeber time relative to this experimental LD cycle, and in an LD cycle of 12h of light and 12h of darkness (LD 12:12) the time of lights on is denoted by ZT0, and the time of lights off by ZT12)), while night-phased metabolites reached peak metabolic abundance mainly before midnight (ZT14-ZT18) (Figure 2B).

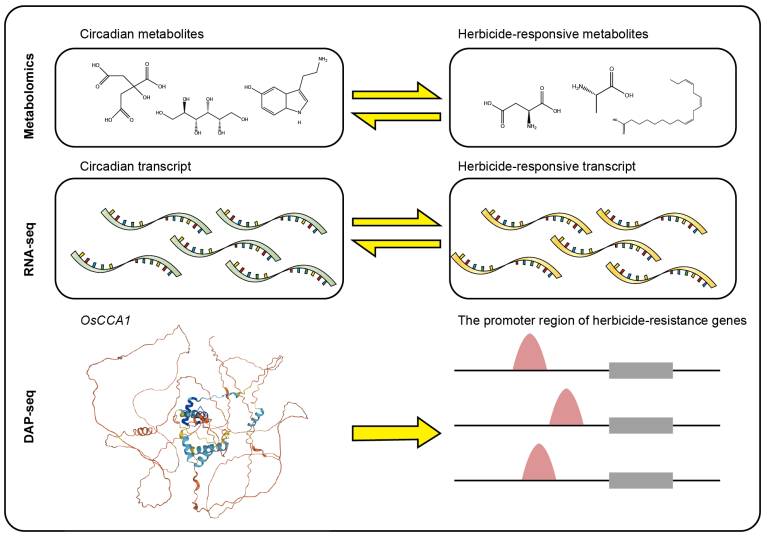



FIGURE 1
Schematic overview of the workflow and analysis pipeline in this study.

Among the daily metabolites, the Serotonin, Sorbitol, Citric Acid and L-Aspartic acid were peaking in the morning (ZT2) and the 6-hydroxynicotinic acid, alpha.-Linolenic acid, L-Alanine and Cytosine were peaking in the afternoon (Figure 2C). And only Malic acid and L-Threonine reached peak expression after the midnight, while the rest peaked from dusk to midnight (Figure 2C). Importantly, 16 (84%) of the CRMs responded to the application of herbicide, which raises the possibility that the circadian clock is

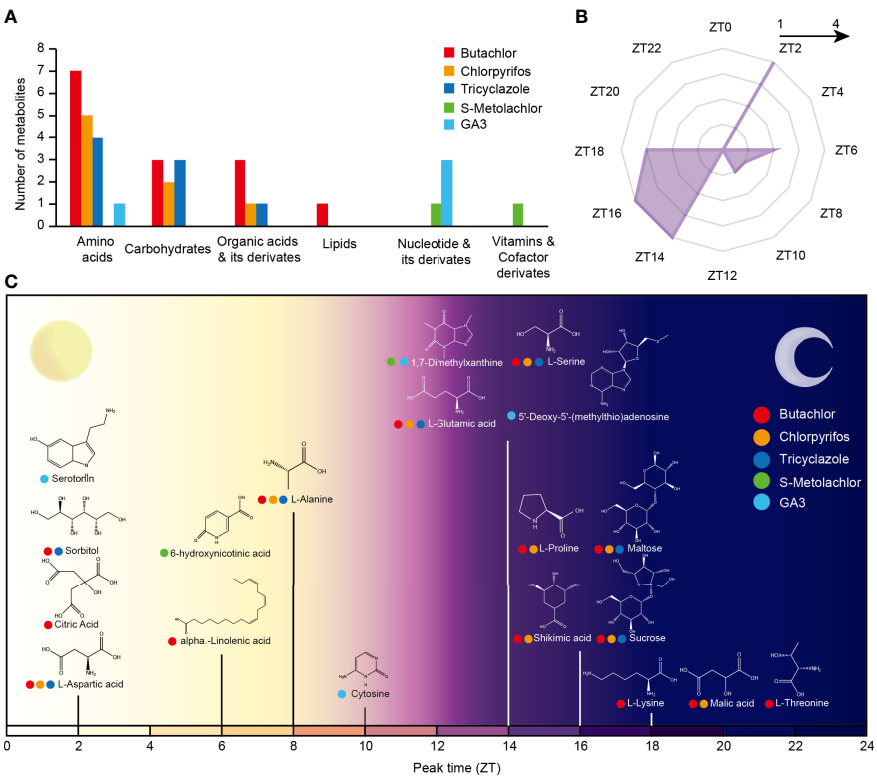



FIGURE 2
The crosstalk of rice circadian rhythm and pesticides response processes at metabolome level. **(A)** Number of various CRMs in response to the application of different pesticides. **(B)** Circadian phase distribution of pesticides responsive CRMs. The coordinates of each angle represent phases ZT0–ZT24. The inner to outer circles represent 1-5 metabolites reaching peak expression at that time point, respectively. **(C)** Time of peak accumulation of pesticides responsive CRMs. The x-axis represents the circadian phase of the metabolite. The different colors of the circles on the left of the metabolite's name represent its response to different pesticides.

directly involved in a number of the metabolic pathways by which rice reacts to the application of herbicide.

## 2.2 A series of herbicide-responsive genes are rhythmic

To investigate the crosstalk between the circadian clock and herbicide application in rice, we identified circadian rhythm genes in three rice varieties in using the previous published RNA-Sequencing data (Table S2) (Li et al., 2021), and defined genes expressed by 24 hours period in at least two varieties as circadian genes (CGs) and genes expressed by 24 hours period in all three rice varieties as conserved circadian genes (CCGs). Totally, we identified 9,974 CGs and 4,560 CCGs (Figure 3A), and found that most CGs and CCGs were phased from dusk to midnight (ZT12-ZT15) (Figures 3B, C). Next, we performed gene ontology (GO) enrichment analysis of CGs (Figure S1) and CCGs (Figure S2), found that both were enriched in genes related to nitrogen synthesis and metabolism as well as small molecule metabolism. This indicated that rhythmic genes, which are conserved in many rice varieties, are mainly involved in nutrient uptake metabolism and plant growth metabolic processes.

Next, we identified circadian rhythmic genes that were influenced by herbicides and herbicide safeners application including glyphosate (Table S2) (Zhai et al., 2020), s-metolachlor, fenclorim (Table S2) (Hu et al., 2020), metcamifen (Table S2) and GA3, and found 1,069 glyphosate-responsive genes (Table S3), 51 s-metolachlor-responsive genes (Table S4), 182 fenclorim-responsive genes (Table S5), 122 metcamifen-responsive genes (Table S6) and 183 GA3-responsive genes that exhibited oscillations under light-dark cycles (Table S7). In other words, we found that 16.39%-29.82% of herbicide-responsive genes had stable circadian rhythms, with the highest percentage of s-metolachlor-induced genes and the lowest percentage of metcamifen-suppressed genes (Figures 3D, F).

For the rhythmic herbicide-induced genes, we found that the phases of s-metolachlor-induced genes showed a scattered distribution throughout the day, while those of 61.86% glyphosate-induced rhythmic genes reached peak abundance levels from dusk to midnight (ZT12-ZT16.5). As for the rhythmic herbicide-suppressed genes, we found that s-metolachlor-suppressed genes were mainly expressed at night, while glyphosate-suppressed genes had the most genes with peak expression around ZT3 and ZT12, respectively (Figure 3E). For the rhythmic herbicide safener-induced genes, we found that the phase of fenclorim-induced genes and GA3-induced genes showed a scattered distribution pattern throughout the day, the metcamifen-induced genes were mainly concentrated their expression peak at ZT3 and all night. And for the rhythmic herbicide safener-suppressed genes, we found 69.73% of fenclorim-suppressed rhythmic genes reached peak abundance from dusk to midnight (ZT12-ZT16.5), the GA3 and metcamifen-suppressed rhythmic genes more peaking in the ZT18-ZT22.5 and ZT16.5-ZT18, respectively (Figure 3G).

To investigate the functions possessed by rhythmic herbicide response genes, we performed GO functional enrichment analysis of rhythmic herbicide-induced or suppressed genes. For herbicides application, we found that rhythmic glyphosate-induced genes are mainly involved in glucan biosynthesis, metabolism and other processes; rhythmic S-Metolachlor-induced genes are mainly involved in small molecule catalytic processes while suppressed genes are mainly involved in carbohydrate biosynthesis processes. For herbicide safeners application, we noticed the rhythmic Fenclorim-induced genes mainly involved in stress response processes, while the suppressed genes are mainly involved in the carbohydrate synthesis and metabolism of carbohydrate such as fatty acids; the rhythmic Metcamifen-induced genes are mainly involved in some transmembrane transport processes while the suppressed genes are also involved in carbohydrate synthesis and metabolism; the GA3-induced genes are mainly involved in some acid catalytic processes while the suppressed genes are mainly involved in some cellular organization processes. This suggests that for different herbicides, the effect of circadian rhythms on their response processes may be different (Figure 3H).

## 2.3 Temporal expression pattern of rhythmic herbicide-responsive genes

To further understand the molecular mechanisms between gene expression patterns and the herbicide response, we performed a weighted gene co-expression network analysis (WGCNA) using 10,467 genes expressed over a 24-period on 48 h time-scale samples. We found that samples obtained during the day could be classified into two clusters, based on whether they were obtained in the morning or afternoon, while samples obtained at night were all clustered together (Figure 4A). These genes were classified into 7 co-expression modules *via* WGCNA (Figure 4B) and the turquoise and black modules had the most (4,480) and least (59) number of genes, respectively. The blue, brown, green, red, and yellow modules had 2,557, 1,777, 488, 170, and 862 genes, respectively (Table S8).

Next, we found that the highest percentages of both herbicide-responsive genes and herbicide safener-responsive genes could be observed in the module black (Figure 4C) and peaked in abundance at ZT18 (Figure S3). Interestingly, it was almost always the induced genes that were in this module except S-Metolachlor application. Besides the module black, the second-most percentage of glyphosate-induced genes occurred in the module brown, and their peak abundance was reached before midnight (ZT15-ZT18 and ZT39-ZT42) (Figure 4C). However, the highest percentage of glyphosate-suppressed genes occurred in the module green and red (Figure 4C), and their peak abundance was reached in the morning ((ZT3-ZT6, ZT27-ZT30) (module green) and ZT30 (module red)) (Figure S3). Another herbicide, S-Metolachlor, which induced genes were most abundant in module red and the most abundant suppressed genes were almost same in module green, red and turquoise in addition to the black module (Figure 4C). For herbicide safener, the second-most percentage of fenclorim-induced genes were the module yellow (Figure 4C), which reached peak abundance at dawn (ZT0, ZT3, ZT24, ZT27 and ZT48) (Figure S3) and the module red had the highest percentage of fenclorim-suppressed genes, and these findings were similar to

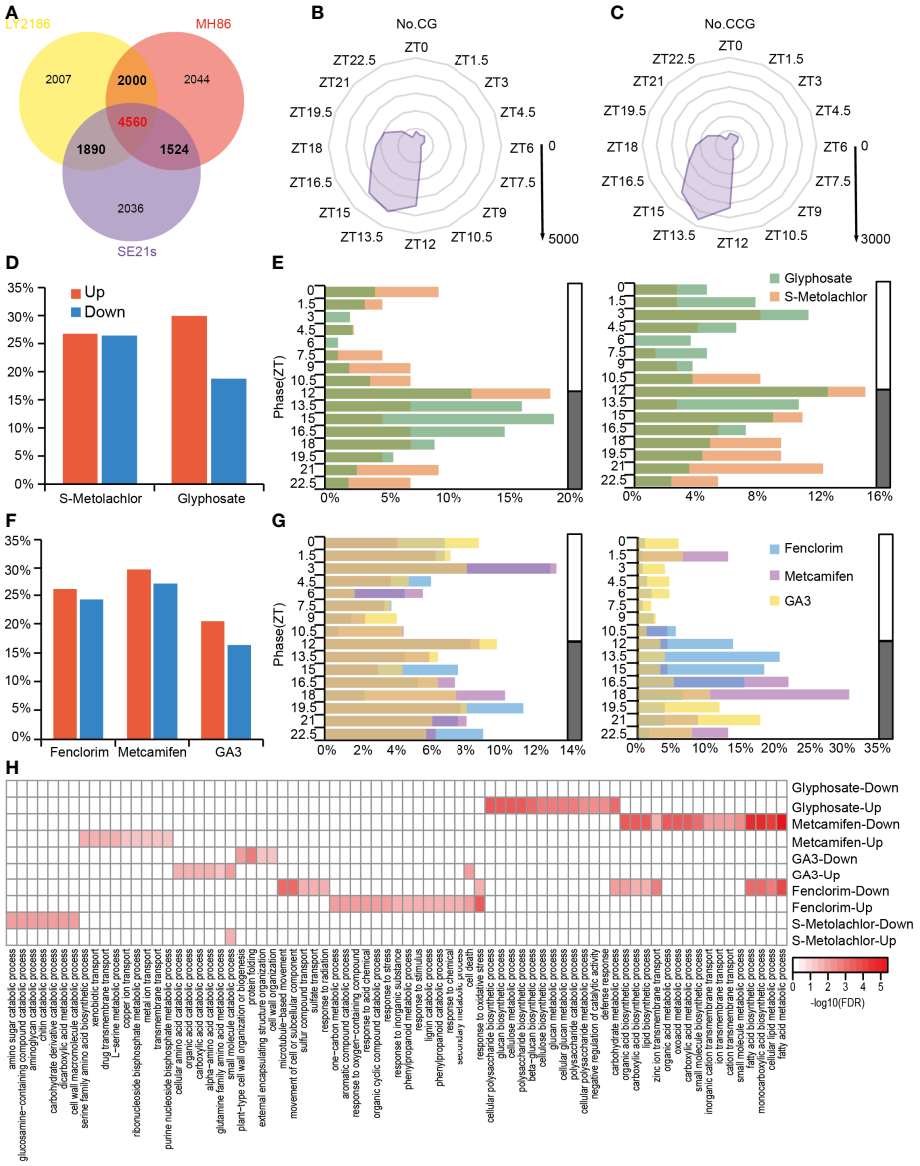



FIGURE 3
The crosstalk of rice circadian rhythm and herbicide response processes at transcriptome level. **(A)** Venn diagram indicating CGs, CCGs and specific genes is shown. Genes with a 24-hour stable circadian rhythm in at least two rice varieties are defined as CG (indicated by bold font); genes with a 24-hour stable circadian rhythm in all three rice varieties are defined as CCG (indicated by red bold font). **(B)** Phase distribution of the CGs. The coordinates of each angle represent phases ZT0–ZT24. The inner to outer circles represent 1,000-5,000 genes reaching peak expression at that time point, respectively. **(C)** Phase distribution of the CCGs. The coordinates of each angle represent phases ZT0–ZT24. The inner to outer circles represent 500-3,000 genes reaching peak expression at that time point, respectively. **(D)** The percentage of rhythmic herbicide-responsive genes in the total herbicide-responsive genes. **(E)** Phase distribution of rhythmic herbicide-responsive genes. The left panel is rhythmic herbicide-induced genes and right panel is rhythmic herbicide-suppressed genes. **(F)** The percentage of rhythmic herbicide safener-responsive genes in the total herbicide safener-responsive genes. **(G)** Phase distribution of rhythmic herbicide safener-responsive genes. Left panel is rhythmic herbicide safener-induced genes and the right panel is rhythmic herbicide safener-suppressed genes. **(H)** Heatmap of GO enrichment analysis of rhythmic herbicide-responsive genes.

those for glyphosate-suppressed genes (Figure 4C), which reached their peak abundance at ZT30 (Figure S3). In addition to the module black, the percentage of metcamifen-induced genes and GA3-induced genes were second highest in the module red. And ga3-surppressed genes were second highest in the module Green (Figure 4C).

Interestingly, we found that the module yellow (phased in ZT0-3; ZT24-27), module red (phased in 6-9; ZT30), and module brown (phased in ZT15-18; ZT 39-42) contained specific genes that were highly expressed in the morning, afternoon, and evening, respectively (Figure S3). In the module yellow, 133, 4, 18, 27 and 28 genes responded to glyphosate, S-Metolachlor, fenclorim, metcamifen and GA3 application, respectively. In addition, we identified rice central oscillator gene *OsCCA1* and 48 transcription factors in this module. Among the transcription factor families, bZIP, GRAS, MYB, bHLH, ERF, and Nin-like families were the most abundant (larger than 3) (Figure 4D). In the module red, 28, 2, 5, 5 and 7 genes showed a response to

10.3389/fpls.2023.1155258

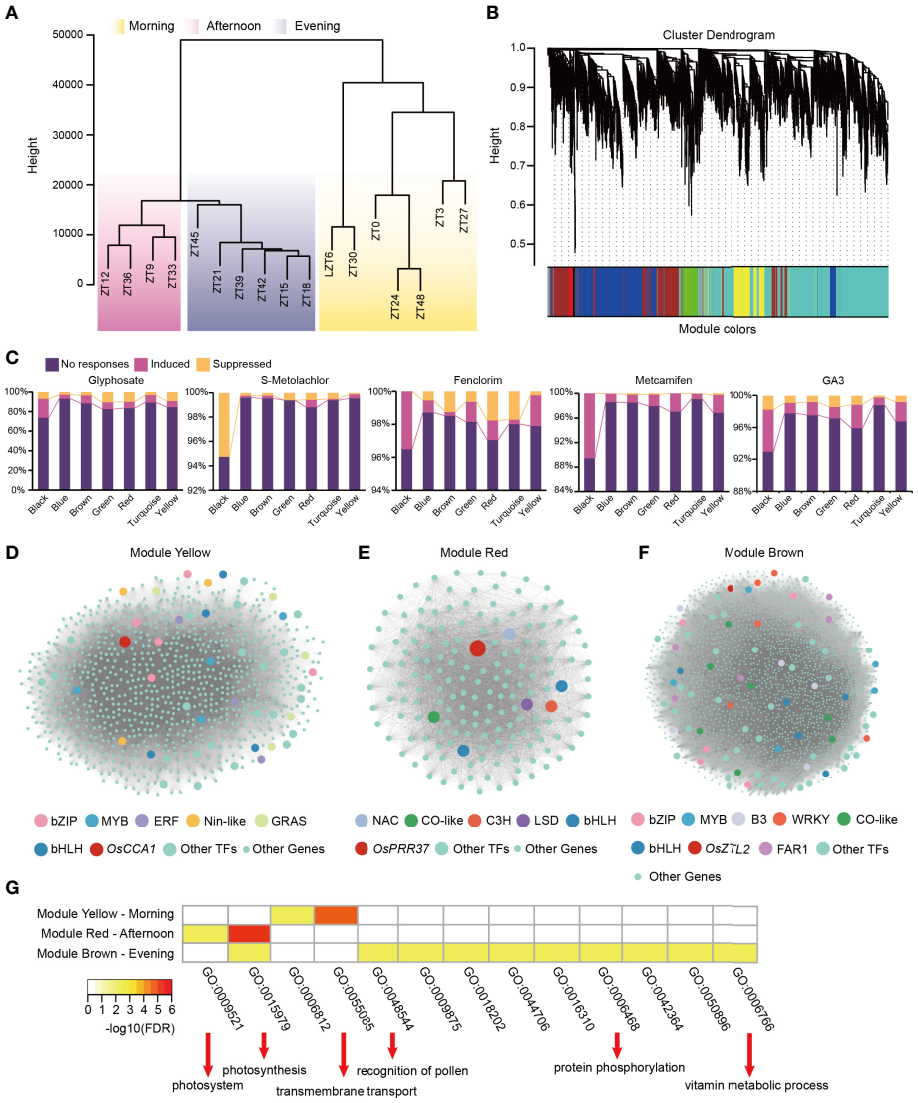



FIGURE 4
Co-expression analyses of 24h periods circadian rhythm genes. **(A)** Hierarchical clustering dendrogram showing samples grouping into morning, afternoon and evening clusters. **(B)** Dendrogram showing transcript co-expression modules (clusters) identified by WGCNA across different times. The major tree branches constitute 7 modules labeled with different colors. **(C)** Proportion of genes in each module in response to herbicide application. **(D–F)** Co-expression network of module yellow, module red and module brown, respectively. The types of each transcription factor family are shown in the figure legend. **(G)** Heatmap of GO enrichment analysis of genes in module yellow (upper), red (medium) and brown (bottom), respectively.

glyphosate, S-Metolachlor, fenclorim, metcamifen and GA3 application, respectively. In addition, we identified rice central oscillator gene *OsPRR37* and 7 transcription factors, including those from the bHLH, C3H, CO-like, LSD, and NAC transcription factor families (Figure 4E). In the module brown, 207, 8, 26, 26 and 44 genes responded to glyphosate, S-Metolachlor, fenclorim, metcamifen and GA3 application, respectively. In addition, we identified rice central oscillator genes, such as *OsZTL2* and *OsLWD2*, and 80 transcription factors in this module. Among the transcription factor families, the CO-like, B3, bHLH, bZIP, FAR1, MYB, and WRKY families occurred most abundantly (larger than 5) (Figure 4F). Further, GO enrichment analysis of genes in the three modules revealed that the genes in the morning-phased module (the module yellow) are mainly involved in transmembrane transport, the genes in the afternoon-phased module (the module red) are mainly involved in photosynthesis-related processes, while the genes in the evening-phased module (the module brown) are involved in recognition of pollen, protein phosphorylation and vitamin metabolic process (Figure 4G).

## 2.4 *OsCCA1* directly regulates a range of herbicide-responsive genes and certain herbicide-resistance genes

*OsCCA1* is a MYB transcription factor and one of the central oscillator genes in the rice circadian clock. It has been shown that *OsCCA1* could regulate several biological processes associated with

rice growth and development (Wang et al., 2020; Sun et al., 2021; Wei et al., 2022). In addition, it has been previously shown that the overexpression of *CCA1* in *Arabidopsis* leads to a disruption in the rhythmic effectiveness of glyphosate application (Belbin et al., 2019). To explore whether *OsCCA1* directly affects the response to herbicides and herbicide safeners, we used DNA affinity purification sequencing (DAP-seq) of *OsCCA1* (Wei et al., 2022). Among the identified binding peaks, 4,078 were located within regions that were 1 kb upstream of the ATG codon and contained 4,031 downstream genes that may be directly regulated.

Among the genes that were possibly regulated by *OsCCA1*, we identified 339 (8.40% in glyphosate-responsive genes), 22 (9.24% in S-Metolachlor-responsive genes), 53 (7.08% in fenclorim-responsive genes), 53 (8.54% in metcamifen-responsive genes) and 63 (9.92% in ga3-responsive genes) genes that showed a response to glyphosate (OsCCA1-Gly), S-Metolachlor (OsCCA1-SMet), fenclorim (OsCCA1-Fen), metcamifen (OsCCA1-Met) and GA3 (OsCCA1-ga3) application, respectively (Figures 5A, B). Of these, 63 genes responded to at least two herbicides (Figure 5A). However, only 7% (4,031/57,585) of genes at the genome-wide level may be directly regulated by *OsCCA1*, suggesting that *OsCCA1* may be involved in the response process to herbicides except fenclorim.

Next, we analyzed the phase distribution of these genes and found that most *OsCCA1*-regulated genes associated with a response to herbicides reached peak abundance levels in the evening. Among these, OsCCA1-Gly, OsCCA1-Fen and OsCCA1-ga3 mainly reached peak expression levels at ZT13.5, OsCCA1-Met mainly reached peak expression levels at ZT16.5 and OsCCA1-SMet mainly enriched in ZT12 (Figure 5C). The most herbicide-responsive genes regulated by *OsCCA1* all reached peak expression mainly at night, suggesting that *OsCCA1*-regulated processes associated with responsiveness to herbicides may play a greater role at night.

Notably, we identified four herbicide resistance genes may be directly regulated by *OsCCA1*, including *OsCYP81A12*, which homologous in *Echinochloa phyllopogon* could confer bensulfuron-methyl and penoxsulam resistance (Iwakami et al., 2014), *OsCYP81E22* which homologs in soybean to confer bentazon resistance (Kato et al., 2020), *OsCYP76C2* and *OsCYP76C4* which homologs in *Arabidopsis thaliana* to confer Monoterpenols and Phenylurea resistance (Hofer et al., 2014) (Figure 5A). We identified significant binding peaks at 295 bp, 863 bp and 549 bp before the start codon of *OsCYP81A12* (Figure S4), *OsCYP76C2* (Figure S5) and *OsCYP76C4* (Figure 5D), respectively. In addition, we identified two significant binding peaks at a location 276 bp upstream of the start codon of *OsCYP81E22* (Figure 5E). Notably, we also identified a significant binding peak located 4,551 bp upstream of the start codon of *OsABCC8*, which facilitated the translocation of glyphosate into the vesicles and protected plants from glyphosate toxicity (Pan et al., 2021) (Figure S6).

Next, we analyzed the response of these herbicide resistance genes, which may be directly regulated by *OsCCA1*, to herbicide and herbicide safener applications. We found that *OsCYP81A12* ($log_2$foldchange = 2.95, FDR = 0.0067) was significantly induced upon glyphosate application, while *OsCYP76C2* ($log_2$foldchange = −3.07, FDR = 0.045 and *OsCYP76C4* ($log_2$foldchange = −1.15, FDR = 1.69e-14) were significantly suppressed upon glyphosate application. And the application of metcamifen, significantly induced the expression of *OsCYP76C2* ($log_2$foldchange = 5.15, FDR = 2.41e-43) (Table S9).

## 2.5 Effect of herbicide application on rice central oscillator

To investigate whether the application of herbicides disrupted the central oscillator of the circadian clock in rice, we extracted central oscillator genes associated with the circadian clock in rice (Hayama and Coupland, 2004; Song et al., 2010; Filichkin et al., 2011; Greenham and McClung, 2015; Steed et al., 2021) and analyzed their expression phased in LY2186 and the changes in their expression levels before and after herbicide application (Figure 6).

For herbicides application, we found that only *OsHY5* ($log_2$foldchange = −1.16, p-value = 1.05e-11) was significantly suppressed upon glyphosate application and no central oscillator gene significantly responds to S-Metolachlor application. For herbicide safeners application, we found four genes were significantly affected by fenclorim application, with *OsLUX* ($log_2$foldchange = 1.27, p-value = 0.0048) and *OsHY5* ($log_2$foldchange = 1.12, p-value = 0.00022) being significantly induced, and *OsPRR37* ($log_2$foldchange = −1.04, p-value = 0.0014) and *OsGI* ($log_2$foldchange = −1.22, p-value = 4.09e-06) being significantly suppressed. And *OsPRR95* was significantly induced by GA3 application. However, no central oscillator genes were significantly affected by the application of metcamifen (Figure 6). This suggests that central oscillator genes respond differently to herbicides.

# 3 Discussion

Previous studies have demonstrated that the effectiveness of herbicides may vary depending on the time of day they are applied. Although numerous exogenous environmental factors in the field could influence this process, time remains the second-most influential factor after ambient temperature (Martinson et al., 2002; Miller et al., 2003). It has also been recently shown that the overexpression of the *Arabidopsis thaliana* central oscillator genes *CCA1* and *TOC1* would disrupt the rhythmic effectiveness of glyphosate (Belbin et al., 2019). Nevertheless, the crosstalk between the circadian clock and the response to herbicides remains largely unknown. A previous study found the oscillation phases of rice metabolites were enriched in the afternoon and late-night (Zhou et al., 2022). Which differed from those reported on the relatively continuous peak phase of rhythmic metabolites in animals (Krishnaiah et al., 2017). Interestingly, we found that the accumulation of 19 CRMs, including eight amino acids, three

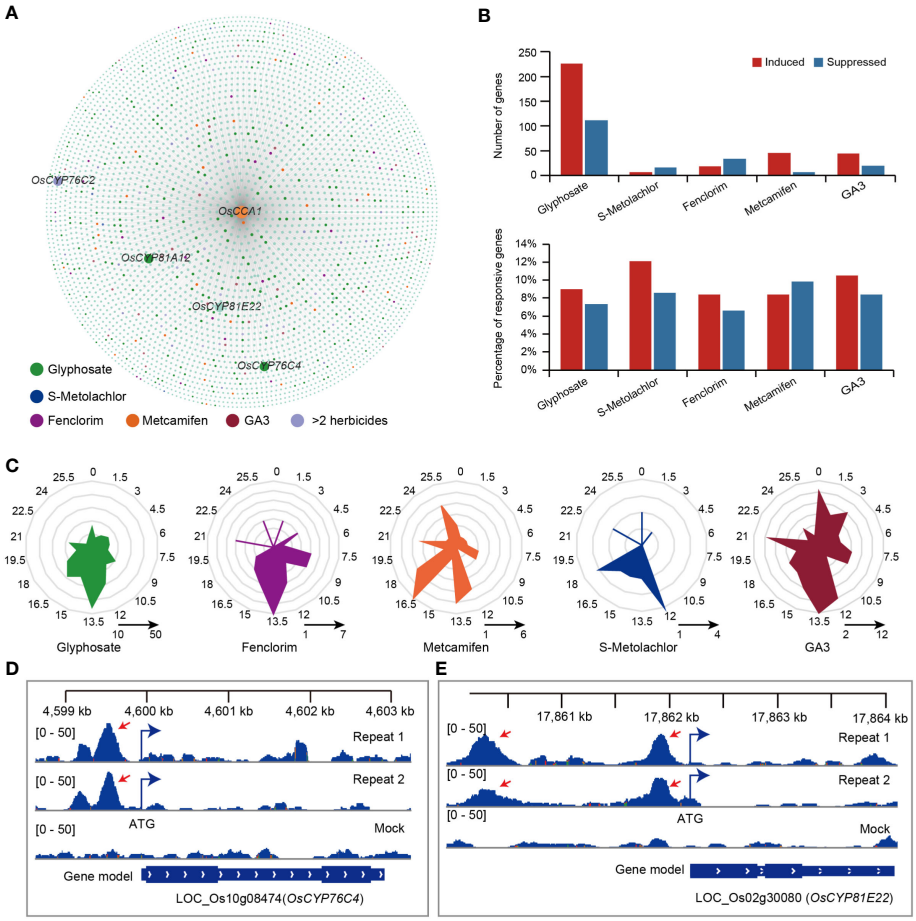



FIGURE 5
The crosstalk of *OsCCA1* target genes and herbicide-responsive genes. **(A)** *OsCCA1* and its target genes (*OsCCA1* has a binding peak within 1000 bp upstream of the start codon of this gene.) association network, the closer to *OsCCA1* gene indicates that the DAP-seq peak is closer to the promoter of the gene. **(B)** The statistics of genes induced or suppressed by herbicide in *OsCCA1* target genes. upper: the number of herbicide-responsive genes in *OsCCA1* target gene sets. bottom: the proportion of *OsCCA1* targeted herbicide-responsive genes in total herbicide-responsive genes. **(C)** Phase distribution of herbicide-responsive *OsCCA1* target genes. **(D)** The binding peaks (repeats 1 and 2) and negative control (mock) of *OsCCA1* in the promoter (−549 bp) of *OsCYP76C4* by DAP-seq. The [0−50] shows the scale bar of binding peak that refers to the height of the peak. **(E)** The binding peaks (repeats 1 and 2) and negative control (mock) of *OsCCA1* in the promoter (−276 bp) of *OsCYP81E22* by DAP-seq. The [0−50] shows the scale bar of binding peak that refers to the height of the peak.

carbohydrates, three Organic acids & its derivates, three Nucleotide & its derivates, a lipid and a Vitamin, could be response to pesticide application (Figure 2A). They reached peak levels of metabolic abundance mainly at ZT2 or ZT14-ZT18 (Figure 2B). Interestingly, most of the pesticide-responsive metabolites could respond to the herbicide (Figure 2C), which suggests that circadian rhythms are likely to influence the response process of rice to herbicides.

Our findings showed that approximately 30% of rice genes exhibit circadian expression, and these findings are similar to those for *Arabidopsis* (Covington et al., 2008). Nearly 20% and 10% of genes in rice were circadian and conserved circadian genes, respectively (Figure 3A). And the function of rhythmic herbicide response genes differed when different herbicides were applied. This suggests that circadian rhythms may differ in response processes to different herbicides. However, this study was only able to explore the effect of genes with circadian rhythms when subjected to herbicide application, but the dataset was unable to explore the differential response to rice produced by herbicide application at different time points, which needs to be explored in future study.

We conducted the hierarchical clustering of temporal transcriptome samples and found that they could be divided into three categories: morning, afternoon, and evening (Figure 4A). This suggests that samples obtained during the day could be divided into two groups, may be based on differences in light intensity, temperature, humidity, and other environmental factors in the morning and afternoon, but all the samples obtained at night were assigned to one group. Among these co-expression modules, we found three modules, each of which contained the central circadian clock oscillator genes *OsCCA1*, *OsPRR37*, and *OsZTL2*, which represented sets of genes that were specifically expressed at high levels in the morning, afternoon, and evening, respectively (Figures 4D–F). In addition, we found that the three modules drive

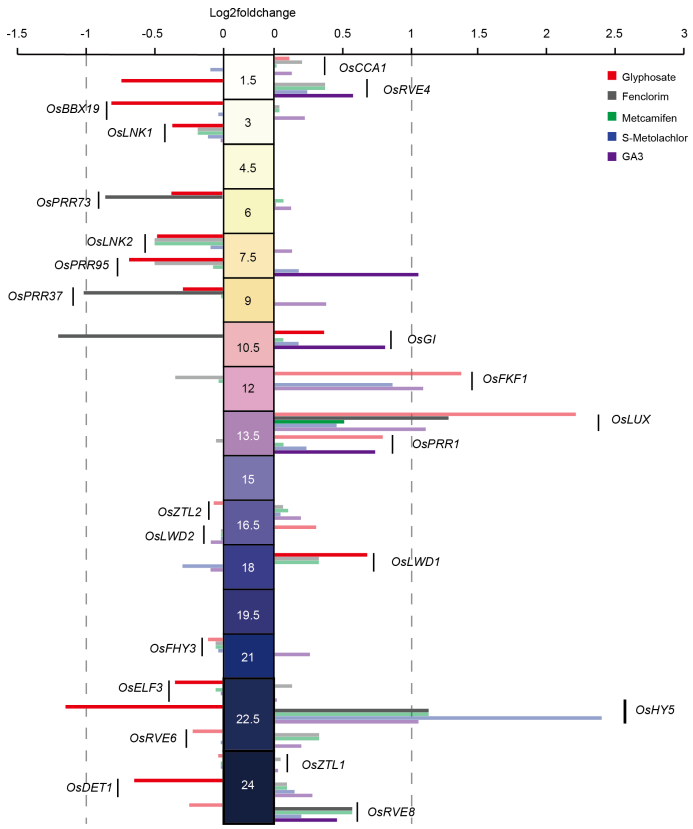



FIGURE 6
Effects of herbicide application on rice central circadian oscillator. The middle values indicate the phase (ZT) of genes. Opaque columns indicate that a significance level has been achieved (FDR<0.05).

different functions. The morning-phased module is mainly involved in the transmembrane transport process, the afternoon-phased module is mainly involved in the photosynthesis-related process, and the evening-phased module is mainly involved in the pollen recognition process, protein phosphorylation process and vitamin metabolic process (Figure 4G).

In *Arabidopsis thaliana*, *CCA1* conferred heterosis for defense against bacteria in hybrids without growth vigor penalties and significantly enhance the growth heterosis of hybrids infected with pathogens (Yang et al., 2021). A previous study has shown that the overexpression of *CCA1* in *Arabidopsis* leads to the disruption of the rhythmic effectiveness of glyphosate application (Belbin et al., 2019). *OsCCA1*, its homolog in rice, is also able to participate in the regulation of many growth and development processes (Wang et al., 2020; Sun et al., 2021; Wei et al., 2022). We have identified 4,031 genes that *OsCCA1* might directly regulate using the DAP-seq data of *OsCCA1*, and identified several genes that exhibited a response to herbicide application (Figure 5A) Notably, we identified significant *OsCCA1* binding peaks in the promoter regions of the four homologous herbicide resistance genes in rice (Figures 5D, E). And a significant binding peak was also identified at a location that was 4,551 bp upstream of the start codon of the rice glyphosate resistance gene *OsABCC8* (Figure S6). This suggests that *OsCCA1* may be directly involved in regulating the response and resistance to herbicides.

# 4 Conclusion

In this study, we have revealed the crosstalk between rice circadian clock and herbicide response processes in epigenomics, transcriptomics and metabolomics level. In addition, four homologs of known herbicide resistance genes in rice including *OsCYP81A12*, *OsCYP76C2*, *OsCYP76C4* and *OsCYP81E22* were found to may be directly regulated by the central oscillator gene *OsCCA1*. Finally, we found that the central oscillator genes respond differently to herbicides. We believe this research will provide a theoretical reference for discovering rhythmic herbicide targets and thus developing new herbicides or breeding herbicide-resistant crops.

# 5 Materials and methods

## 5.1 Transcriptome analysis

The RNA sequencing reads were removed of adapters and trimmed for low-quality bases using fastp (Chen et al., 2018) (version 0.21.0). Clean reads were mapped to the rice genome (MSU-release 7) (Kawahara et al., 2013) using HISAT2 (Kim et al., 2015) (version 2.1.0) with default parameters. The JTK_CYCLE (Hughes et al., 2010) was used to identify circadian genes. The JTK_CYCLE algorithm is an available computational R script from

https://openwetware.org/wiki/HughesLab: JTK_Cycle. We applied the JTK_CYCLE algorithm to estimate the period length, phase, and amplitude for the genes. Only genes with a (P-value < 0.05 and period length = 24) were considered circadian genes.

## 5.2 Co-expression network construction

The Stringtie software (version 1.3.6) (Pertea et al., 2015) was used to calculate the TPM value of genes and the R package WGCNA (Langfelder and Horvath, 2008) were used to identify co-expression modules based on the TPM data.

## 5.3 Analysis of differential expressed genes

The software featurecount (version 1.6.0) (Liao et al., 2014) was used to calculate the raw count of genes and the R package DESeq2 (Love et al., 2014) was used to identify differentially expressed genes. Log2foldchange > 1 or < -1 and adjusted p-values (FDR) < 0.05 were defined as differential expressed genes.

## 5.4 Functional annotation and classification

Gene Ontology (GO, http://www.geneontology.org/) enrichment analysis of the genes was using R package topGO (Alexa and Rahnenfuhrer, 2010).

## 5.5 DAP-seq data analysis

DAP-seq reads were aligned to the reference genome using Bowtie2 (version 2.4.2)(Langmead and Salzberg, 2012). The software MACS2 (version 2.2.7.1) was used to callpeak (Zhang et al., 2008) and IDR (version 2.0.2) software (Li et al., 2011) was used to merge the peaks of the two biological replicates with P<0.05. The bound peaks were annotated using Homer software (Heinz et al., 2010).

## 5.6 Identification of rhythmic agrochemical-responsive metabolites

We used the lists of metabolites that herbicide-induced or suppressed (Liu and Zhu, 2020), the list of circadian rhythmic metabolites (Zhou et al., 2022) and our metabolomic data about s-metolachlor and GA3 application. We found 19 metabolites that not only had stable circadian rhythms and were able to respond to pesticide application. Each of these metabolites was also assigned to bins of 2 h according to its circadian phase.

## 5.7 Identification of the rice central oscillator genes

Sequences mentioned in this article can be download from Rice Genome Annotation Project Database (http://rice.plantbiology.msu.edu/) under the following accession numbers: *OsCCA1*, LOC_Os08g06110; *OsPRR1*, LOC_Os02g40510; *OsPRR95*, LOC_Os09g36220; *OsPRR37*, LOC_Os07g49460; *OsPRR73*, LOC_Os03g17570; *OsLUX*, LOC_Os01g74020; *OsELF3*, LOC_Os06g05060; *OsZTL1*, LOC_Os02g05700; *OsZTL2*, LOC_Os06g47890; *OsFKF1*, LOC_Os11g34460; *OsGI*, LOC_Os01g08700. Other genes were derived from blast with Arabidopsis genes: *OsBBX19*, LOC_Os09g35880; *OsDET1*, LOC_Os01g01484; *OsFHY3*, LOC_Os02g39540; *OsHY5*, LOC_Os02g10860; *OsLWD1*, LOC_Os02g45810; *OsLWD2*, LOC_Os02g32430; *OsLNK1*, LOC_Os03g27019; *OsLNK2*, LOC_Os01g31360; *OsRVE4*, LOC_Os02g45670; *OsRVE6*, LOC_Os06g45840; *OsRVE8*, LOC_Os06g01670.

## 5.8 metabolite extraction

Biological samples are freeze-dried by vacuum freeze-dryer (Scientz-100F). The freeze-dried sample was crushed using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. Dissolve 100 mg of lyophilized powder with 1.2 mL 70% methanol solution, vortex 30 seconds every 30 minutes for 6 times in total, place the sample in a refrigerator at 4°C overnight. Following centrifugation at 12000 rpm for 10 min, the extracts were filtrated (SCAA-104, 0.22 μm pore size; ANPEL, Shanghai, China, http://www.anpel.com.cn/) before UPLC-MS/MS analysis.

## 5.9 UPLC-MS/MS conditions and qualitative

The sample extracts were analyzed using an UPLC-ESI-MS/MS system (UPLC, SHIMADZU Nexera X2, www.shimadzu.com.cn/; MS, Applied Biosystems 4500 Q TRAP, www.appliedbiosystems.com.cn/). The analytical conditions were as follows, UPLC: column, Agilent SB-C18 (1.8 μm, 2.1 mm * 100 mm); The mobile phase was consisted of solvent A, pure water with 0.1% formic acid, and solvent B, acetonitrile with 0.1% formic acid. Sample measurements were performed with a gradient program that employed the starting conditions of 95% A, 5% B. Within 9 min, a linear gradient to 5% A, 95% B was programmed, and a composition of 5% A, 95% B was kept for 1 min. Subsequently, a composition of 95% A, 5.0% B was adjusted within 1.1 min and kept for 2.9 min. The flow velocity was set as 0.35 mL per minute; The column oven was set to 40°C; The injection volume was 4 μL. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (QTRAP)-MS. LIT and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), AB4500 Q TRAP UPLC/MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in positive and negative ion mode and controlled by Analyst 1.6.3 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 550°C; ion spray voltage (IS) 5500 V (positive ion mode)/-4500 V (negative ion mode); ion source gas I (GSI), gas II (GSII), curtain gas (CUR) was set at 50, 60, and 25.0 psi, respectively; the collision-activated dissociation (CAD) was high. Instrument tuning and mass calibration were performed with 10 and 100 μmol/L polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were acquired as MRM experiments with collision gas

(nitrogen) set to medium. DP and CE for individual MRM transitions was done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period. Metabolite data were log2-transformed for statistical analysis to improve normality and were normalized.

## 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

LFW, TL, LB., and KC. designed the study. KC and XS performed the analysis. KC, HY, and YP wrote the manuscript. KC, LFW, LMW, and ZZ revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by grants from the National Natural Science Foundation of China (No. 32272564 and No. 32001923), the National Key R&D Program of China (No. 2021YFD1700101), the science and Technology Innovation Program of Hunan Province (No. 2020WK2014 and No. 2020WK2023), the Training Program for Excellent Young Innovators of Changsha (kq2106079), and the China Agriculture Research System of MOF and MARA (CARS-16-E19).

## 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.2023.1155258/full#supplementary-material

## References

Alexa, A., and Rahnenfuhrer, J. (2010). topGO: enrichment analysis for gene ontology. *R Package version* 2.

Banerjee, B. D., Seth, V., and Ahmed, R. S. (2001). Pesticide-induced oxidative stress: perspective and trends. *Rev. Environ. Health* 16, 1–40. doi: 10.1515/reveh.2001.16.1.1

Belbin, F. E., Hall, G. J., Jackson, A. B., Schanschieff, F. E., Archibald, G., Formstone, C., et al. (2019). Plant circadian rhythms regulate the effectiveness of a glyphosate-based herbicide. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-019-11709-5

Bhatnagar-Mathur, P., Vadez, V., and Sharma, K. K. (2008). Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. *Plant Cell Rep.* 27, 411–424. doi: 10.1007/s00299-007-0474-9

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Cheng, Y., Shi, Z.-P., Jiang, L.-B., Ge, L.-Q., Wu, J.-C., and Jahn, G. C. (2012). Possible connection between imidacloprid-induced changes in rice gene transcription profiles and susceptibility to the brown plant hopper nilaparvata lugens stål (Hemiptera: Delphacidae). *Pesticide Biochem. Physiol.* 102, 213–219. doi: 10.1016/j.pestbp.2012.01.003

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

Ding, H., Lu, H., Lavoie, M., Xie, J., Li, Y., Lv, X., et al. (2014). Unraveling the toxicity mechanisms of the herbicide diclofop-methyl in rice: modulation of the activity of key enzymes involved in citrate metabolism and induction of cell membrane anion channels. *J. Agric. Food Chem.* 62, 10654–10660. doi: 10.1021/jf503974t

Edgar, R. S., Green, E. W., Zhao, Y., Van Ooijen, G., Olmedo, M., Qin, X., et al. (2012). Peroxiredoxins are conserved markers of circadian rhythms. *Nature* 485, 459–464. doi: 10.1038/nature11088

Filichkin, S. A., Breton, G., Priest, H. D., Dharmawardhana, P., Jaiswal, P., Fox, S. E., et al. (2011). Global profiling of rice and poplar transcriptomes highlights key conserved circadian-controlled pathways and cis-regulatory modules. *PloS One* 6, e16907. doi: 10.1371/journal.pone.0016907

Greenham, K., and McClung, C. R. (2015). Integrating circadian dynamics with physiological processes in plants. *Nat. Rev. Genet.* 16, 598–610. doi: 10.1038/nrg3976

Hayama, R., and Coupland, G. (2004). The molecular basis of diversity in the photoperiodic flowering responses of arabidopsis and rice. *Plant Physiol.* 135, 677–684. doi: 10.1104/pp.104.042614

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol. Cell* 38, 576–589. doi: 10.1016/j.molcel.2010.05.004

Hofer, R., Boachon, B., Renault, H., Gavira, C., Miesch, L., Iglesias, J., et al. (2014). Dual function of the cytochrome P450 CYP76 family from arabidopsis thaliana in the metabolism of monoterpenols and phenylurea herbicides. *Plant Physiol.* 166, 1149–1161. doi: 10.1104/pp.114.244814

Hu, L., Yao, Y., Cai, R., Pan, L., Liu, K., and Bai, L. (2020). Effects of fenclorim on rice physiology, gene transcription and pretilachlor detoxification ability. *BMC Plant Biol.* 20, 1–12. doi: 10.1186/s12870-020-2304-y

Hughes, M. E., Hogenesch, J. B., and Kornacker, K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J. Biol. rhythms* 25, 372–380. doi: 10.1177/0748730410379711

Iwakami, S., Endo, M., Saika, H., Okuno, J., Nakamura, N., Yokoyama, M., et al. (2014). Cytochrome P450 CYP81A12 and CYP81A21 are associated with resistance to two acetolactate synthase inhibitors in echinochloa phyllopogon. *Plant Physiol.* 165, 618. doi: 10.1104/pp.113.232843

Kato, S., Yokota, Y., Suzuki, R., Fujisawa, Y., and Ishimoto, M. (2020). Identification of a cytochrome P450 hydroxylase, CYP81E22, as a causative gene for the high sensitivity of soybean to herbicide bentazon. *Theor. Appl. Genet.* 133, 2105–2115. doi: 10.1007/s00122-020-03580-6

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., Mccombie, W. R., Ouyang, S., et al. (2013). Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 1–10. doi: 10.1186/1939-8433-6-4

Kim, W.-Y., Ali, Z., Park, H. J., Park, S. J., Cha, J.-Y., Perez-Hormaeche, J., et al. (2013). Release of SOS2 kinase from sequestration with GIGANTEA determines salt tolerance in arabidopsis. *Nat. Commun.* 4, 1–13. doi: 10.1038/ncomms2357

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

Krishnaiah, S. Y., Wu, G., Altman, B. J., Growe, J., Rhoades, S. D., Coldren, F., et al. (2017). Clock regulation of metabolites reveals coupling between transcription and metabolism. *Cell Metab.* 25, 1206. doi: 10.1016/j.cmet.2017.04.023

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

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, Q., Brown, J. B., Huang, H., and Bickel, P. J. (2011). Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* 5, 1752–1779. doi: 10.1214/11-AOAS466

Li, Y., Lu, Y., Zhou, Y., Wei, X., Peng, Y., Dai, Y., et al. (2021). Diurnal transcriptomics analysis reveals the regulatory role of the circadian rhythm in super-hybrid rice LY2186. *Genomics* 113, 1281–1290. doi: 10.1016/j.ygeno.2020.12.046

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, N., and Zhu, L. (2020). Metabolomic and transcriptomic investigation of metabolic perturbations in oryza sativa l. triggered by three pesticides. *Environ. Sci. Technol.* 54, 6115–6124. doi: 10.1021/acs.est.0c00425

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, 1–21. doi: 10.1186/s13059-014-0550-8

Maggi, F., Tang, F. H., La Cecilia, D., and Mcbratney, A. (2019). PEST-CHEMGRIDS, global gridded maps of the top 20 crop-specific pesticide application rates from 2015 to 2025. *Sci. Data* 6, 1–20. doi: 10.1038/s41597-019-0169-4

Martinson, K. B., Sothern, R. B., Koukkari, W. L., Durgan, B. R., and Gunsolus, J. L. (2002). Circadian response of annual weeds to glyphosate and glufosinate. *Chronobiol. Int.* 19, 405–422. doi: 10.1081/cbi-120002877

Miller, R. P., Martinson, K. B., Sothern, R. B., Durgan, B. R., and Gunsolus, J. L. (2003). Circadian response of annual weeds in a natural setting to high and low application rates of four herbicides with different modes of action. *Chronobiol. Int.* 20, 299–324. doi: 10.1081/cbi-120019344

Pan, L., Yu, Q., Wang, J., Han, H., Mao, L., Nyporko, A., et al. (2021). An ABCC-type transporter endowing glyphosate resistance in plants. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2100136118. doi: 10.1073/pnas.2100136118

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, 290–295. doi: 10.1038/nbt.3122

Sasaki, T., and Burr, B. (2000). International rice genome sequencing project: the effort to completely sequence the rice genome. *Curr. Opin. Plant Biol.* 3, 138–142. doi: 10.1016/s1369-5266(99)00047-3

Song, Y. H., Ito, S., and Imaizumi, T. (2010). Similarities in the circadian clock and photoperiodism in plants. *Curr. Opin. Plant Biol.* 13, 594–603. doi: 10.1016/j.pbi.2010.05.004

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

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

Wang, F., Han, T., Song, Q., Ye, W., Song, X., Chu, J., et al. (2020). The rice circadian clock regulates tiller growth and panicle development through strigolactone signaling and sugar sensing. *Plant Cell* 32, 3124–3138. doi: 10.1105/tpc.20.00289

Wei, H., Wang, X., He, Y., Xu, H., and Wang, L. (2021). Clock component OsPRR73 positively regulates rice salt tolerance by modulating OsHKT2; 1-mediated sodium homeostasis. *EMBO J.* 40, e105086. doi: 10.15252/embj.2020105086

Wei, H., Xu, H., Su, C., Wang, X., and Wang, L. (2022). Rice CIRCADIAN CLOCK ASSOCIATED 1 transcriptionally regulates ABA signaling to confer multiple abiotic stress tolerance. *Plant Physiol.* 190, 1057–1037. doi: 10.1093/plphys/kiac196

Yang, L., Liu, P., Wang, X., Jia, A., Ren, D., Tang, Y., et al. (2021). A central circadian oscillator confers defense heterosis in hybrids without growth vigor costs. *Nat. Commun.* 12, 1–14. doi: 10.1038/s41467-021-22268-z

Zhai, R., Ye, S., Zhu, G., Lu, Y., Ye, J., Yu, F., et al. (2020). Identification and integrated analysis of glyphosate stress-responsive microRNAs, lncRNAs, and mRNAs in rice using genome-wide high-throughput sequencing. *BMC Genomics* 21, 1–15. doi: 10.1186/s12864-020-6637-6

Zhang, C., Gao, M., Seitz, N. C., Angel, W., Hallworth, A., Wiratan, L., et al. (2019). LUX ARRHYTHMO mediates crosstalk between the circadian clock and defense in arabidopsis. *Nat. Commun.* 10, 1–14. doi: 10.1038/s41467-019-10485-6

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, 1–9. doi: 10.1186/gb-2008-9-9-r137

Zhou, J., Liu, C., Chen, Q., Liu, L., Niu, S., Chen, R., et al. (2022). Integration of rhythmic metabolome and transcriptome provides insights into the transmission of rhythmic fluctuations and temporal diversity of metabolism in rice. *Sci. China Life Sci.* 65, 1794–1810. doi: 10.1007/s11427-021-2064-7

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 April 2023
DOI 10.3389/fpls.2023.1125065



OPEN ACCESS

EDITED BY
Milen I. Georgiev,
Bulgarian Academy of Sciences, Bulgaria

REVIEWED BY
Soheil Mahmoud,
University of British Columbia, Canada
Bhagwat Dadarao Nawade,
Kongju National University,
Republic of Korea
Sandrine Moja,
Université Jean Monnet, France
Michael Phillips,
University of Toronto, Canada

*CORRESPONDENCE
B. Markus Lange
lange-m@wsu.edu

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

RECEIVED 15 December 2022
ACCEPTED 20 March 2023
PUBLISHED 14 April 2023

CITATION
Lange BM, Srividya N, Lange I, Parrish AN, Benzenberg LR, Pandelova I, Vining KJ and Wüst M (2023) Biochemical basis for the formation of organ-specific volatile blends in mint.
*Front. Plant Sci.* 14:1125065.
doi: 10.3389/fpls.2023.1125065



# Biochemical basis for the formation of organ-specific volatile blends in mint

B. Markus Lange[1]*, Narayanan Srividya[1], Iris Lange[1], Amber N. Parrish[1], Lukas R. Benzenberg[1,2], Iovanna Pandelova[3], Kelly J. Vining[3] and Matthias Wüst[2]

[1]Institute of Biological Chemistry and M.J. Murdock Metabolomics Laboratory, WashingtonState University, Pullman, WA, United States, [2]Institut für Ernährungs- und Lebensmittelwissenschaften, Rheinische Friedrich Wilhelms-UniversitätBonn, Bonn, Germany, [3]Department of Horticulture, Oregon State University, Corvallis, OR, United States

Above-ground material of members of the mint family is commercially distilled to extract essential oils, which are then formulated into a myriad of consumer products. Most of the research aimed at characterizing the processes involved in the formation of terpenoid oil constituents has focused on leaves. We now demonstrate, by investigating three mint species, peppermint (*Mentha* $^{x}$ *piperita* L.), spearmint (*Mentha spicata* L.) and horsemint (*Mentha longifolia* (L.) Huds.; accessions CMEN 585 and CMEN 584), that other organs – namely stems, rhizomes and roots – also emit volatiles and that the terpenoid volatile composition of these organs can vary substantially from that of leaves, supporting the notion that substantial, currently underappreciated, chemical diversity exists. Differences in volatile quantities released by plants whose roots had been dipped in a *Verticillium dahliae*-spore suspension (experimental) or dipped in water (controls) were evident: increases of some volatiles in the root headspace of mint species that are susceptible to Verticillium wilt disease (peppermint and *M. longifolia* CMEN 584) were detected, while the quantities of certain volatiles decreased in rhizomes of species that show resistance to the disease (spearmint and M. longifolia CMEN 585). To address the genetic and biochemical basis underlying chemical diversity, we took advantage of the newly sequenced *M. longifolia* CMEN 585 genome to identify candidate genes putatively coding for monoterpene synthases (MTSs), the enzymes that catalyze the first committed step in the biosynthesis of monoterpenoid volatiles. The functions of these genes were established by heterologous expression in *Escherichia coli*, purification of the corresponding recombinant proteins, and enzyme assays, thereby establishing the existence of MTSs with activities to convert a common substrate, geranyl diphosphate, to (+)-α-terpineol, 1,8-cineole, γ-terpinene, and (–)-bornyl diphosphate, but were not active with other potential substrates. In conjunction with previously described MTSs that catalyze the formation of (–)-β-pinene and (–)-limonene, the product profiles of the MTSs identified here can explain the generation of all major monoterpene skeletons represented in the volatiles released by different mint organs.

KEYWORDS

**Mentha, mint, monoterpene synthase, terpene, Verticillium, volatile**

## Introduction

Members of the mint family are the source of valuable consumer products, including essential oils. Commonly grown cultivars in temperate climates are peppermint (*Mentha* x *piperita* L.) and spearmint (*Mentha spicata* L.). The dominant monoterpenoids in these oils are derived from a common biosynthetic intermediate, (–)-limonene, but the pathway is branched, thereby leading to the accumulation of (–)-carvone in spearmint, while (–)-menthone and (–)-menthol are major constituents of peppermint (Lange, 2015a) (Figure 1). The Black Mitcham peppermint and Native spearmint cultivars, grown widely across the United States, are polyploid, sterile hybrids, and diploid ancestral species, such as horsemint (*Mentha longifolia* (L.) Huds.), have thus been advanced as model species for genetic analyses (Vining et al., 2005). Substantial chemical diversity of distilled oils has been reported for members of this genus but there is also considerable variation in the resistance to Verticillium wilt (Vining et al., 2005; Vining et al., 2022), caused by the soil-borne fungus *Verticillium dahliae* Kleb., the most destructive disease of mint in the United States (Dung, 2020). The fungus initially infects roots, subsequently invades the vascular system, and ultimately causes occlusion and, thus, restricted water movement (Daayf, 2015).

The commercial process for obtaining mint oils starts with cutting above-ground material and leaving it in the field to dry (Denny and Lawrence, 2007). The mint is then chopped and blown into a large container (mint tub) that sits on a trailer. The trailer is pulled to a still, where the tub is connected to a boiler *via* a hose at the bottom and to a condenser line at the top. Steam from the boiler is introduced into the tub and distributed through pipes on the tub floor. Volatilized mint constituents rise with the introduced steam and enter the condenser line, which carries the water-oil mixture through a series of water-cooled smaller lines, thus forming a condensate that is collected in a separator tank. The lower density oil forms a separate phase on top of the water and is drained into galvanized drums (Denny and Lawrence, 2007). Oil distillations can be performed on a laboratory scale in purpose-built glassware, such as a Likens-Nickerson apparatus, which was originally developed for capturing hop oil and uses two boiling flasks connected to a mixer/condenser (Likens and Nickerson, 1964). One flask is loaded with mint leaves in water, while the other contains an immiscible organic solvent (usually *n*-hexane) (Vining et al., 2019). Each flask is heated to the boiling point of the contained liquid, thus sending vapors along the arms of the apparatus to the central cold finger condenser where they are mixed and condensed. The lighter organic phase is then collected in a separatory trap with a stopcock (Likens and Nickerson, 1964). This approach collects volatiles in an organic solvent volume of typically 8 to 10 ml, of which only a small fraction (usually 1 µl) is injected onto a Gas Chromatograph (GC) for subsequent analysis. Mint leaves have anatomical structures called glandular trichomes that store copious amounts of volatiles and other constituents, and the substantial dilution of analytes by laboratory-scale distillation is thus acceptable for volatile analyses (Amelunxen, 1965; Amelunxen et al., 1969; Lange and Turner, 2013; Lange, 2015a). One main goal of the current study was to assess if organs other than leaves may produce volatiles that have not been reported to occur in mint before, thus enhancing our understanding of the chemical diversity. Roots, rhizomes, and stems have very small glandular trichomes or none (Kowalski et al., 2019) and volatile concentrations would thus be expected to be comparatively low. In such cases, trapping of volatilized metabolites from the headspace above samples, enabled by techniques such as Solid Phase Microextraction (SPME), is generally more suitable (Tholl et al., 2006). Several studies have demonstrated that SPME is a reliable method for mint volatile analysis (Rohloff, 1999; Cordero et al., 2012; Silva and Câmara, 2013; Taherpour et al., 2017), and we thus developed a SPME-based approach for assessing volatiles in selected mint species.

To further explore chemical diversity, we assessed if the inoculation of roots with *V. dahliae* has effects on the volatile repertoire of mint organs relevant for the early stages of infection (roots, rhizomes, and stems). It is well known that mint-fungal interactions, both beneficial and detrimental, can have dramatic effects on the production of terpenoid volatiles (Karagiannidis et al., 2011; Arango et al., 2012; Bharti et al., 2013; Johnson et al., 2013; del Rosario Cappellari et al., 2015; Urcoviche et al., 2015; Khalvandi

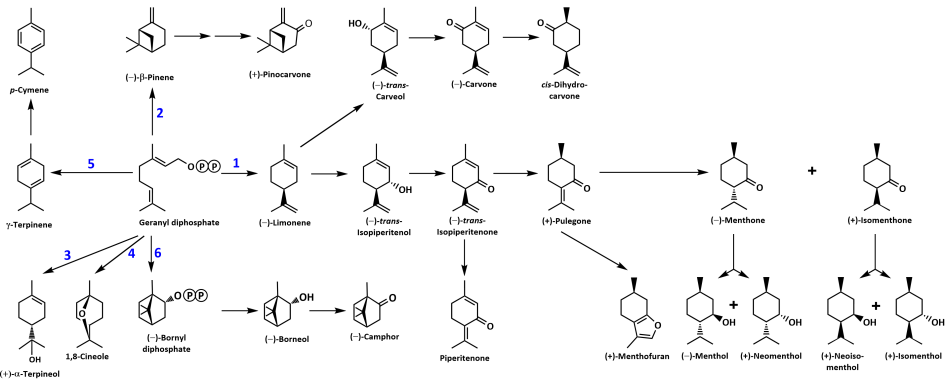



FIGURE 1
Overview of the mint monoterpene biosynthetic pathway. The circled letter P denotes a phosphate moiety. Reactions catalyzed by MTSs are numbered: 1, (–)-limonene synthase; 2, (–)-β-pinene synthase; 3, (–)-α-terpineol synthase; 4, 1,8-cineole synthase; 5, γ-terpinene synthase; and 6, (–)-bornyl diphosphate synthase.

et al., 2019; Merlin et al., 2020; Khalvandi et al., 2021). The interactions between mint and fungi can be mutualistic (e.g., the mycorrhizal fungus *Glomus fasciculatum*) (Gupta et al., 2002), endosymbiotic (Mucciarelli et al., 2007) or pathogenic (Dung, 2020). Following the analysis of volatiles released by different organs of mint plants grown under various conditions, we took advantage of the recently completed chromosome-level assembly of the *M. longifolia* CMEN 585 genome and published transcriptome data sets for several mint species (Ahkami et al., 2015; Vining et al., 2017; Chen et al., 2021; Vining et al., 2022) to identify candidate genes coding for monoterpene synthases (MTSs), the enzymes that catalyze the first committed step of monoterpenoid volatile biosynthesis. The products generated by these as yet uncharacterized MTSs, in conjunction with those characterized previously, can explain the volatile profiles observed in our experiments.

# Materials and methods

## Plant materials and greenhouse growth conditions

Peppermint (Black Mitcham cultivar) and spearmint (Native cultivar) were supplied by the Mint Industry Research Council (Salem, OR, USA). *M. longifolia* accessions CMEN 585 (PI 557767) and CMEN 584 (PI 557769) were obtained from the U.S. Department of Agriculture National Clonal Germplasm Repository (Corvallis, OR, USA). Greenhouse-grown plants were maintained under the following conditions: temperature at 23–27° C; humidity at 60–75%; natural lighting supplemented with sodium vapor lights to generate a consistent 16 h day/8 h night cycle; average light intensity during the day was 250 µE $m^{-2}$ $s^{-1}$; daily watering and fertilizer treatment once a week with Peters 20/20/20 (R.J. Peters Inc., Chesterfield, MI, USA).

## Growth chamber growing conditions for *V. dahliae* inoculations and mock-inoculated controls

Plant material for cultivation in growth chambers was established from 6-10 cm cuttings taken off apical and auxiliary stems of pre-flowering greenhouse-grown plants. Cuttings were placed in plastic trays filled with autoclaved Sunshine Mix #1-Fafard-1P (Sun Gro Horticulture; Agawam, MA, USA), moistened with autoclaved double-distilled water, and the trays placed in plastic flats with an 8 cm humidity dome. Trays were moved to a E-75L1 growth chamber (Percival Scientific, Perry, IA, USA) fitted with a programmable light timer (Woods Model # 59745, supplied by Coleman Cable, Waukegan, IL, USA). The growth chamber was set at 22°C, 10 h light/14 h dark for peppermint and spearmint; 22° C, 8 h light/16 h dark for *M. longifolia* accessions CMEN 584 and CMEN 585 (average light intensity during illumination was 150 µE $m^{-2}$ $s^{-1}$). Flats were rotated daily and moistened with autoclaved, double-distilled water. The 8 cm humidity dome was replaced with a 20 cm humidity dome after one week. Humidity domes were removed completely after two weeks, and plants were not watered the day before performing root dips to aid in soil removal from roots. The *V. dahliae* strain used in our experiments had been collected from infected mint fields in Oregon (Vining and Pandelova, 2022). Fungal suspensions were initiated by placing a 1 $cm^2$ agar plug into 125 ml of liquid Czapek-Dox broth (PhytoTechnology Laboratories, Lenexa, KS, USA). The liquid cultures were placed in a designated shaker at 180 rpm and 23°C for 7 – 9 days. Fungal suspensions for inoculation were prepared by filtering the culture through autoclaved Miracloth (Calbiochem, La Jolla, CA, USA) using an autoclaved Büchner funnel and flask setup with a UN810 FTP Laboport vacuum pump (KNF Neuberger, Trenton, NJ, USA). The filtered conidia suspension was evenly distributed across centrifuge tubes and spun down at 10,000 x *g* for 5 min. Autoclaved, double-distilled water was added to the pellets, mixed, and diluted to a final concentration of $10^7$ conidia per ml solution (determined by use of a Bright-Line hemacytometer (Hausser Scientific, Horsham, PA, USA) under a light microscope (DM LM, Leica Microsystems, Wetzlar, Germany). Root dips were performed by shaking off soil from plant roots and dipping roots into glass jars of either 10 ml autoclaved, double-distilled water (controls) or 10 ml *V. dahliae* suspension (experimental treatment). Roots were held in the jars for 5 min, before being planted in new plastic trays prefilled with soil as described above. Control and experimental treatment flats were placed on the same growth chamber platform and rotated daily for 6 d before harvesting (growing conditions as outlined above). Plants were watered sparingly to facilitate inoculation.

## Harvest of plant material and capturing volatiles by hydrodistillation

Leaves were picked from stems, avoiding the top and bottom nodes (approximately 3 g fresh weight). Hydrodistillations were performed in a modified Likens-Nickerson apparatus for 60 min using *n*-hexane as carrier solvent based on a previously published method (Ringer et al., 2003). An aliquot of the *n*-hexane fraction, which contained the volatile oil constituents, was transferred to a 2-mL glass vial. A 1 ml aliquot of this extract was analyzed by Gas Chromatography – Mass Spectrometry (GC–MS) as described under "Separation, Detection and Analysis of Volatiles".

## Harvest of plant material, trapping of volatiles by static headspace SPME, and analysis of volatiles by GC–MS

Plants were removed from soil and briefly shaken. Rhizomes and roots were rinsed with autoclaved double-distilled water before separating them with scissors. Additional portions of soil were removed from roots with tweezers, then rinsed again, and gently dabbed dry between two sheets of clean paper towels to remove excess water. Rhizome segments of 0.5 cm were cut away from the lowest portion of stem and dabbed on a separate paper towel to

remove excess water. Stem segments of 0.5 cm were cut with scissors at 2.5 cm above the soil line, avoiding nodes. Scissor blades and tweezer tips were wiped down using Kimwipes (Kimberly-Clark-Kimtech, Irving, TX, USA) first wetted with 70% EtOH, then wiped down again using Kimwipes saturated with autoclaved double-distilled water. Oil samples from leaves were taken by gently rubbing the surfaces of three leaves with a single 2.5 cm x 1 cm piece of clean paper towel. This was necessary because the quantity of volatiles released from leaves was very high and quickly saturated the SPME fiber. All plant tissue samples were weighed in pre-tared 20 ml screw-thread glass vials (Microliter Analytical Supplies, Suwanee, GA, USA), while the paper leaf rubbings were loosely folded and dropped into untarred vials. All vials were sealed within 60 seconds using 16 mm screw-top caps (butyl rubber septa with polytetrafluoroethylene coating (Wheaton MicroLiter; septa were prebaked overnight at 150°C). A Leap Technologies Combi PAL autosampler was outfitted with SPME syringe adapter, agitator module, and fiber conditioning station (CTC Analytics AG, Zwingen, Switzerland). SPME fiber assemblies were 23 gauge 65 µm PDMS/DVB StableFlex (Supelco; Bellefonte, PA, USA) used in conjunction with a Merlin Microseal High Pressure Nut Kit (Restek, Bellefonte, PA, USA) and 0.75 mm id Ultra Inert splitless liner with SPME taper (Agilent Technologies, Inc.; Santa Clara, CA, USA). Samples were preincubated at 60°C in the agitator (rotating at 250 rpm) for 10 min, before a SPME fiber was inserted (vial penetration of 22 mm) and exposed to headspace volatiles for 15 min. The SPME fiber was then inserted into the GC inlet at an injection depth of 54 mm and volatiles were desorbed for 10 s at 220°C in splitless injection mode, followed by a purge step (1 min at 20 ml/min) (6890N system; Agilent Technologies, Santa Clara, CA, USA). The fiber was reconditioned under a stream of $N_2$ for 3 min at 250°C right after the injection.

## Separation, detection and analysis of volatiles

The separation of headspace volatiles was achieved over a DB-5MS column (30 m $^{x}$ 0.25 mm, 0.25 µm film thickness; J&W Scientific, Folsom, CA, USA) at a flow rate of 1 ml/min He. The oven program started at 165°C, with a ramp of 25°C/min to 215°C, and a final hold at this temperature for 3 min. Through a transfer line, the column end reached into the ion source of a quadrupole MS detector (5973 *inert* system; Agilent Technologies, Santa Clara, CA, USA). The ion source was set to 230°C with an ionization voltage of 70 eV (electron ionization). The scan range was set to 40 – 350 *m/z*. The plant material after volatile capture was dried in an oven at 50°C overnight and weighed to determine the dry weight. Chiral analyses were performed on the same GC–MS system using a Cyclodex-B column (30 m $^{x}$ 0.25 mm, 0.25 µm film thickness; J&W Scientific, Folsom, CA, USA). The following GC settings were used: injector temperature 250°C; injector split 1:20; sample volume 1 µl; flow rate 2 ml/min with He as carrier gas; oven heating program with an initial ramp from 40°C to 120°C at 2°C/min, followed by a second ramp to 200°C at 50°C/min, and a final hold at 200°C for 2 min. Identification of volatiles was achieved by matching retention times and target ions to those of authentic standards.

Data analysis was performed with the ChemStation E.02.00.493 software (Agilent Technologies, Santa Clara, CA, USA), with peak annotation and relative quantitation being achieved based on a combination of i) ion extraction *via* ChemStation's QEdit function, ii) a dedicated spectral library for volatiles (Adams, 2007), iii) the NIST 20 library (National Institute of Standards and Technology, Gaithersburg, MD, USA), and iv) an extensive in-house library of authentic standards for volatiles (purchased from commercial sources, isolated from plant oils or obtained by chemical synthesis; the identity of all compounds in this library has been confirmed using GC–MS and nuclear magnetic resonance spectroscopy). The QEdit function in MSD ChemStation was specifically used to identify peaks of closely eluting compounds using pre-defined retention times and m/z values. Obtaining absolute values (ng per g tissue or similar) when acquiring static headspace SPME data is extremely challenging and only possible for a small number of selected analytes. It would require optimization, for each individual analyte, of fiber affinity, sensitivity, reproducibility, and linear range; for complex matrices, one would also have to assess competition for all analytes; additional parameters for optimization would be the extraction time and temperature for each analyte; finally, data for an external standard curve would have to be acquired for each analyte present in a matrix that mimics the samples but does not contain the analyte (Roberts et al., 2000). Since the goal of the current study was to assess chemical diversity (which includes a larger number of analytes), we chose to employ a simplified quantitation approach. Tissue used for volatile analyses was subsequently dried (48 h at 60° C) and the weight determined, which allowed the quantitation of analytes as peak abundance per unit weight, which is comparable across samples and the closest we can get to absolute quantitation within the scope of the study. Statistical analyses were performed using a two-tailed Mann-Whitney U test in Excel (Microsoft 365, Redmond, WA, USA).

## Phylogenetic analysis of MTSs

The sequences of functionally characterized MTSs from Lamiaceae were obtained from the National Center for Biotechnology Information (NCBI) protein database. MTS candidates of *M. longifolia* CMEN 585 had been identified previously based on the recently completed chromosome-level genome assembly (Chen et al., 2021; Vining et al., 2022). One MTS candidate from *M.* $^{x}$ *piperita* (Black Mitcham cultivar) was identified by searching the translated transcriptome consensus assembly for leaf glandular trichomes of this species (NCBI Short Read Archive BioProject Identifier PRJNA 245825) against functionally characterized 1,8-cineole synthases of the Lamiaceae with a 60% sequence identity cutoff (sequences provided in Supplemental Table S1). The evolutionary history of MTSs was inferred by using the Maximum Likelihood method and JTT matrix-based model (Jones et al., 1992). The bootstrap consensus

tree inferred from 1,000 replicates (Felsenstein, 1985) was taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. Initial trees for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the JTT model (Jones et al., 1992). This analysis involved 40 amino acid sequences (Supplemental Table S1). Evolutionary analyses were conducted in the MEGA software (version 11) (Tamura et al., 2021).

## Functional characterization of MTSs

Synthetic cDNAs corresponding to candidate transcripts Ml_15999 and Ml_14850 (excluding plastidial targeting sequences) were obtained from Twist Bioscience (San Francisco, CA, USA or Biomatik (Wilmington, DE, USA), while the cDNA corresponding to candidate Ml_38055 (also excluding plastidial targeting sequence) was amplified from first-strand cDNA representing transcripts present in isolated *M. longifolia* CMEN 585 glandular trichomes (forward primer: ATGCGACGATCCGGAAATTACAAACCTAC; reverse primer: CTAAACAATAGGTTCAAATATGAGACCTAAAATG). A synthetic gene corresponding to peppermint candidate Mp_12413 (excluding plastidial targeting sequence) was ordered from Twist Bioscience (San Francisco, CA, USA). The accession numbers of all cDNAs are provided at the end of this article. All genes were integrated into the multiple cloning site of the pET28b expression vector (Novagen, Madison, WI, USA). Plasmids were transformed into chemically competent cells of *E. coli* strain BL21, which were then grown in 5 ml of liquid LB medium and move to 150 ml culture at 37°C with shaking at 350 rpm (Series 25, New Brunswick, ThermoFisher Scientific, Waltham, MA, USA) to an $OD_{600}$ of 0.8. Expression of MTS genes was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (GoldBio, St. Louis, MO, USA) and cells were grown for another 24 h at 16°C and 350 rpm (Series 25, New Brunswick, ThermoFisher Scientific, Waltham, MA, USA). Bacterial cells were harvested by centrifugation at 5,000 x *g* and resuspended in 300 ul MOPSO buffer, pH 7.0, supplemented with 1 mM dithiothreitol (DTT; GoldBio, St. Louis, MO, USA). Cells were lysed using a model 475 sonicator (VirTis, Gardiner, NY, USA), with three 15 s bursts and cooling on ice for 45 s between bursts. The resulting homogenate was centrifuged at 15,000 x *g* for 30 min at 4°C, and recombinant protein present in the supernatant purified over $Ni^{2+}$ affinity columns according to the manufacturer's instructions (Marvelgent Biosciences, Canton, MA, USA). *In vitro* assays were performed in 2-ml glass vials containing 200 µg purified enzyme in MOPSO buffer (50 mM) containing DTT (1 mM) and MgCl2 (50 mM) (total volume of reaction mixture 100 µl). A prenyl diphosphate substrate (geranyl diphosphate, neryl diphosphate, (*E,E*)-farnesyl diphosphate or (*Z,Z*)-farnesyl diphosphate) was added to a final concentration of 0.5 mM. The assay mixtures were overlaid with 100 µl *n*-hexane (containing 15 ng/µL camphor as internal standard for subsequent analyses) (Avantor, Center Valley, PA, USA) and incubated at 30°C for 16 h on a multi-tube rotator (Labquake, Barnstead Thermolyne, Ramsey, MN, USA). Following addition of 20 µL of a solution containing 3 units of potato apyrase and 0.5 units of wheat germ acid phosphatase (SigmaAldrich, St. Louis, MO, USA) in 1 M sodium acetate (pH 5.0), the mixture was rotated for another 24 h at 31°C. To assess if bornyl diphosphate was generated as a product (which was then hydrolyzed to borneol by the addition of apyrase and acid phosphatase) or borneol was produced directly, we also performed assays without the addition of apyrase and acid phosphatase (no borneol was detected). Furthermore, negative control assays with boiled enzyme extracts were performed (no band visible on gels, no activity detected). Enzymatic reactions were stopped by vigorous mixing of the contents of the tubes, followed by 30 min at -20°C for phase separation (aqueous phase solidifies and organic phase remains liquid on top). The organic phase was removed and transferred to glass vial inserts and stored in GC vials at -20°C until further analysis. Assay products were analyzed by chiral GC analysis as described previously (Srividya et al., 2016). The composition of assay products was determined based on calibration curves obtained with authentic standards.

# Results

## Pilot experiment and experimental design considerations

To test the suitability of static headspace SPME for the analysis of mint volatile diversity, we performed a pilot experiment using greenhouse-grown plants. Leaves were subjected to hydrodistillation (*n*-hexane as carrier solvent) with subsequent analysis by GC–MS. The extract obtained after hydrodistillation was processed in one of two ways: i) direct injection onto the GC or ii) placing an aliquot of the extract in a glass vial to allow volatiles released into the headspace to be absorbed by a SPME fiber, which was then inserted into the GC inlet for volatile desorption. Representatives of four accessions were analyzed in this pilot experiment: peppermint (Black Mitcham cultivar), spearmint (Native cultivar), and *M. longifolia* accessions CMEN 585 and CMEN 584. The profile of the main terpenoid volatiles was generally very similar across these samples (Table 1 and Supplemental Table S2). (–)- Menthone and (+)-menthofuran were the most prominent constituents of oils extracted from peppermint (55 – 58% and 14 – 16%, respectively), (–)-carvone was the principal volatile of spearmint and *M. longifolia* CMEN 584 oils (79 – 84% and 70 – 81%, respectively), and (+)-pulegone was the most abundant terpenoid of *M. longifolia* CMEN 585 oils (69 – 74%). However, a substantially lower amount of germacrene D (< 1%) was detected by headspace SPME when compared to the directly injected liquid extract (5%) in *M. longifolia* CMEN 585 (Table 1 and Supplemental Table S2), indicating that SPME introduces a bias against higher-boiling sesquiterpenes, which are only minor constituents of the mint species of interest in the present study. A third sample type was generated by gently rubbing leaves against a rectangular piece of paper towel, which was then placed in a glass vial, and volatiles released into the headspace were trapped by SPME. This adjustment was necessary because even the smallest quantity of leaf material we could weigh consistently (0.5 mg) emitted quantities of volatiles that exceeded the capacity of the SPME fiber. Distilled peppermint extracts measured *via* SPME had a lower quantity of

TABLE 1 Terpenoid profiles of mint leaves observed when employing different volatile capturing methods (expressed as percent of total volatiles with standard error; n = 3-5).

| Species (Cultivar/Accession) | Terpenoid | Hexane Extract (Direct Injection) | Hexane Extract (SPME) | Leaf Rub (SPME) |
|---|---|---|---|---|
| *Mentha* x *piperita* L. | (-)-Menthone | 54.7 ± 0.5 | 57.7 ± 1.6 | 63.5 ± 2.4 |
| (Black Mitcham) | (+)-Menthofuran | 14.4 ± 0.3 | 16.4 ± 0.5 | 0.3 ± 0.1 |
| | (-)-Menthol | 6.6 ± 0.1 | 4.7 ± 1.5 | 13.0 ± 1.3 |
| *Mentha spicata* L. | (-)-Carvone | 79.3 ± 0.9 | 83.6 ± 2.4 | 83.1 ± 8.5 |
| (Native) | (-)-Limonene | 7.7 ± 0.5 | 5.3 ± 1.3 | 2.5 ± 1.4 |
| *Mentha longifolia* (L.) Huds. | (+)-Pulegone | 68.9 ± 1.5 | 74.3 ± 5.4 | 90.1 ± 0.9 |
| (CMEN 585) | Piperitenone | 9.8 ± 0.5 | 3.4 ± 2.5 | 0.6 ± 0.1 |
| | Germacrene D | 5.2 ± 0.1 | 0.9 ± 0.6 | 0.2 ± 0.1 |
| *Mentha longifolia* (L.) Huds. | (-)-Carvone | 70.0 ± 1.0 | 80.5 ± 0.8 | 85.7 ± 1.7 |
| (CMEN 584) | (-)-Limonene | 14.1 ± 1.3 | 7.6 ± 1.2 | 2.8 ± 0.9 |
| | (-)-Borneol | 5.7 ± 0.1 | 4.6 ± 1.2 | 4.9 ± 1.3 |

For details see text. SPME, head space solid phase microextraction.

(–)-menthone (58% versus 64%) and higher amount of (+)-menthofuran (16% versus 0.3%), and a lower amount of (–)-menthol (5% versus 13%) compared to those processed by static headspace SPME of leaf rubs (Table 1). (–)-Limonene was generally higher in distilled samples (all accessions). (+)-Pulegone levels were lower in distilled samples of *M. longifolia* CMEN 585 (74% versus 90%) (Table 1). While the relative percentages of volatiles differed between extracts and headspace samples, each tested method was highly reproducible (Supplemental Table S2); because our main goal was to investigate the chemical diversity of volatile blends (which was captured by all methods), we thus proceeded with headspace SPME to allow direct comparisons across organs.

## Mint plants emit species- and organ-specific blends of terpenoid volatiles

Volatile bouquets emitted by stems and rhizomes from greenhouse-grown peppermint plants were characterized by high amounts of (+)-menthofuran (> 75% of detected volatiles), when compared to leaves, which released (–)-menthone (> 28%) and (–)-menthol as the most abundant monoterpenoids (> 20%) (Figure 2A and Supplemental Table S3). Root samples of peppermint had a unique volatile profile that, in addition to the presence of (+)-menthofuran (53%), was characterized by high amounts of (–)-borneol (11%), (–)-camphor (5%), (–)-menthyl acetate (5%), and (+)-pulegone (5%). Volatile emission from roots was particularly low compared to all other sample types (Figure 2B).

(–)-Carvone was the most abundant monoterpenoid in the volatile blends emitted by spearmint plants across organs (≥ 80% in leaf rubs, stems, 76% in rhizomes; 25% in roots) (Figure 2C and Supplemental Table S3). Leaf samples also released sesquiterpenes (most prominently germacrene D (7%) and (*E*)-β-farnesene (4%)) as volatiles of note. Stem samples had (–)-*trans*-carveol and *cis*-dihydrocarvone as additional noteworthy, volatilized metabolites. Rhizomes formed (–)-limonene (12%) and *p*-cymene (4%) as further prominent volatiles. The root volatile profile was more complex, with substantial contributions, in addition to (–)-carvone, coming from (–)-limonene (16%), *p*-cymene (7%), (–)-borneol (7%), and (+)-pinocarvone (6%) (Figure 2C and Supplemental Table S3). Volatile emission from leaf and stem samples was substantially higher than that of rhizomes and roots (Figure 2D).

The CMEN 585 accession of *M. longifolia* emitted (+)-pulegone as signature volatile from all samples (84% in leaves, 66% in stems, 39% in rhizomes, and 47% in roots) (Figure 2E and Supplemental Table S3). Other notable constituents of leaf volatile blends were piperitenone (6%) and germacrene D (5%). Stems released (+)-isomenthone (18%) and (–)-borneol (4%) as additional volatiles of note. The headspace above rhizome samples contained, in addition to (+)-pulegone, (+)-isomenthone (36%), (–)-menthone (9%), and (+)-menthofuran (5%) as abundant volatiles. Roots released (–)-borneol (14%) and (+)-isomenthone (13%) as noteworthy metabolites (in addition to (+)-pulegone) (Figure 2E and Supplemental Table S3). Volatile emission from leaves was substantially higher than that of all other sample types (Figure 2F).

The leaf profile of the *M. longifolia* accession CMEN 584 resembled that of spearmint, with (–)-carvone being dominant in the blends released by leaf rubbings (88%), stems (89%) and rhizomes (79%), with a slightly higher amount in roots (35%) (Figure 2G and Supplemental Table S3). The only other notable constituent in leaf surface oils was germacrene D (6%). Notable volatile constituents in rhizomes, aside from (–)-carvone, were

(–)-limonene (9%) and (–)-borneol (5%). Stems emitted, in addition to (–)-carvone, smaller quantities of (–)-*trans*-carveol, *cis*-dihydrocarvone, and (–)-limonene (slightly above 2% each). The most abundant volatiles of roots were (–)-borneol (43%), (–)-carvone (35%), and (–)-camphor (9%) (Figure 2G and Supplemental Table S3). Volatile emission from leaf and stem samples was substantially higher than that of rhizomes and roots (Figure 2H).

In summary, our analyses by SPME–GC–MS showed substantial differences regarding volatiles emitted by different mint species, consistent with prior analyses using hydrodistillation (Ahkami et al., 2015; Vining et al., 2022). In addition, even within a species, each organ produced a unique blend of volatiles, which is novel information adding to our knowledge of the repertoire for volatile formation.

## Experimental conditions to assess the effects of *V. dahliae* inoculation increase volatile chemical diversity

Our previous analysis of the *M. longifolia* genome (accession CMEN 585) indicated the presence of genes coding for more than fifty terpene synthases (Chen et al., 2021). It is well known that the expression of terpene synthase genes is not only organ-, tissue, and cell type-specific, but they are also inducible by different environmental and experimental conditions (Degenhardt et al., 2009). To evaluate the inducibility of terpene emission, we designed a study to assess volatile blends released after inoculation of roots with the wilt pathogen *V. dahliae*. Two accessions with significant wilt resistance (spearmint and *M. longifolia* CMEN 585) and two susceptible accessions

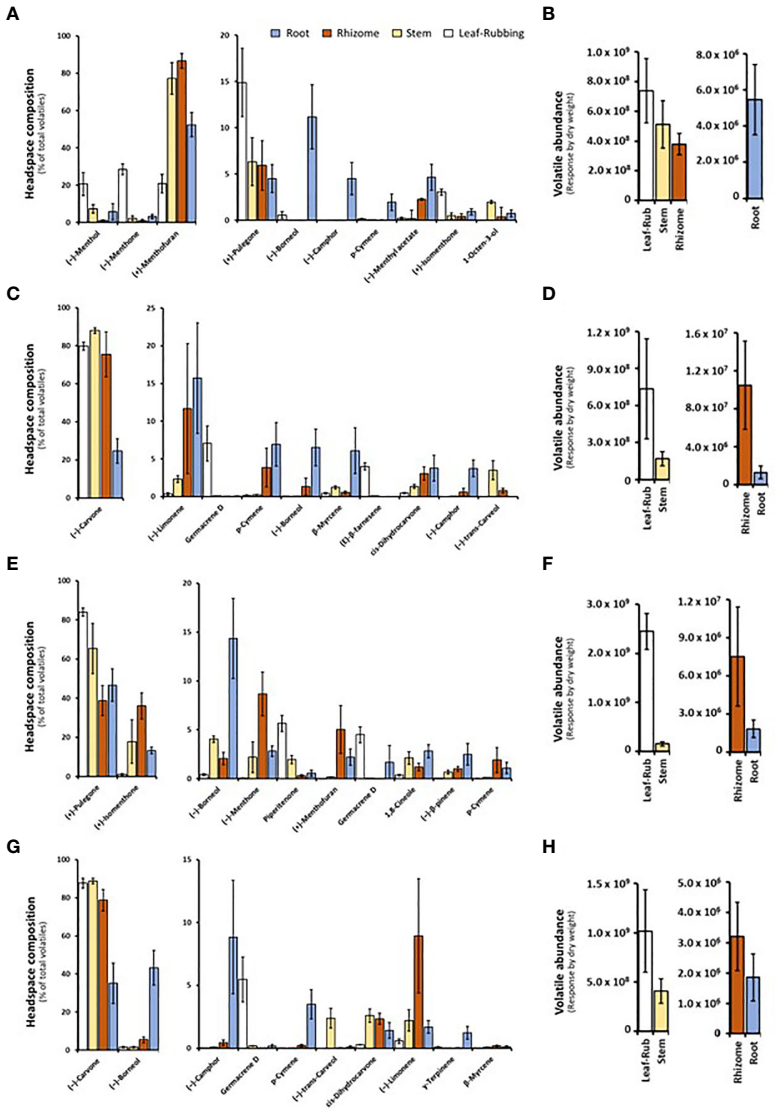



FIGURE 2
Monoterpenoid volatiles released by several organs of greenhouse-grown mint species as determined by SPME–GC–MS analysis (columns represent means of 6 biological replicates, with bars showing the standard error of the mean). **(A)** peppermint; **(C)** spearmint; **(E)** *M. longifolia* CMEN 585; and **(G)** *M. longifolia* CMEN 584. The volatile abundance is shown in **(B)** (peppermint), **(D)** (spearmint), **(F)** (*M. longifolia* CMEN 585), and **(H)** (*M. longifolia* CMEN 584). The color code of columns is as follows: white, leaves; yellow, stems; orange, rhizomes; and blue, roots.

(peppermint and *M. longifolia* CMEN 584) were included in this work. The experiment was performed in climate-controlled growth chambers to ensure pathogen containment, necessitating growing conditions that were substantially different from those for the data sets acquired with greenhouse-grown plants as discussed above (see Materials and Methods for details). The experimental design also had to be modified substantially: i) plants were carefully taken out of their flats and soil shaken off; ii) roots were dipped in either sterilized water (controls) or a solution containing a known concentration of *V. dahliae* inoculum (experimental plants); and iii) plants were re-planted separately in pots containing sterilized soil.

In mock-inoculated control samples, the release of volatiles from peppermint roots was markedly lower than that from rhizomes (roughly 200-fold less in both control and experimental samples) and volatile emission from rhizomes was roughly half of that of stems (Figure 3A and Supplemental Table S4). The emission from peppermint roots, rhizomes and stems was slightly reduced (not statistically significant) in *V. dahliae*-dipped experimental root samples when compared to the corresponding water-dipped controls (Figure 3A). The main constituents of volatile blends emitted by peppermint root controls were (–)-menthol (25% of total volatiles; 37,700 peak area units (PAU) per mg dry weight (DW)), (+)-menthofuran (18%), *p*-cymene (8%), 1,8-cineole (6%), and 1-octen-3-ol (4%) (Figure 3B and Supplemental Table S4). *V. dahliae* inoculation resulted in a slight reduction (not statistically significant) in the quantities of major emitted volatiles [(–)-menthol, (+)-menthofuran, *p*-cymene, and 1,8-cineole], dramatic reductions in the release of (+)-pulegone (25-fold from 3,698 to 139 PAU per mg DW; *P*-Value of 0.02) and (+)-isomenthone (from 549 PAU per mg DW to undetectable levels; *P*-Value of 0.02), and a striking 10-fold increase in the volatilization of (+)-α-terpineol (from 480 to 4,794 PAU per mg DW; *P*-value of 0.03) (Figure 3B and Supplemental Table S4). Peppermint rhizome controls emitted primarily (+)-menthofuran (83%; $2.5 \cdot 10^7$ PAU per mg DW), with (–)-menthol (5%), (–)-menthyl acetate (3%), and 1,8-cineole (2%) as other noteworthy, volatilized constituents (Figure 3B and Supplemental Table S4). As a response to *V. dahliae* treatment, the release of the major constituents (+)-menthofuran and (–)-menthol decreased slightly (not statistically significant), while the emission of the minor component (+)-α-terpineol increased 7-fold (from 461 to 3,453 PAU per mg DW; *P*-value of 0.03). The mixture of volatiles released by peppermint stem controls was dominated by (+)-menthofuran (67%; $4.5 \cdot 10^7$ PAU per mg DW), with (–)-menthol (9%), 1-octen-3-ol (9%), and 1,8-cineole (3%) being other noteworthy constituents (Figure 3B and Supplemental Table S4). No significant changes in volatile bouquets were detected when comparing water-dipped stem controls to the corresponding *V. dahliae*-inoculated samples.

The abundance of volatiles released by mock-inoculated spearmint root controls ($5.2 \cdot 10^4$ PAU per mg DW) was roughly 100-fold lower than that from rhizomes ($5.4 \cdot 10^6$ PAU per mg DW), and the volatile emission from rhizomes was 15-fold lower than that from stems ($8.1 \cdot 10^7$ PAU per mg DW) (Figure 3C and Supplemental Table S4). The overall quantity of volatiles increased slightly in *V. dahliae*-inoculated spearmint roots, while decreasing slightly in rhizomes and stems of experimental plants when compared to controls (not statistically significant) (Figure 3C). Roots of spearmint controls emitted (+)-pinocarvone as the most prominent volatile (26% of total volatiles; 10,519 PAU per mg DW), with *p*-cymene (13%), (–)-borneol (11%), terpinen-4-ol (10%), 1,8-cineole (7%), and γ-terpinene (7%) as other notable monoterpenes (Figure 3D and Supplemental Table S4). Slight increases (not statistically significant) in the emission of several volatiles [(-)-borneol,(+)-pinovarvone, terpinen-4-ol, and 1,8-cineole] were observed as a response to *V. dahliae* inoculation; a non-significant decrease of (-)-carvone was also detected. (–)-Carvone (35% of total volatiles; $2.9 \cdot 10^6$ PAU per mg DW) was the most abundant volatile emitted by spearmint rhizome controls, with (–)-limonene (19%), *cis*-dihydrocarvone (15%), and β-myrcene (7%) being other notable constituents (Figure 3D and Supplemental Table S4). A *V. dahliae* treatment resulted in decreases in the release of *p*-cymene (3.3-fold from 88,424 to 26,991 PAU per mg DW *P*-Value 0.03), *trans*-dihydrocarvone (3.7-fold from 18,981 to 5,077 PAU per mg DW; *P*-Value 0.02), and (+)-pinovarvone (2.7-fold from 11,270 to 4,194 PAU per mg DW *P*-Value 0.02) compared to control samples. In headspace samples of spearmint stems (controls), (–)-carvone was dominant (84% of total volatiles; $7.1 \cdot 10^7$ PAU per mg DW), with (–)-limonene (9%), *trans*-carveol and β-myrcene (both at < 2%) as other noteworthy contributors to volatiles blends(Figure 3D and Supplemental Table S4). A slight reduction (not statistically significant) in the concentrations of these volatiles were detected in stems of *V. dahliae*-inoculated samples (when compared to mock-inoculated controls).

The overall release of volatiles from mock-inoculated roots ($1.8 \cdot 10^5$ PAU per mg DW) of the *M. longifolia* CMEN 585 accession was 780-fold lower than from rhizomes ($1.4 \cdot 10^8$ PAU per mg DW), and volatile emission from rhizomes were 2-fold lower than those of stems ($2.9 \cdot 10^8$ PAU per mg DW) (Figure 3E). A *V. dahliae* treatment did not affect total root and stem volatiles significantly, but there was a 2.8-fold decrease in total volatile emissions from rhizomes when compared to those of corresponding controls (*P*-Value 0.03) (Figure 3E). The volatile blends emitted by *M. longifolia* CMEN 585 roots consisted of primarily (+)-menthofuran (37% of total volatiles; 63,416 PAU per mg DW), (–)-borneol (22%), and α-terpineol (13%), with (+)-pulegone (7%), δ-terpineol (7%), terpinen-4-ol (5%) as additional notable constituents (Figure 3F and Supplemental Table S4). There were no statistically significant differences between *V. dahliae*-inoculated and control root volatiles. The headspace above rhizome samples of *M. longifolia* CMEN 585 control plants was rich in (+)-isomenthone (65%), (+)-pulegone (14%), and (–)-menthone (11%). The quantities of three rhizome volatiles decreased as a response to *V. dahliae* inoculation when compared to controls: (+)-isomenthone was reduced by 3.2-fold (from $9.5 \cdot 10^7$ to $2.9 \cdot 10^7$ PAU per mg DW; *P*-Value 0.01), (–)-menthone was 4.4-fold lower (from $1.6 \cdot 10^7$ to $3.6 \cdot 10^6$ PAU per mg D; *P*-Value 0.03), and the concentration of piperitone was decreased 4.4-fold (from $2.6 \cdot 10^6$ to $5.9 \cdot 10^5$ PAU per mg DW; *P*-Value 0.01). Stems of *M. longifolia* CMEN 585 control samples

10.3389/fpls.2023.1125065

emitted primarily (+)-pulegone (75%) and (+)-isomenthone (15%), with (-)-borneol (3%) and (–)-menthone (1%) being the only other noteworthy constituents (Figure 3F and Supplemental Table S4). A *V. dahliae* treatment resulted in slight increases (not statistically significant) in the emission of (+)-isomenthone, (–)-menthone and (–)-borneol from stems.

The release of volatiles from roots of *M. longifolia* accession CMEN 584 ($2.4 \cdot 10^5$ PAU per mg DW) was 255-fold lower than that from rhizomes ($6.1 \cdot 10^7$ PAU per mg DW), and volatile emission from rhizomes was 5-fold less than from stems ($3.1 \cdot 10^8$ PAU per mg DW) (Figure 3H). The experimental *V. dahliae* treatment caused a slight increase of volatile emissions from rhizomes and a slight decrease of volatile release from stems (not statistically significant), and no changes in roots (Figure 3G and Supplemental Table S4). Roots emitted a volatile blend rich in (–)-borneol (33% of total volatiles; 1,208,008 PAU per mg DW) and

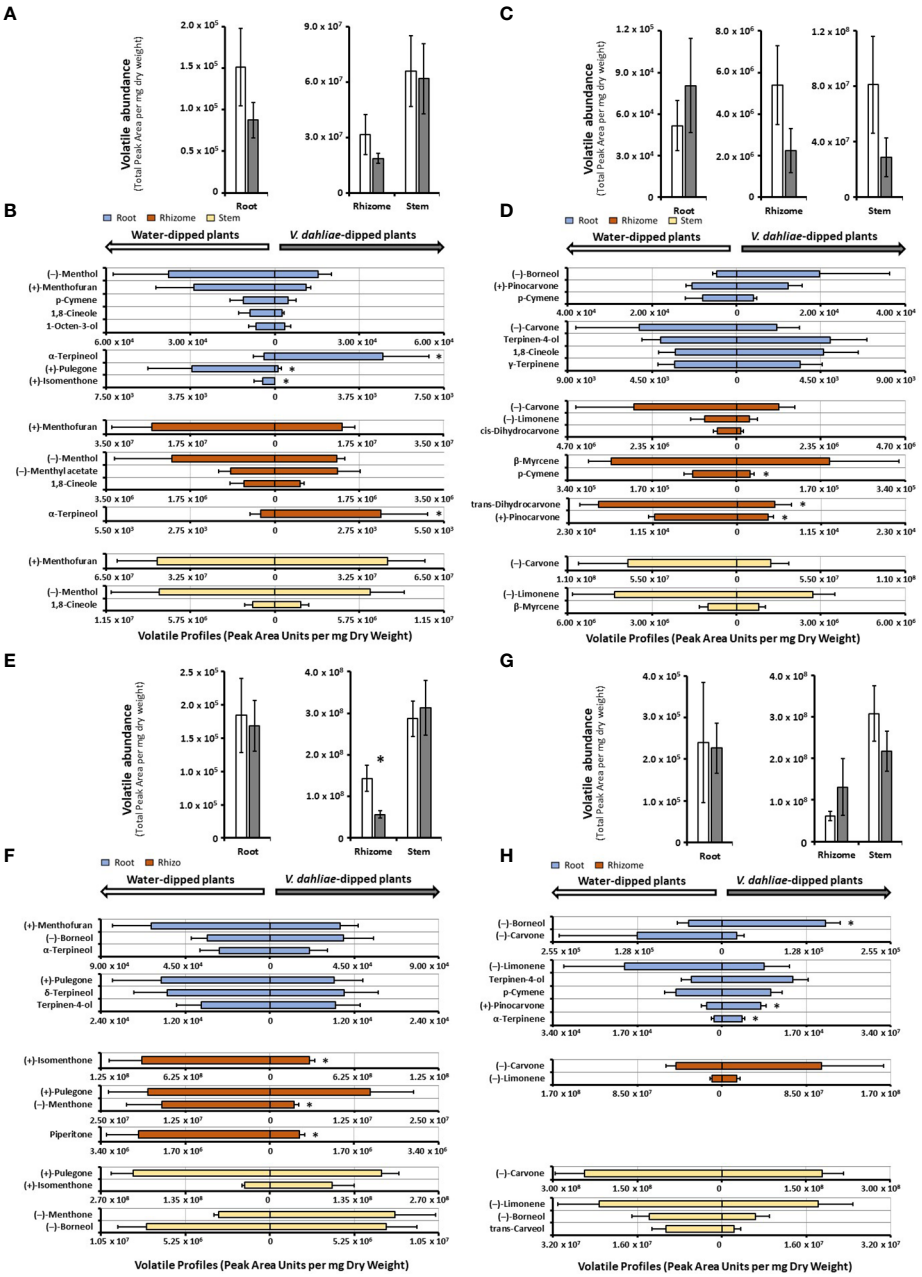



FIGURE 3
Monoterpenoid volatile profiles, as determined by SPME–GC–MS analysis, released by several organs of mint species maintained in growth chambers and prepared for *V. dahliae* inoculation (columns represent means of 4 – 6 biological replicates, with bars showing the standard error of the mean). The total volatile abundance of volatiles emitted by each organ is shown in panels **(A)** (peppermint), **(C)** (spearmint), **(E)** (M. longifolia CMEN 585), and **(G)** (M. longifolia CMEN 584, with white columns for controls and gray columns for experimental samples. The left panel of each butterfly diagram shows data for mock-inoculated controls, while the right panel has data for plants whose roots were inoculated with *V. dahliae*, with **(B)**, peppermint; **(D)**, spearmint; **(F)**, *M. longifolia* CMEN 585; and **(H)**, *M. longifolia* CMEN 584. The color code of columns is as follows: yellow, stems; orange, rhizomes; and blue, roots. An asterisk represent *P*-value < 0.5.

(–)-carvone (24%), with (–)-limonene (9%), *p*-cymene (6%), terpinen-4-ol (4%), and *cis*-dihydrocarvone (4%), (+)-pinocarvone (2%), and α-terpinene (1%). A *V. dahliae* treatment-induced increase in the emission of (-)-borneol (3.1-fold; *P*-Value 0.02), (+)-pinocarvone (2.6-fold; *P*-Value 0.03), and α-terpinene (2.5-fold; *P*-Value 0.02) was observed. The dominant monoterpenoid emitted from rhizomes of control plants was (–)-carvone (75%), with only (–)-limonene (18%) being another substantial contributor to the profile (all other constituents at < 1%). *V. dahliae* inoculation led to slight increases (not statistically significant) in (–)-carvone and (–)-limonene release. Stems of controls emitted high quantities of (–)-carvone (80%; $2.4 \cdot 10^8$ PAU per mg DW) and lower amounts of (–)-limonene (8%), (-)-borneol (4%), and *trans*-carveol (3%) (Figure 3G and Supplemental Table S4). Slight decreases (not statistically significant) in the quantities of these metabolites were observed as a response to *V. dahliae* inoculation.

## Cloning and functional characterization of candidate MTSs involved in mint terpenoid volatile emissions

Despite the large body of published work on the biosynthesis of functionalized monoterpenoids in mint (Lange and Srividya, 2019), surprisingly, only two MTSs have been functionally characterized: (–)-limonene synthase from spearmint (Colby et al., 1993) and (–)-β-pinene synthase from *M. longifolia* CMEN 585 (Kim et al., 2022). These MTSs are involved in the biosynthesis of *p*-menthane monoterpenoids, such as (–)-menthol of peppermint, (–)-carvone of spearmint, and (+)-pulegone of *M. longifolia* CMEN 585 (all derived from (–)-limonene) and (+)-pinocarvone (derived from (–)-β-pinene) (Figure 1). However, MTSs must also exist to produce other volatiles detected in our experiments, including α-terpineol, 1,8-cineole, *p*-cymene (through γ-terpinene as precursor), as well as (–)-borneol and (–)-camphor (both through (–)-bornyl diphosphate as precursor) (Figure 1). To identify candidate genes coding for these MTSs, we had taken advantage of the recently published chromosome-level assembly of *M. longifolia* accession CMEN 585 and performed subsequent bioinformatic analyses (Chen et al., 2021; Vining et al., 2022). Translated peptide sequences of candidate genes with high identity to previously characterized MTSs across the Lamiaceae (mint family) were then subjected to a phylogenetic analysis. *M. longifolia* CMEN 585 candidate Ml_15999 occupied a clade with previously characterized α-terpineol synthases of Cretan thyme (*Thymus caespititius* Brot.) (Lima et al., 2013) (referred to as ATERS1 and ATERS2 in Figure 4). Candidate Ml_14850 clustered with γ-terpinene synthases of the genera *Thymus* and *Origanum* (Mendes et al., 2014; Tohidi et al., 2020) (labeled as GTS4 and GTS, respectively, in Figure 4). Candidate Ml_38055 formed a clade with previously characterized bornyl diphosphate synthases of *Lavandula angustifolia* Mill. and *Phyla dulcis* (Trevir.) Moldenke (Despinasse et al., 2017; Hurd et al., 2017) (indicated as BPPS in Figure 4). Candidate Mp_12413 from peppermint clustered with

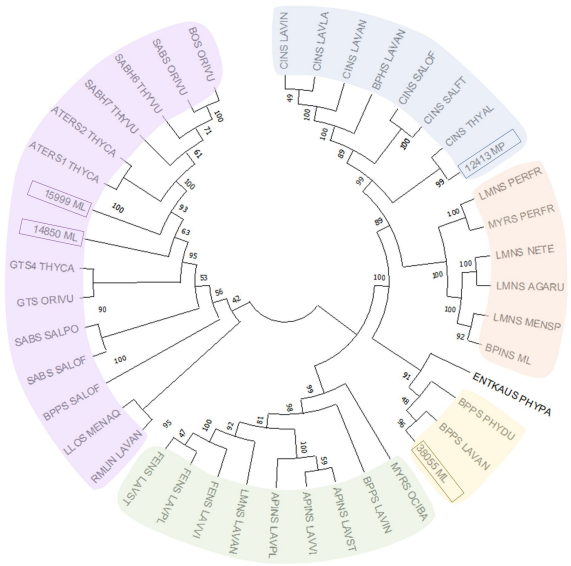



FIGURE 4
Evolutionary analysis of characterized MTSs of the Lamiaceae by the Maximum Likelihood method. Methodological details are provided in Materials and Methods, while sequences and abbreviations for MTSs are given in Supplemental Table S1.

1,8-cineole synthases (labeled as CINS in Figure 4) of several species of the Lamiaceae (Wise et al., 1998; Kampranis et al., 2007; Demissie et al., 2012; Ali et al., 2022).

Each of the four candidate cDNAs was cloned into an expression vector (pET 28b), the construct transformed into *Escherichia coli*, transgene expression induced, the recombinant protein purified, a functional assay performed, and products were analyzed using a GC equipped with a flame ionization detector (as described in Srividya et al., 2016). The dominant product (63%) of Ml_15999 from *M. longifolia* was (+)-α-terpineol (> 95 enantiomeric excess (ee)), with (–)-linalool and (+)-linalool as prominent side-products (each roughly 14%) (Figures 5A–C). Ml_14850 of *M. longifolia* generated γ-terpinene as primary product (76%), with (–)-β-pinene (9.3%) and (–)-thujene (10%) as notable byproducts (Figures 6A–C). The principal product of Ml_38055 was (–)-bornyl diphosphate (56%; detected as (-)-borneol after hydrolysis at >99% ee), with (–)-camphene (18.5%), (–)-α-pinene (10%), and (–)-β-pinene (5%) as significant side-products (Figures 7A–C). 1,8-Cineole was detected in the volatile blends emitted by several peppermint samples (Figure 3), and we thus searched previously published peppermint transcriptome data sets for additional MTS candidate genes. One of these candidates, Mp_12413, formed 1,8-cineole as main product (72%), with (+)/(–)-α-pinene (total of 11%), (+)/(–)-β -pinene (total of 11%) and (+)-sabinene (6%) as the most prominent side products (Figures 8A–C). All MTS candidates were assayed with neryl diphosphate, (*Z,Z*)-farnesyl diphosphate, and (*E,E*)-farnesyl diphosphate as alternative substrates, but none of these reactions were productive (Figures 5C, 6C, 7C, 8C).

# Discussion

## Unique blends of volatiles are emitted by different mint organs

Most volatile analyses of mint species have been performed with oils obtained by steam distillation of above-ground material (Lawrence, 2006). Since a research-scale hydrodistillation that uses a carrier solvent is only suitable for analytes that are accumulated at fairly high levels in a tissue or organ of interest (Likens and Nickerson, 1964), we employed volatile capture by SPME as method to assess both above-ground and below-ground samples. These analyses indicated the presence of monoterpenoids that had previously been reported only as trace constituents with low annotation confidence. For example, *p*-cymene has occasionally (but not consistently) been reported as a 0.1% constituent of steam-distilled peppermint oil (Lawrence, 2008). We observed similar levels in both hydrodistilled and headspace SPME analyses of peppermint leaves, while the headspace above root samples of all accessions had a much higher proportion of this volatile (1.9% in peppermint, 6.9% in spearmint, 1.1% in *M. longifolia* CMEN 585, and 3.5% in *M. longifolia* CMEN 584). While trace quantities of (–)-borneol and (–)-camphor were observed in some leaf samples (but they were mostly undetectable), these volatiles were present at higher levels in volatile blends emitted from roots of peppermint (11 and 5%, respectively), spearmint (7 and 4%, respectively), and *M. longifolia* CMEN 584 (43 and 9%, respectively), while *M. longifolia* CMEN 585 roots emitted (–)-borneol (14%), but not (–)-camphor. To the best of our knowledge, neither borneol nor camphor had been described as mint constituents before. (+)-Pinocarvone, which has also not been reported as a mint oil constituent before, was detected in the headspace above spearmint root samples (6% of total volatiles).

The bulk of the mint leaf volatiles is synthesized and accumulated in specialized anatomical structures called peltate glandular trichomes, but such structures are not present on stems, rhizomes, or roots (Lange, 2015b). The formation of volatiles by all organs investigated as part of the current study may thus come as a surprise. However, it should be noted that the pathways for the formation of terpenoids and other volatiles are not restricted to glandular trichomes. The biosynthesis of volatiles in flowers often occurs in epidermal cells (Effmert et al., 2006). Volatiles can be produced throughout the leaf, as illustrated by the release of sesquiterpenes at the site of damage caused by herbivore feeding (Köllner et al., 2013). Terpenes can also be formed in many types of roots cells, as demonstrated for the formation of 1,8-cineole in *Arabidopsis thaliana* L. (Chen et al., 2004). We showed earlier that genes involved in monoterpene biosynthesis are responsive to treatment of *M. longifolia* with *V. dahliae* (Vining and Pandelova, 2022). Having demonstrated the formation of terpene volatiles by several mint organs in the present work, it will be interesting to investigate the localization of the relevant pathways outside glandular trichomes in future studies.

While the concentrations of volatiles emitted by mint roots were substantially lower than those released by stems and rhizomes, the discovery of unique blends formed in roots increases the known monoterpenoid repertoire of mint. This observation also begs the question if mint root volatiles might play roles in plant-plant and/or plant-microbe interactions, akin to evidence gathered as part of soil biome studies with other plants (Junker and Tholl, 2013; Peñuelas et al., 2014; Delory et al., 2016). The experiments discussed in the

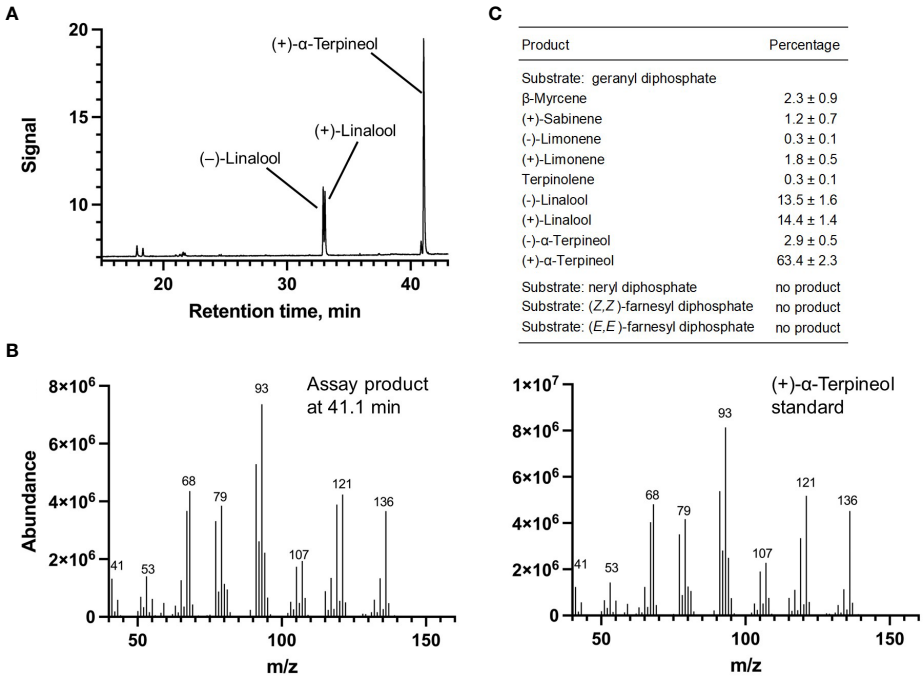



FIGURE 5
Characterization of *M. longifolia* CMEN 585 candidate Ml_15999. **(A)** GC chromatogram with annotation for major peaks; **(B)** mass spectrum of the main enzymatically generated product and the matching authentic standard [(+)-α-terpineol); and **(C)** tabulated product profile.

upcoming paragraph begin to address this question, but more work will be necessary in the future. The cloning of MTSs from mint that collectively generate the major root volatiles detected in our experiments (more details below) is an important first step to provide a better understanding of the molecular basis for volatile formation in mint organs other than leaves.

## Expansion of mint volatile portfolio through environmental and experimental conditions

A follow-up experiment was performed to assess if *V. dahliae* inoculation would induce the formation of volatiles not seen in control samples. Our design involved the removal of plants from sterile soil, inoculation of roots with a suspension containing fungal conidia (or water for control samples), and then reestablishing plants in new sterile soil. Plants were maintained in growth chambers to ensure fungal containment. Several differences were observed when comparing volatiles blends emitted from roots of mint plants maintained in greenhouses versus those that were kept in growth chambers and prepared for *V. dahliae* inoculation, in some cases even in control plants. For example, (–)-borneol and (–)-camphor were among the more abundant volatiles emitted by peppermint roots of greenhouse-grown plants, while *p*-cymene and 1,8-cineole were released at relatively high levels by roots of growth chamber-maintained plants. (–)-Carvone and (–)-limonene were the most prominent volatiles detected in the headspace above spearmint roots of greenhouse-grown plants, whereas (–)-borneol, (+)-pinocarvone, and *p*-cymene were the signature volatiles emitted by roots of growth-chamber-maintained plants. The most abundant volatiles released by roots of greenhouse-grown *M. longifolia* CMEN 585 were (+)-pulegone and (+)-isomenthone, while volatile blends emitted by roots of growth chamber-maintained plants were enriched in (+)-menthofuran, (–)-borneol, and α-terpineol. (–)-Camphor was among the more abundant volatiles in the headspace above roots from greenhouse-grown *M. longifolia* CMEN 584 plants, while (–)-limonene was released at much higher levels from roots of growth chamber-maintained plants. Interestingly, the overall volatile profiles emitted by rhizomes and stems were more comparable across all species and growth conditions investigated here, indicating that differences across root samples are more likely to result from the temporary removal of plants from the soil as an inevitable step for the experiment performed in growth chambers. Evaluating the differences in root volatile profiles in greenhouse-grown and growth chamber-maintained plants was a valuable exercise in the context of our project goal to evaluate terpenoid chemical diversity.

Statistically significant differences in volatile quantities released by plants whose roots were dipped in a *V. dahliae*-spore suspension (experimental) or dipped in water (controls) were evident in roots of mint species that are susceptible to *Verticillium* wilt disease (peppermint and *M. longifolia* CMEN 584), while rhizome volatiles were affected in species that show resistance to the disease (spearmint and *M. longifolia* CMEN 585). Our observations are certainly interesting, as volatiles could play roles in the communication between mint roots and the soil-borne *V. dahliae* pathogen. For example, soil treatments with terpene-containing formulations have been shown to reduce *V. dahliae* growth (Wei et al., 2016). Contrastingly, it has been shown that the

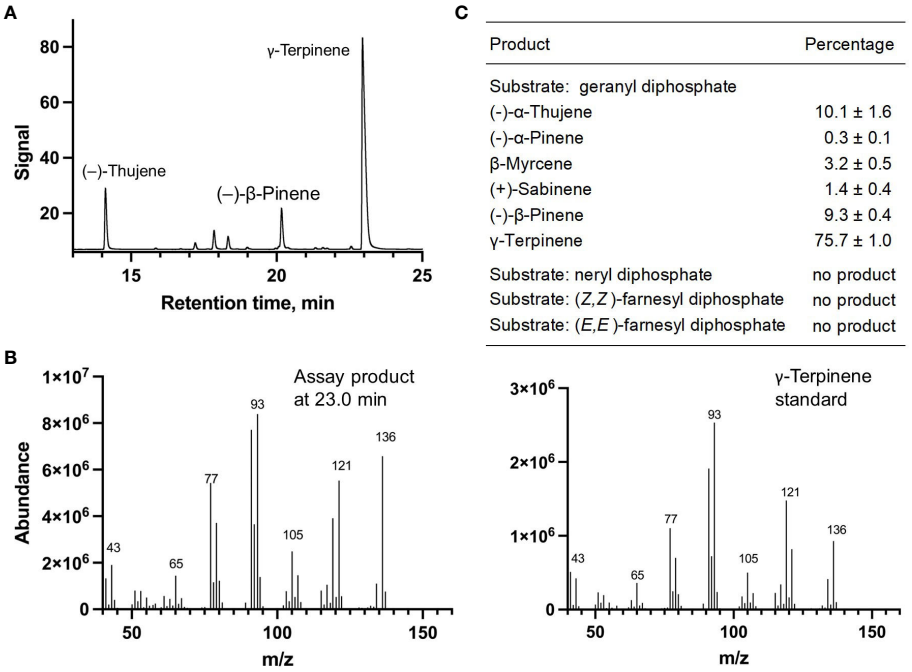



FIGURE 6
Characterization of *M. longifolia* CMEN 585 candidate Ml_14850. **(A)** GC chromatogram with annotation for major peaks; **(B)** mass spectrum of the main enzymatically generated product and the matching authentic standard (γ-terpinene); and **(C)** tabulated product profile.

overexpression of genes involved in monoterpene biosynthesis in transgenic *Arabidopsis thaliana* plants leads to an enhanced colonization by *V. longisporum*, a species closely related to *V. dahliae* (Roos et al., 2015). It is thus conceivable that specific volatile blends have functions in susceptibility or resistance to *Verticillium* wilt disease, which is an interesting hypothesis for further testing, but beyond the scope of the current study. In the present work, our emphasis was primarily to assess if experimental conditions could be used to increase the diversity of volatiles released by mint. While the quantities of volatile blends emitted by mint roots were low, the observed chemical diversity was surprising, and we thus embarked on efforts to characterize the genes underlying volatile production in mint, as discussed in the next paragraph.

## Mint MTSs are expressed in multiple organs, thus accounting for emitted volatiles

A (–)-limonene-forming MTS from spearmint (LMNS) had been cloned and characterized decades ago, with orthologs of peppermint and *M. longifolia* CMEN 585 having been characterized as well (Colby et al., 1993; Ahkami et al., 2015; Chen et al., 2021; Vining et al., 2022). LMNS generates the precursor for the main pathways active in peppermint, spearmint, *M. longifolia* CMEN 585, and *M. longifolia* CMEN 584, accounting for the formation of (–)-limonene, (–)-menthol, (–)-menthone, (–)-menthyl acetate, (+)-isomenthone, (+)-menthofuran, (+)-pulegone, piperitenone, (–)-*trans*-carveol, (–)-carvone, and *cis*-dihydrocarvone (Figures 1, 2). LMNS is expressed at high levels in leaf glandular trichomes of all mint species investigated thus far (Lange et al., 2000; Ahkami et al., 2015; Vining et al., 2022). In *M. longifolia* CMEN 585, LMNS expression levels (represented by two genes that code for different isoforms (Ml_17628 and Ml_17636); Vining et al., 2022) were low in roots and high in stems (Supplemental Table S5; Vining and Pandelova, 2022), consistent with the volatile quantities emitted from these organs as determined in the present study. The same gene expression patterns and volatile emission characteristics were observed for *M. longifolia* CMEN 584 (Supplemental Tables S4, S5).

We recently reported on the characterization of an MTS from *M. longifolia* CMEN 585 that generates (–)-β-pinene as its main product (Kim et al., 2022), which is an intermediate in the biosynthesis of (+)-pinocarvone (Figure 1). Both volatiles were emitted by stems but not from roots, which correlated well with gene expression patterns (the transcript for (–)-β-pinene synthase (Ml_39080) was expressed at relatively high levels in stems but barely detectable in roots of *M. longifolia* CMEN 585 and CMEN 584; Vining and Pandelova, 2022) (Supplemental Tables S4, S5). The gene coding for (–)-bornyl diphosphate synthase (Ml_38055) was expressed at very low levels in roots and high levels in stems, which correlated with the release of borneol (high concentration in stem volatiles for both *M. longifolia* CMEN 585 and CMEN 584; Vining and Pandelova (2022); low abundance in volatile blends emitted by roots) (Supplemental Tables S4, S5). Transcript levels for γ-terpinene synthase (Ml_14850) were previously reported to be low in both roots and stems of *M. longifolia* CMEN 585 and CMEN

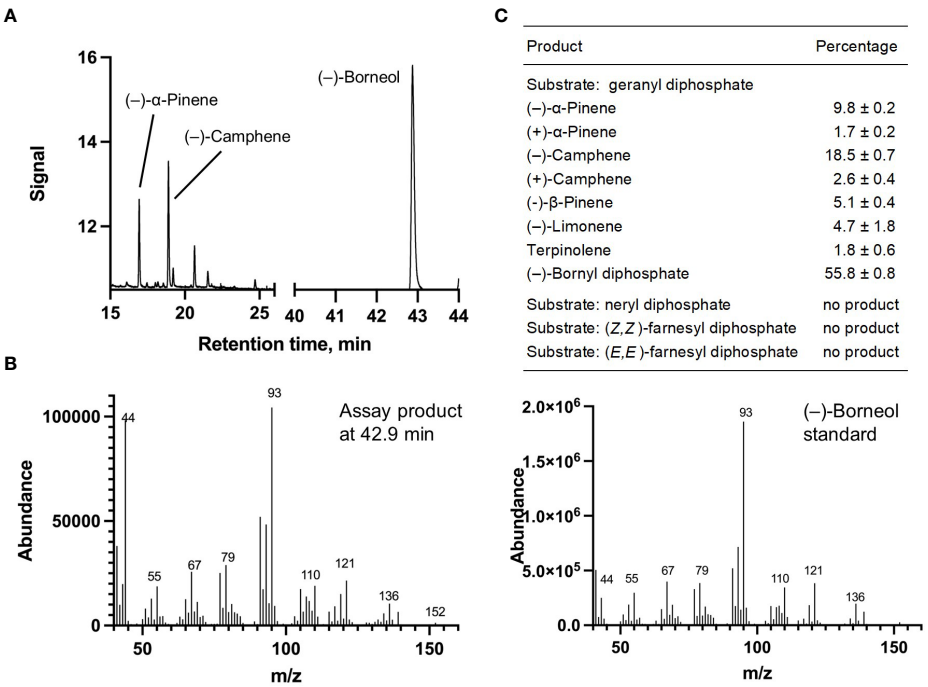



FIGURE 7
Characterization of *M. longifolia* CMEN 585 candidate Ml_38055. **(A)** GC chromatogram with annotation for major peaks; **(B)** mass spectrum of the main enzymatically generated product (following hydrolysis) and the matching authentic standard [(–)-borneol, which is hydrolytically released from bornyl diphosphate]; and **(C)** tabulated product profile.

584 (Vining and Pandelova, 2022), which correlated with low levels of the corresponding volatiles (γ-terpinene and p-cymene) in the same samples, as reported in the current study (Supplemental Tables S4, S5). The α-terpineol synthase gene (Ml_15999) was demonstrated earlier to be expressed at high levels in roots of *M. longifolia* CMEN 585 (Vining and Pandelova, 2022), where α-terpineol is a major volatile; expression levels of this gene in stems of *M. longifolia* CMEN 585 and both roots and stems of *M. longifolia* CMEN 584 were significantly lower (Vining and Pandelova, 2022), coinciding with a lower emission of α-terpineol from the same samples (Supplemental Tables S4, S5). In summary, there was a tight correlation with MTS gene expression patterns and the volatiles generated by these enzymes across two mint accessions and multiple organs.

## Comparison of putative active site residues across functionally diverged MTSs

(+)-α-Terpineol, which is released prominently by roots of peppermint and *M. longifolia* CMEN 585 when maintained in growth chambers while being prepared for *V. dahliae* inoculation (Figures 1, 3). A phylogenetic analysis placed (+)-α-terpineol synthase (ML_15999) in the same clade as two previously characterized α-terpineol synthases (ATERS1 and ATERS2 in Figure 4) of Cretan thyme (Lima et al., 2013). The *M. longifolia* candidate Ml_15999 is the first α-terpineol synthase for which the product stereochemistry has been established (with an enantiomeric excess of 95.4% for the (+)- over the (–)-enantiomer). Since (+)-linalool (14.4%) and (–)-linalool (13.5%) are other notable products, ML_15999 generates almost exclusively monoterpenoid alcohols (94.2%). 1,8-Cineole was detected as a volatile emitted by greenhouse-grown peppermint and *M. longifolia* CMEN 585 but also by roots of *M. longifolia* CMEN 585 maintained in growth chambers (Figures 1, 3), and, in the current study, we identified and functionally characterized a 1,8-cineole synthase (Mp_12413). This enzyme clustered with previously characterized 1,8-cineole synthases of the Lamiaceae (Wise et al., 1998; Kampranis et al., 2007; Demissie et al., 2012). Interestingly, while 1,8-cineole synthases and α-terpineol synthases of the Solanaceae were demonstrated to be very closely related (Fähnrich et al., 2011), those of the Lamiaceae occupy more distant branches of the phylogenetic tree (Figure 4). Signature sequence motifs also differed between members of these enzyme classes: variable region 1 was I/T-I/L-L-I-T/S in α-terpineol synthases and N-A-L-V-T in 1,8-cineole synthases; the metal ion-binding DDxxD motif was D-D-V-Y-D in α-terpineol synthases and D-D/E-V/I-F/Y-D in 1,8-cineole synthases; and variable region 2 was S-I-G/A-S-L-T/P-I in α-terpineol synthases and S-I-G-G-L/V/I-P-I in 1,8-cineole synthases (Supplemental Figure S1).

A gene coding for a γ-terpinene synthase from *M. longifolia* CMEN 585 was identified and characterized as part of the present work (ML_14850), which is likely responsible for the accumulation of γ-terpinene as a constituent of root volatile blends emitted by *M. longifolia* CMEN 584 and could also serve to provide the precursor for *p*-cymene formation in all species investigated here (Figures 1 and 2). Based on a phylogenetic analysis, the *M. lognifolia* γ-terpinene synthase was closely related to previously characterized γ-terpinene synthases (GTSs) from *Thymus caespititius* and *Ocimum vulgare* (Crocoll et al., 2010; Mendes et al., 2014),

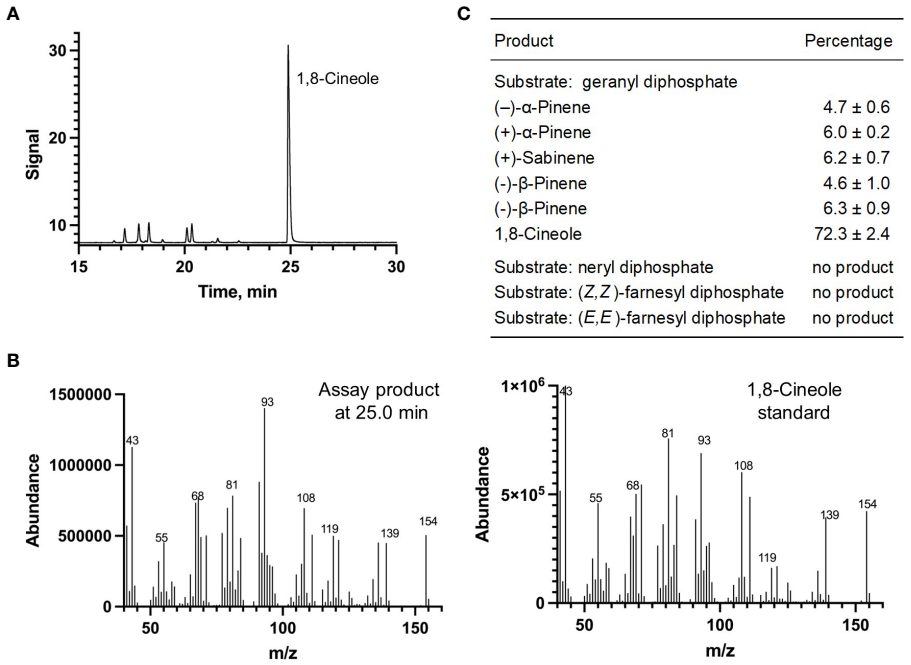



FIGURE 8
Characterization of *M. longifolia* CMEN 585 candidate Mp_12413. **(A)** GC chromatogram with annotation for major peaks; **(B)** mass spectrum of the main enzymatically generated product and the matching authentic standard (1,8-cineole); and **(C)** tabulated product profile.

positioned in a sister clade to α-terpineol synthases (Figure 4). Signature sequence motifs were very similar across members of these enzyme classes but there were also some notable differences: variable region 1 was I/T-I/L-L-I-T/S in α-terpineol synthases and I-T-F-V-T in γ-terpinene synthases; the metal ion-binding DDxxD motif was D-D-V-Y-D in both; and variable region 2 was S-I-G/A-S-L-T/P-I in α-terpineol synthases and S-I-S-S-P-T-I in γ-terpinene synthases. It will now be interesting to investigate the structure-function relationships underlying these different, but closely related, activities for the formation of α-terpineol, γ-terpinene, and 1,8-cineole (Supplemental Figure S2).

(–)-Borneol and (–)-camphor were detected in volatiles blends released by all species investigated here, and the newly discovered MTS from *M. longifolia* accession CMEN 585 catalyzing the formation of (–)-bornyl diphosphate (ML_38055) would provide the precursor for these volatiles (Figures 1–3). This enzyme clusters with characterized bornyl diphosphate synthases (BPPSs) of *Lavandula angustifolia* and *Phyla dulcis* for which the product stereochemistry had not been assessed (Despinasse et al., 2017; Hurd et al., 2017). It is interesting to note that ML_38055 is positioned on a different branch than *Salvia officinalis* (+)-bornyl diphosphate synthase (Wise et al., 1998) (Figure 4). Signature sequence motifs differed between members of this enzyme class that generate the (+)-isomer and those that cluster with the (–)-isomer-releasing ML_38055: conserved region 1 was I-I-V-L-A in (+)-bornyl diphosphate synthase but L-F-V/I-L/F-I in those that cluster with (–)-bornyl diphosphate synthase; the metal ion-binding DDxxD motif was D-D-I-Y/F/S-D across all bornyl diphosphate synthases; and conserved region 2 was S-V-A-S-P-A-I in but S-I-S-A/T-H/P-L/T-I in those that cluster with (–)-bornyl diphosphate synthase (Supplemental Figure S3). Follow-up work beyond the scope of the current study can now be undertaken to investigate how these sequence differences affect the stereochemistry of the product.

## Conclusions

The present work demonstrates that unique blends of volatiles are emitted from roots, rhizomes and stems of several mint species, organs that had not previously been investigated in this context. Above-ground material of members of the mint family is commercially distilled to extract essential oils, which are then formulated into a myriad of consumer products. Most of the research aimed at characterizing the processes involved in the formation of terpenoid oil constituents has focused on leaves. We now demonstrate, by investigating three mint species, peppermint (*Mentha* × *piperita* L.), spearmint (*Mentha spicata* L.) and horsemint (*Mentha longifolia* (L.) Huds.; accessions CMEN 585 and CMEN 584), that other organs – namely stems, rhizomes and roots – also emit volatiles and that the terpenoid volatile composition of these organs can vary substantially from that of leaves, supporting the notion that substantial, currently underappreciated, chemical diversity exists. Differences in volatile quantities released by plants whose roots were dipped in a *V. dahliae*-spore suspension (experimental) or dipped in water (controls) were evident: in mint species that are susceptible to *Verticillium* wilt disease (peppermint and *M. longifolia* CMEN 584) roots showed increases in the emission of some volatiles, while the experimental treatment decreased the release of certain volatiles from rhizomes of species that show resistance to the disease (spearmint and *M. longifolia* CMEN 585). In the present study, we then identified and functionally characterized four MTSs whose expression was closely correlated with the emission of (+)-α-terpineol, 1,8- cineole, γ-terpinene, and (–)-bornyl diphosphate. In conjunction with previously described MTSs that catalyze the formation of (–)-β-pinene and (–)-limonene, the product profiles of the MTSs identified here can explain the generation of all major monoterpene skeletons represented in the volatiles released by different mint organs.

## Accession numbers

Nucleotide sequences of genes characterized as part of this study were deposited in GenBank and received the accession numbers: OP784758, (–)-bornyl diphosphate synthase (*M. longifolia*); OP784759, 1,8-cineole synthase (*M. piperita*); OP784760, (+)-α-terpineol synthase (*M. longifolia*); and OP784761, γ-terpinene synthase (*M. longifolia*).

## Data availability statement

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

## Author contributions

BL designed the research; AP, IL, LB, and NS performed the experiments; IP and KV provided training for the *V. dahliae* inoculation experiments; AP, BL, IL, LB, and NS analyzed the data; BL and MW served as co-mentors for LB; BL wrote the manuscript, with contributions from all authors. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by a grant from the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, U.S. Department of Energy (grant no. DE–SC0001553 to BL and NS). The Mint Industry Research Council provided funds for the methods development phase of the volatile analyses.

## Acknowledgments

The authors would like to thank Julie Thayer and Devon Thrasher for maintaining plants in the greenhouse. We are also

indebted to Prof. Rikard Unelius (Linnaeus University, Kalmar, Sweden) for gifts of pinocarvone enantiomers and Vigon International (East Stroudsburg, PA, USA) for several additional monoterpene standards.

## Conflict of interest

BL has financial interests in Dewey Scientific LLC, a biotechnology company focused on cannabis breeding, which is unrelated to the studies presented here.

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1125065/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Alignment of sequences of characterized α-terpineol synthases and 1,8-cineole synthases of the Lamiaceae.

SUPPLEMENTARY FIGURE 2
Alignment of sequences of characterized α-terpineol synthases and γ-terpinene synthases of the Lamiaceae.

SUPPLEMENTARY FIGURE 3
Alignment of sequences of characterized bornyl diphosphate synthases of the Lamiaceae.

SUPPLEMENTARY TABLE 1
Accession numbers and amino acid sequences of terpene synthases included in a phylogenetic analysis presented in Figure 4.

SUPPLEMENTARY TABLE 2
Complete terpenoid profiles for the data sets included in Table 1.

SUPPLEMENTARY TABLE 3
Complete terpenoid profiles for the data sets included in Table 2.

SUPPLEMENTARY TABLE 4
Complete terpenoid profiles for the data sets included in Table 3.

## References

Adams, R. P. (2007). *Identification of essential oil components by gas chromatography/mass spectrometry* (Carol Stream, IL: Allured publishing corporation).

Ahkami, A., Johnson, S. R., Srividya, N., and Lange, B. M. (2015). Multiple levels of regulation determine monoterpenoid essential oil compositional variation in the mint family. *Mol. Plant* 8, 188–191. doi: 10.1016/j.molp.2014.11.009

Ali, M., Alshehri, D., Alkhaibari, A. M., Elhalem, N. A., and Darwish, D. B. E. (2022). Cloning and characterization of 1,8-cineole synthase (SgCINS) gene from the leaves of salvia guaranitica plant. *Front. Plant Sci.* 13, 869432–869432. doi: 10.3389/fpls.2022.869432

Amelunxen, F. (1965). Elektronenmikroskopische untersuchungen den den druesenschuppen von mentha piperita l. *Planta Med.* 13, 457–473. doi: 10.1055/s-0028-1100141

Amelunxen, F., Wahlig, T., and Arbeiter, H. (1969). Über den nachweis des ätherischen Öls in isolierten drüsenhaaren und drüsenschuppen von mentha piperita l. *Z Pflanzenphysiol* 61, 68–72.

Arango, M. C., Ruscitti, M., Ronco, M. G., and Beltrano, J. (2012). Mycorrhizal fungi inoculation and phosphorus fertilizer on growth, essential oil production and nutrient uptake in peppermint (Mentha piperita l.). *Rev. Bras. Plantas Med.* 14, 692–699. doi: 10.1590/S1516-05722012000400018

Bharti, N., Baghel, S., Barnawal, D., Yadav, A., and Kalra, A. (2013). The greater effectiveness of glomus mosseae and glomus intraradices in improving productivity, oil content and tolerance of salt-stressed menthol mint (Mentha arvensis). *J. Sci. Food Agricult* 93, 2154–2161. doi: 10.1002/jsfa.6021

Chen, Z., Vining, K. J., Qi, X., Yu, X., Zheng, Y., Liu, Z., et al. (2021). Genome-wide analysis of terpene synthase gene family in mentha longifolia and catalytic activity analysis of a single terpene synthase. *Genes* 12, 518. doi: 10.3390/genes12040518

Chen, F., Ro, D. K., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E., and Tholl, D. (2004). Characterization of a root-specific Arabidopsis terpene synthase responsible for the formation of the volatile monoterpene 1, 8-cineole. *Plant Physio.* 135(4), 1956–1966.

Colby, S. M., Alonso, W. R., Katahira, E. J., McGarvey, D. J., and Croteau, R. (1993). 4S-limonene synthase from the oil glands of spearmint (Mentha spicata). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. *J. Biol. Chem.* 268, 23016–23024. doi: 10.1016/S0021-9258(19)49419-2

Cordero, C., Zebelo, S. A., Gnavi, G., Griglione, A., Bicchi, C., Maffei, M. E., et al. (2012). HS-SPME-GC×GC-qMS volatile metabolite profiling of chrysolina herbacea frass and mentha spp. leaves. *Anal. Bioanal Chem.* 402, 1941–1952. doi: 10.1007/s00216-011-5600-4

Crocoll, C., Asbach, J., Novak, J., Gershenzon, J., and Degenhardt, J. (2010). Terpene synthases of oregano (Origanum vulgare l.) and their roles in the pathway and regulation of terpene biosynthesis. *Plant Mol. Biol.* 73, 587–603. doi: 10.1007/s11103-010-9636-1

Daayf, F. (2015). Verticillium wilts in crop plants: Pathogen invasion and host defence responses. *Can. J. Plant Pathol.* 37, 8–20. doi: 10.1080/07060661.2014.989908

Degenhardt, J., Kollner, T. G., and Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 70, 1621–1637. doi: 10.1016/j.phytochem.2009.07.030

Delory, B. M., Delaplace, P., Fauconnier, M.-L., and Du Jardin, P. (2016). Root-emitted volatile organic compounds: Can they mediate belowground plant-plant interactions? *Plant Soil* 402, 1–26. doi: 10.1007/s11104-016-2823-3

del Rosario Cappellari, L., Santoro, M. V., Reinoso, H., Travaglia, C., Giordano, W., and Banchio, E. (2015). Anatomical, morphological, and phytochemical effects of inoculation with plant growth-promoting rhizobacteria on peppermint (Mentha piperita). *J. Chem. Ecol.* 41, 149–158. doi: 10.1007/s10886-015-0549-y

Demissie, Z. A., Cella, M. A., Sarker, L. S., Thompson, T. J., Rheault, M. R., and Mahmoud, S. S. (2012). Cloning, functional characterization and genomic organization of 1, 8-cineole synthases from lavandula. *Plant Mol. Biol.* 79, 393–411. doi: 10.1007/s11103-012-9920-3

Denny, E. F. K., and Lawrence, B. M. (2007). "The distillation of mint oils: History, current theory, and practice," in *Mint: The genus mentha*. Ed. B. M. Lawrence (Boca Raton, FL, USA: CRC Press, Taylor & Francis Group), 165–184.

Despinasse, Y., Fiorucci, S., Antonczak, S., Moja, S., Bony, A., Nicolè, F., et al. (2017). Bornyl-diphosphate synthase from lavandula angustifolia: A major monoterpene synthase involved in essential oil quality. *Phytochemistry* 137, 24–33. doi: 10.1016/j.phytochem.2017.01.015

Dung, J. K. (2020). Verticillium wilt of mint in the united states of America. *Plants* 9, 1602. doi: 10.3390/plants9111602

Effmert, U., Buss, D., Rohrbeck, D., and Piechulla, B. (2006). Localization of the synthesis and emission of scent compounds within the flower. *Biol. of Floral Scent* 105–124.

Fähnrich, A., Krause, K., and Piechulla, B. (2011). Product variability of the 'cineole cassette' monoterpene synthases of related nicotiana species. *Mol. Plant* 4, 965–984. doi: 10.1093/mp/ssr021

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Gupta, M. L., Prasad, A., Ram, M., and Kumar, S. (2002). Effect of the vesicular–arbuscular mycorrhizal (VAM) fungus glomus fasciculatum on the essential oil yield related characters and nutrient acquisition in the crops of different cultivars of menthol mint (Mentha arvensis) under field conditions. *Biores Technol.* 81, 77–79. doi: 10.1016/S0960-8524(01)00109-2

Hurd, M. C., Kwon, M., and Ro, D.-K. (2017). Functional identification of a lippia dulcis bornyl diphosphate synthase that contains a duplicated, inhibitory arginine-rich motif. *Biochem. Biophys. Res. Commun.* 490, 963–968. doi: 10.1016/j.bbrc.2017.06.147

Johnson, D. A., Baker, R., and Boydston, R. A. (2013). Field evaluation of mutant and hybrid lines of mint for resistance to verticillium wilt and yield. *Crop Prot.* 43, 1–6. doi: 10.1016/j.cropro.2012.08.010

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275

Junker, R. R., and Tholl, D. (2013). Volatile organic compound mediated interactions at the plant-microbe interface. *J. Chem. Ecol.* 39, 810–825. doi: 10.1007/s10886-013-0325-9

Kampranis, S. C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N. A., et al. (2007). Rational conversion of substrate and product specificity in a salvia monoterpene synthase: Structural insights into the evolution of terpene synthase function. *Plant Cell* 19, 1994–2005. doi: 10.1105/tpc.106.047779

Karagiannidis, N., Thomidis, T., Lazari, D., Panou-Filotheou, E., and Karagiannidou, C. (2011). Effect of three Greek arbuscular mycorrhizal fungi in improving the growth, nutrient concentration, and production of essential oils of oregano and mint plants. *Sci. Hortic. (Amsterdam)* 129, 329–334. doi: 10.1016/j.scienta.2011.03.043

Khalvandi, M., Amerian, M., Pirdashti, H., and Keramati, S. (2021). Does co-inoculation of mycorrhiza and piriformospora indica fungi enhance the efficiency of chlorophyll fluorescence and essential oil composition in peppermint under irrigation with saline water from the Caspian Sea? *PLoS One* 16, e0254076. doi: 10.1371/journal.pone.0254076

Khalvandi, M., Amerian, M., Pirdashti, H., Keramati, S., and Hosseini, J. (2019). Essential oil of peppermint in symbiotic relationship with piriformospora indica and methyl jasmonate application under saline condition. *Ind. Crop Prod.* 127, 195–202. doi: 10.1016/j.indcrop.2018.10.072

Kim, H., Srividya, N., Lange, I., Huchala, E. W., Ginovska, B., Lange, B. M., et al. (2022). Determinants of selectivity for the formation of monocyclic and bicyclic products in monoterpene synthases. *ACS Catal.* 12, 7453–7469. doi: 10.1021/acscatal.2c01836

Kowalski, R., Kowalska, G., Jankowska, M., Nawrocka, A., Kałwa, K., Pankiewicz, U., et al. (2019). Secretory structures and essential oil composition of selected industrial species of lamiaceae. *Acta Sci. Pol. Hort Cult* 18, 53–69. doi: 10.24326/asphc.2019.2.6

Köllner, T. G., Lenk, C., Schnee, C., Köpke, S., Lindemann, P., Gershenzon, J., and Degenhardt, J. (2013). Localization of sesquiterpene formation and emission in maize leaves after herbivore damage. *BMC Plant Biol.* 13(1), 1–10.

Lange, B. M. (2015a). The evolution of plant secretory structures and emergence of terpenoid chemical diversity. *Annu. Rev. Plant Biol.* 66, 139–159. doi: 10.1146/annurev-arplant-043014-114639

Lange, B. M. (2015b). Biosynthesis and biotechnology of p-menthane monoterpenes, including menthol, carvone and limonene. *Adv. Biochem. Eng. Biotechnol.* 148, 319–353. doi: 10.1007/10_2014_289

Lange, B. M., and Srividya, N. (2019). Enzymology of monoterpene functionalization in glandular trichomes. *J. Exp. Bot.* 70, 1095–1108. doi: 10.1093/jxb/ery436

Lange, B., and Turner, G. (2013). Terpenoid biosynthesis in trichomes–current status and future opportunities. *Plant Biotechnol. J.* 11, 2–22. doi: 10.1111/j.1467-7652.2012.00737.x

Lange, B. M., Wildung, M. R., Stauber, E. J., Sanchez, C., Pouchnik, D., and Croteau, R. (2000). Probing essential oil biosynthesis and secretion by functional evaluation of expressed sequence tags from mint glandular trichomes. *Proc. Natl. Acad. Sci. U.S.A.* 97, 2934–2939. doi: 10.1073/pnas.97.6.2934

Lawrence, B. M. (2006). *Mint: The Genus Mentha* (Boca Raton, FL, USA: CRC Press).

Lawrence, B. M. (2008). *Peppermint Oil* (Carol Stream, IL, USA: Allured Publishing Corporation).

Likens, S., and Nickerson, G. (1964). Detection of certain hop oil constituents in brewing products. *Proceedings Annu. Meeting Am. Soc. Brew Chemi*, 5–13. doi: 10.1080/00960845.1964.12006730

Lima, A. S., Schimmel, J., Lukas, B., Novak, J., Barroso, J. G., Figueiredo, A. C., et al. (2013). Genomic characterization, molecular cloning and expression analysis of two terpene synthases from thymus caespititius (Lamiaceae). *Planta* 238, 191–204. doi: 10.1007/s00425-013-1884-2

Mendes, M. D., Barroso, J. G., Oliveira, M. M., and Trindade, H. (2014). Identification and characterization of a second isogene encoding γ-terpinene synthase in thymus caespititius. *J. Plant Physiol.* 171, 1017–1027. doi: 10.1016/j.jplph.2014.04.001

Merlin, E., Melato, E., Lourenco, E. L. B., Jacomassi, E., Junior, A. G., da Cruz, R. M. S., et al. (2020). Inoculation of arbuscular mycorrhizal fungi and phosphorus addition increase coarse mint (Plectranthus amboinicus lour.) plant growth and essential oil content. *Rhizosphere* 15, 100217. doi: 10.1016/j.rhisph.2020.100217

Mucciarelli, M., Camusso, W., Maffei, M., Panicco, P., and Bicchi, C. (2007). Volatile terpenoids of endophyte-free and infected peppermint (Mentha piperita l.): chemical partitioning of a symbiosis. *Microb. Ecol.* 54, 685. doi: 10.1007/s00248-007-9227-0

Peñuelas, J., Asensio, D., Tholl, D., Wenke, K., Rosenkranz, M., Piechulla, B., et al. (2014). Biogenic volatile emissions from the soil. *Plant Cell Environ.* 37, 1866–1891. doi: 10.1111/pce.12340

Ringer, K. L., McConkey, M. E., Davis, E. M., Rushing, G. W., and Croteau, R. (2003). Monoterpene double-bond reductases of the (-)-menthol biosynthetic pathway: isolation and characterization of cDNAs encoding (-)-isopiperitenone reductase and (+)-pulegone reductase of peppermint. *Arch. Biochem. Biophys.* 418, 80–92. doi: 10.1016/S0003-9861(03)00390-4

Roberts, D. D., Pollien, P., and Milo, C. (2000). Solid-phase microextraction method development for headspace analysis of volatile flavor compounds. *J. Agric. Food Chem.* 48, 2430–2437.

Rohloff, J. (1999). Monoterpene composition of essential oil from peppermint (Mentha × piperita l.) with regard to leaf position using solid-phase microextraction and gas chromatography/mass spectrometry analysis. *J. Agr Food Chem.* 47, 3782–3786. doi: 10.1021/jf981310s

Roos, J., Bejai, S., Mozūraitis, R., and Dixelius, C. (2015). Susceptibility to verticillium longisporum is linked to monoterpene production by TPS 23/27 in arabidopsis. *Plant J.* 81, 572–585. doi: 10.1111/tpj.12752

Silva, C. L., and Câmara, J. S. (2013). Profiling of volatiles in the leaves of lamiaceae species based on headspace solid phase microextraction and mass spectrometry. *Food Res. Int.* 51, 378–387. doi: 10.1016/j.foodres.2012.12.040

Srividya, N., Lange, I., and Lange, B. M. (2016). Generation and functional evaluation of designer monoterpene synthases. *Methods Enzymol.* 576, 147–165. doi: 10.1016/bs.mie.2016.03.013

Taherpour, A. A., Khaef, S., Yari, A., Nikeafshar, S., Fathi, M., and Ghambari, S. (2017). Chemical composition analysis of the essential oil of mentha piperita l. from kermanshah, Iran by hydrodistillation and HS/SPME methods. *J. Anal. Sci. Technol.* 8, 11. doi: 10.1186/s40543-017-0122-0

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

Tholl, D., Boland, W., Hansel, A., Loreto, F., Röse, U. S., and Schnitzler, J. P. (2006). Practical approaches to plant volatile analysis. *Plant J.* 45, 540–560. doi: 10.1111/j.1365-313X.2005.02612.x

Tohidi, B., Rahimmalek, M., Arzani, A., and Trindade, H. (2020). Sequencing and variation of terpene synthase gene (TPS2) as the major gene in biosynthesis of thymol in different thymus species. *Phytochemistry* 169, 112126. doi: 10.1016/j.phytochem.2019.112126

Urcoviche, R. C., Gazim, Z. C., Dragunski, D. C., Barcellos, F. G., and Alberton, O. (2015). Plant growth and essential oil content of mentha crispa inoculated with arbuscular mycorrhizal fungi under different levels of phosphorus. *Ind. Crop Prod.* 67, 103–107. doi: 10.1016/j.indcrop.2015.01.016

Vining, K. J., Johnson, S. R., Ahkami, A., Lange, I., Parrish, A. N., Trapp, S. C., et al. (2017). Draft genome sequence of mentha longifolia (L.) huds. and development of resources for mint cultivar improvement. *Mol. Plant* 10, 323–339. doi: 10.1016/j.molp.2016.10.018

Vining, K. J., and Pandelova, I. (2022). Dynamic tissue–specific transcriptome changes in response to verticillium dahliae in wild mint species mentha longifolia. *Plants* 11, 674. doi: 10.3390/plants11050674

Vining, K. J., Pandelova, I., Hummer, K., Bassil, N., Contreras, R., Neill, K., et al. (2019). Genetic diversity survey of mentha aquatica l. and mentha suaveolens ehrh., mint crop ancestors. *Genet. Resour. Crop Evol.* 66, 825–845. doi: 10.1007/s10722-019-00750-4

Vining, K. J., Pandelova, I., Lange, I., Parrish, A., Lefors, A., Kronmiller, B., et al. (2022). Chromosome-level assembly of the mentha longifolia l. reveals gene organization underlying disease resistance and essential oil traits. *G3 Genes Genomes Genet.* 12, jkac112. doi: 10.1093/g3journal/jkac112

Vining, K. J., Zhang, Q., Tucker, A. O., Smith, C., and Davis, T. M. (2005). Mentha longifolia (L.): A model species for mint genetic research. *Hortsci* 40, 1225–1229. doi: 10.21273/HORTSCI.40.5.1225

Wei, F., Passey, T., and Xu, X. (2016). Effects of individual and combined use of bio-fumigation-derived products on the viability of verticillium dahliae microsclerotia in soil. *Crop Prot.* 79, 170–176. doi: 10.1016/j.cropro.2015.09.008

Wise, M. L., Savage, T. J., Katahira, E., and Croteau, R. (1998). Monoterpene synthases from common sage (Salvia officinalis) - cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. *J. Biol. Chem.* 273, 14891–14899. doi: 10.1074/jbc.273.24.14891

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 June 2023
DOI 10.3389/fpls.2023.1174450



OPEN ACCESS

EDITED BY
Giuseppe Forlani,
University of Ferrara, Italy

REVIEWED BY
Zhiyong Liu,
Chinese Academy of Sciences (CAS), China
Wei Wei,
University of California, Berkeley,
United States

*CORRESPONDENCE
Rong Liu
liurong@yibinu.edu.cn
Yu Wu
wuyu@cib.ac.cn

†These authors have contributed equally to this work

RECEIVED 26 February 2023
ACCEPTED 05 May 2023
PUBLISHED 05 June 2023

CITATION
Liu R, Lv X, Wang X, Yang L, Cao J, Dai Y, Wu W and Wu Y (2023) Integrative analysis of the multi-omics reveals the stripe rust fungus resistance mechanism of the *TaPAL* in wheat. *Front. Plant Sci.* 14:1174450. doi: 10.3389/fpls.2023.1174450



# Integrative analysis of the multi-omics reveals the stripe rust fungus resistance mechanism of the *TaPAL* in wheat

Rong Liu[1*†], Xue Lv[2†], Xiaohua Wang[1], Li Yang[3], Jun Cao[1], Ya Dai[1], Wang Wu[1] and Yu Wu[2*]

[1]Faculty of Agriculture, Forestry and Food Engineering of Yibin University, Yibin, China, [2]Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China, [3]Wuhan Metware Biotechnology, Wuhan, Wuhan, China

Wheat is one of the major food crops in the world. However, stripe rust fungus significantly decreases wheat yield and quality. In the present study, transcriptomic and metabolite analyses were conducted in R88 (resistant line) and CY12 (susceptible cultivar) during Pst-CYR34 infection due to the limited availability of information regarding the underlying mechanisms governing wheat–pathogen interactions. The results revealed that Pst infection promoted the genes and metabolites involved in phenylpropanoid biosynthesis. The key enzyme gene *TaPAL* to regulate lignin and phenolic synthesis has a positive resistance contribution to Pst in wheat, which was verified by the virus-induced gene silencing (VIGS) technique. The distinctive resistance of R88 is regulated by the selective expression of genes involved in the fine-tuning of wheat–Pst interactions. Furthermore, metabolome analysis suggested that lignin biosynthesis-related metabolite accumulation was significantly affected by Pst. These results help to elucidate the regulatory networks of wheat–Pst interactions and pave the way for durable resistance breeding in wheat, which may ease environmental and food crises around the world.

KEYWORDS

**wheat, stripe rust, lignin, TaPAL, CYR34**

## 1 Introduction

Plants are unceasingly under attack by pathogenic microbes (or pathogens) that have activated an additional plant immune system to defend against invading pathogens (Harvey et al., 2020). Stripe or yellow rust (Yr) of wheat, caused by *Puccinia striiformis* f. sp. *tritici* (Pst), is the most globally devastating fungal disease, which has massively decreased wheat production and quality and ranged from 10% to 70% in most of the wheat-growing areas (Kiran et al., 2017; Wu et al., 2022). In China, stripe rust is usually threatening major wheat-growing regions in the Northern, Southwestern, and

Northwestern provinces along the Huai and Yellow River regions (Chen, 2005; Han et al., 2015; Jiang et al., 2022). Studies have found that the most effective and environment-friendly way to control wheat stripe rust is the cultivation of resistant varieties (Zheng et al., 2017; Chen et al., 2019), which can reduce the indiscriminate use of fungicides that cause serious environmental problems and health hazards to animals and humans. Wheat cultivars with race-specific resistance maintain resistance for only a few years due to the rapid evolution of new Pst races (Xu et al., 2019). For example, a new Pst race (V26, CYR34) virulent to the resistance gene *Yr26* has spread in several major wheat-growing regions where Yr26 is widely used in wheat breeding programs (Han et al., 2015). A recent study has reported that most current wheat cultivars and breeding lines (76%) are susceptible to Pst-CYR34 (Liu et al., 2010). Therefore, it is necessary to explore the mechanism of disease resistance and develop new strategies to improve disease resistance in wheat.

Certain disease-resistance-related enzymes and genes, including *PAL*, *peroxidase* (*POD*), and lignin, have been reported to play an important role in protecting plants against pathogen infection (Dias et al., 2017; Jamieson et al., 2017; Singh and Jha, 2017). It is well known that the phenylalanine ammonia-lyase (PAL) is usually involved in plants under biotic and abiotic stresses (Cass et al., 2015; Jamieson et al., 2017). Previous reports found that the *PAL* gene is considered to be involved in the main resistance processes in wheat (Van Eck et al., 2010). The enzymes of POD catalyze the polymerization of monolignols to generate lignin, which is an aromatic heteropolymer synthesized by phenylpropanoid metabolism (Zhang et al., 2020). The current perspective on plant immunity has evolved towards a comprehensive view of plant–pathogen interactions (Dodds and Rathjen, 2010). However, the gene expression profiles associated with the responses to stripe rust races in both resistant and susceptible wheat are still lacking, and most of the regulatory networks and the molecular mechanisms of the resistant gene to Pst remain unknown.

Therefore, in this study, the metabolomic and transcriptomic responses of wheat cv. Chuanyu12 (CY12) and the resistance breeding line R88 were investigated under Pst-CYR34 infection. In order to investigate the gene expression and metabolite accumulation in Pst response pathways and reveal the underling mechanisms governing wheat resistance to Pst, the candidate gene *TaPAL* was cloned and validated by virus-induced gene silencing (VIGS)-induced gene transient silencing, which proved that *TaPAL* plays an important role in wheat to Pst resistance.

# 2 Materials and methods

## 2.1 Plant materials and fungal treatment

The wheat cv. CY12 and the resistance breeding line R88 were used in this study. With the emergence of new Pst races, CY12 is highly susceptible to the currently predominant stripe rust. Wheat line R88 is an elite wheat germplasm from the offspring of CY12 with high all-stage resistance to CYR32 and CYR34. Artificial inoculation was conducted under controlled greenhouse conditions at the Gansu Academy of Agricultural Sciences at the wheat seedling stage. The wheat cv. Mingxian169 was used to monitor inoculation efficiency. The wheat leaves of CY12 and R88 were collected at 0 h (a), 24 h (b), and 72 h (c) post-inoculation, respectively (each time point with three biological replicates) (Wang et al., 2019; Francesconi and Balestra, 2020). All leaf samples were frozen in liquid nitrogen and stored in a −80° C freezer for subsequent RNA isolation, transcriptome sequencing, and metabolome detection.

## 2.2 RNA isolation and cDNA library construction for RNA sequencing

Total RNA of wheat leaves was extracted using the mirVana miRNA Isolation Kit (Ambion) following the manufacturer's protocol. Then, mRNA was broken into short fragments at a suitable temperature in a thermomixer. Purified fragmented mRNA was used to synthesize first- and second-strand cDNA. The libraries were constructed using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Then, 18 cDNA libraries were sequenced on the Illumina sequencing platform (HiSeqTM 2500 or Illumina HiSeq X Ten), and 125 bp/150 bp paired-end reads were generated. Raw reads containing poly-N and low-quality reads were removed to obtain clean reads. All clean reads were used for subsequent analyses and mapped to the reference genome using HISAT2.

## 2.3 Differential expression genes, GO, and KEGG analysis

The fragments per kilobase of transcript per million mapped reads (FPKM) value of each gene was calculated, and the read counts of each gene were obtained by htseq-count. Differentially expressed genes (DEGs) were analyzed and identified using the DEGseq R package and selected with $|\log2\ FC|>1$ and $Q<0.005$. GO enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of DEGs were performed using R based on the hypergeometric distribution during CYR34 infection. GO classification was performed by mapping the relation between SwissProt and GO terms. The major signaling pathways involved in DEGs were enriched in the KEGG database (Kanehisa et al., 2008).

## 2.4 Metabolite extraction and HPLC-MS/MS analysis

The freeze–dried wheat leaf samples were crushed and extracted with 1.0 ml of 70% methanol aqueous solution overnight at 4°C. The sample extracts were analyzed using a liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) system (Q TRAP, Applied Biosystems API 6500 Q TRAP LC/MS/MS System, Foster City, CA, USA). The analytical conditions were quantified as previously described (Jost et al., 2019).

Metabolites were quantified by calculating the area of each individual peak, and significant differences in content were set with thresholds of variable importance in projection ≥1 and fold change ≥1.5 or ≤0.5. Three biological and three technical replicates were performed in the metabolomics analysis.

## 2.5 Determination of lignin content and enzyme activity

R88 and CY12 wheat leaves were collected at 0, 24, and 72 hpi (Wang et al., 2019; Francesconi and Balestra, 2020). The lignin content was determined by the lignin-thioglycolic acid reaction (dos Santos et al., 2015). The determination of PAL enzyme activity was performed according to a previous study (Wang et al., 2016). The enzyme activities of POD and catalase (CAT) were measured by the increase in absorbance at 470 and 240 nm, respectively, according to the methods of Liu et al. (2019). Lipid peroxidation was evaluated by measuring malondialdehyde content (MDA). According to the method of the previous study (Liu et al., 2019), MDA content and proline (Pro) concentration were also determined with a slightly modified method.

## 2.6 qRT-PCR analysis

qRT-PCR was performed with a SYBR Green PCR kit (Qiagen, 204054) according to the manufacturer's instructions. The experimental conditions were set as follows: 45 cycles at 95°C for 20 s, 55°C for 20 s, and 72°C for 20 s. All genes were subjected to three technical repetitions in this study. The mRNA expression level of the genes was calculated with the $2^{-\Delta\Delta Ct}$ method, and *GAPDH* was used as the internal reference gene. Each wheat leaf sample was repeated three times. The correlation between the RNA-seq and qRT-PCR results was analyzed using the R package version 3.1.3 (http://cran.r-project.org/). The normalized values of relative expression and FPKM values were calculated using log2 (fold change) measurements.

## 2.7 TaPAL gene cloning and BSMV-mediated gene silencing

Based on the gene sequence from the wheat genome (IWGSC v1.1), the *TaPAL* homologous genes (*TraesCS6A02G222900*, *TraesCS6B02G258400*, and *TraesCS6D02G212200*) of wheat were amplified from the cDNA of R88 and CY12 by three pairs of gene-specific primers (Supplementary Table S1), followed by cloning into the pEASY-T1 simple cloning vector (Transgene Biotech, Beijing, China) before sequencing. PCR was carried out as follows: initial denaturation at 98°C for 2 min, followed by 35 cycles of 98°C for 10 s, 52°C for 30 s, and 72°C for 1 min, followed by a final extension at 72°C for 5 min, followed by a 4°C hold. The purified PCR product was sequenced by Tsingke Co., Ltd. (www.tsingke.net).

Barley stripe mosaic virus (BSMV), consisting of RNAs α, β, and γ, induced gene silencing was conducted in this study. BSMV construct vectors were obtained from Professor Jie Liu's lab at Northwest Agriculture & Forestry University. The 221 bp (V-TaPAL1as) and 202 bp (V-TaPAL2as) TaPAL gene fragments were amplified with PacI and NotI restriction sites using the primer pair V-TaPAL F/R (Supplementary Table S1) and inserted into the original BSMV:γ vector to produce the recombinant plasmid BSMV : TaPAL. Using the same method, we constructed the BSMV : TaPDS expression vector as the positive control. Plasmids BSMV : TaPAL, BSMV : TaPDS, and BSMV:γ were linearized followed by transcribing and capping *in vitro* using the RiboMA Large-Scale RNA Production System-T7 (Promega, USA) according to the manufacturer's protocol. The BSMV:α, BSMV:β, and BSMV:γ, BSMV : TaPDS, and BSMV : TaPAL were separately mixed in a 1: 1: 1 ratio and then added to the FES buffer for mechanically rubbed inoculation in the dark for 24 h. Ten days after BSMV inoculation, the phenotype of the fourth leaves of wheat caused by Pst infection was measured. All primers used for gene cloning, vector construction, and qRT-PCR analysis are listed in Supplementary Table S1.

# 3 Results

## 3.1 Analysis of physiological indexes in wheat after Pst infection

Proline (Pro), malondialdehyde (MDA), lignin, PAL, POD, and CAT play important roles in plant response to biotic and abiotic stresses. Therefore, in this experiment, these physiological and biochemical indices were detected at different time points of wheat infection with CYR34 (Figure 1). The results showed that these physiological indexes were significantly increased in CY12 and R88 with the increase in Pst infection time. There was no significant change in proline content in both the 0- and 24-h Pst infections at CY12 and R88. Compared with CY12, the proline content of R88 increased significantly after 72 h of Pst treatment (Figure 1A). On the contrary, MDA content in R88 treated with Pst was significantly lower than that in CY12 (Figure 1B). In addition, the changes in lignin content, PAL, POD, and CAT enzyme activity were similar to those in CY12 and R88 at 24 and 72 h after inoculation with Pst. Among them, before inoculation (a), the basal lignin level, POD, and CAT enzyme activity levels of CY12 and R88 were similar; only PAL enzyme activity was significantly higher in R88 than in CY12 (Figure 1). At 24 h after Pst treatment (b), lignin, PAL, POD, and CAT increased significantly in R88 by 1.5, 1.5, 2.0, and 1.9 times, respectively. At 72 h after Pst treatment (c), lignin, PAL, POD, and CAT were significantly increased by 1.5, 1.4, 1.2, and 1.8 times in R88, respectively (Figures 1C–F). These results indicated that Pst-resistant wheat could significantly increase PAL, POD, and CAT enzyme activities, accumulate Pro and lignin content, and inhibit the increase in MDA to the maximum extent to protect its own cells from damage.

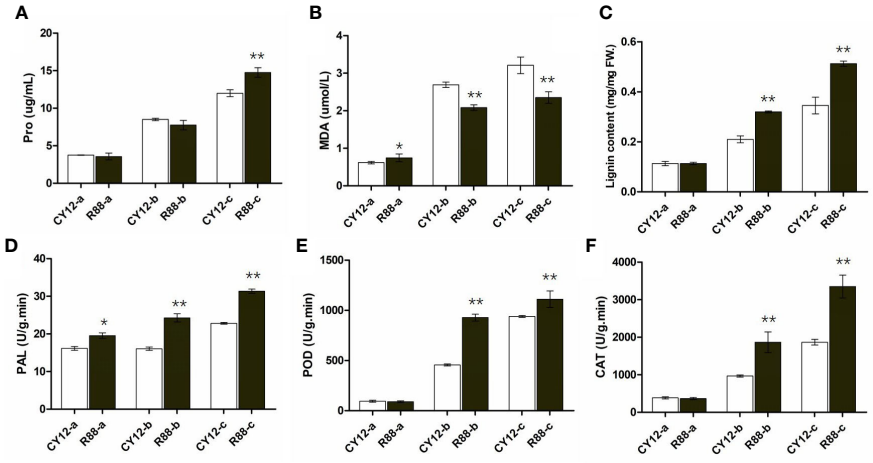



FIGURE 1
The physiological indices of the susceptible and resistant wheat seedlings were detected at different time points before and after Pst inoculation. Proline content **(A)**, malondialdehyde (MDA) content **(B)**, total lignin content **(C)**, and PAL, POD, and CAT enzyme activities **(D–F)** in wheat leaves. In each bar chart, a, b, and c represent 0 h before inoculation, 24 h after inoculation, and 72 h after inoculation, respectively. ** on the bar chart indicated the significant differences between R88 and CY12 after CYR34 inoculation at each time point (P<0.01).

## 3.2 Transcriptome analysis and DEGs identification of wheat

To elucidate the molecular basis for the differential stripe rust response in CY12 and R88, comparative transcriptome analysis was conducted through RNA sequencing (RNA-seq). The 24- and 72-h Pst treatment time points were selected to investigate the wheat response to Pst-CYR34 at different infection stages. Eighteen cDNA libraries were characterized by Illumina HiSeq to detect the transcriptome level of gene expression information. After removing low-quality reads, clean reads were obtained, and more than 87% of the clean reads per library could be mapped to the wheat reference genome (IWGSC RefSeq V1.1). There are more than 60,000 transcripts with a length of 1,201–1,800 bp, and nearly 80,000 transcripts with a length of >1,800 (Supplementary Figure S1A). These transcripts were functionally annotated using the GO, KEGG, COG, NR, Swiss-Port, and Pfam databases (Supplementary Figure S1B).

Principal component analysis (PCA) showed that the first principal component (PC1) could explain 47.63% of the total variance and distinguish samples based on the time of Pst-CYR34 inoculation, and the second principal component (PC2) could explain 13.91% of the total variance and separate each groups according to the three time points (Figure 2A). The heat map of sample correlations (sample to sample clustering) showed that gene expression values among 18 wheat samples were reproducible between the three biological replicates, and batch effects were controlled (Figure 2B). The differentially expressed genes (DEGs) were identified by comparing the FPKM values of each gene between CY12 and R88 (at 0-a, 24 hpi-b, and 72 hpi-c) with the criteria of fold change ≥2 and $p < 0.05$. Then, DEGs involved in the Pst-CYR34 infection process were screened. There were 4,607 DEGs between CY12 and R88 with non-Pst infection, while the number of DEGs changed to 4,834 and 4,456 after 24 and 72 hpi with CYR34 treatment, respectively (Figure 2C). There were 2,592 upregulated and 2,015 downregulated genes in R88-a vs CY12-a (Figure 2D), 2,721 upregulated and 2,113 downregulated genes in R88-b vs. CY12-b (Figure 2E), and 2,252 upregulated and 2,204 downregulated genes in R88-c vs. CY12-c (Figure 2F).

## 3.3 GO enrichment and KEGG analysis of DEGs

The DEGs contributed to the phenotypic differences in Pst-CYR34 infection between CY12 and R88. To identify the major functional terms under Pst-CYR34 treatment, GO enrichment analysis was carried out on the DEGs (Figure 3A). GO enrichment results showed that Pst-CYR34-related DEGs were mainly enriched in metabolic process, cellular process, response to stimulus, and biological regulation in the biological process category; cell part, organelle part, and membrane part in the cellular component category; and catalytic activity, binding, and transporter activity in the molecular function category.

DEGs were also enriched in KEGG pathways in this study. At 24 hpi, Pst-CYR34-induced DEGs in CY12 and R88 were notably enriched in necroptosis (ko04217), phenylpropanoid biosynthesis (ko00940), phenylalanine metabolism (ko00360), plant–pathogen interaction (ko04626), and MAPK signaling pathway–plant (ko04016) (Figure 3B). A total of 60 DEGs in phenylpropanoid biosynthesis were upregulated, including *PAL*, *POD*, *PTAL*, *COMT*, and *REF1*. Phenylalanine metabolism is the upstream pathway of phenylpropanoid biosynthesis, and most DEGs were also upregulated, including the enzyme encoding genes *amiE* and *MIF*. At 72 hpi, Pst-CYR34-induced DEGs were also enriched in phenylalanine metabolism, phenylpropanoid biosynthesis, flavonoid biosynthesis (ko00941), and plant–pathogen

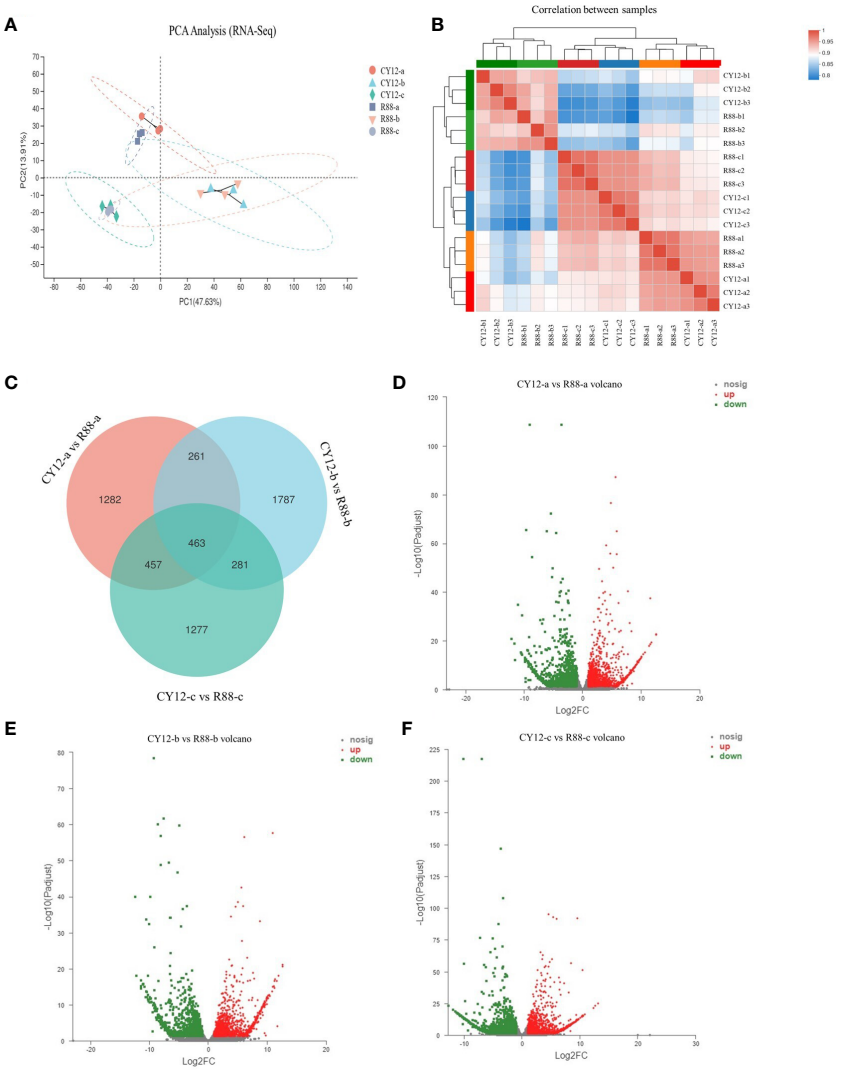



FIGURE 2
Principal component analysis (PCA) **(A)**, sample to sample correlation analysis **(B)**, Venn diagram **(C)**, and volcano diagram **(D–F)** analysis for differentially expressed genes (DEGs) in wheat after Pst infection. CY12-a and R88-a represent non-inoculated Pst on susceptible and resistant wheat seedlings. CY12-b and R88-b represent inoculated with Pst for 24 h on susceptible and resistant wheat seedlings. CY12-c and R88-c represent inoculated with Pst for 72 h on susceptible and resistant wheat seedlings. In the volcano diagram, red dots represent upregulated DEGs, green dots represent downregulated DEGs, and gray dots represent genes with insignificant differences in expression levels.

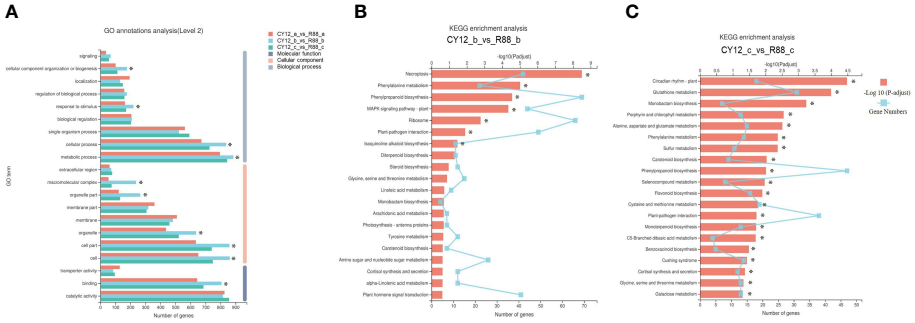



FIGURE 3
The Gene Ontology (GO) annotations **(A)** and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis in CY12-b vs R88-b **(B)** and CY12-c vs R88-c **(C)** comparison groups.

interactions in CY12 and R88 (Figure 3C). These results indicate that phenylpropanoid biosynthesis is important for the Pst response in wheat.

## 3.4 Metabolome analysis in wheat after inoculation with Pst

PCA of metabolome indicated that the PC1 could explain 59.43% of total variance and distinguish wheat seedlings based on the time of Pst-CYR34 inoculation, and the PC2 could explain 34.84% of the total variance (Figure 4A). The hierarchical clustering heat map showed each metabolite accumulation values among CY12 and R88 infected by Pst (0 and 24 h) (Figure 4B). The results showed that there was no significant difference in the level of metabolite accumulation between (a) R88 and CY12 before Pst inoculation. However, at 24 h after Pst inoculation (b), metabolites in CY12 and R88 accumulated significantly, and CY12 and R88 had their own metabolite content changes in response to Pst infection (Supplementary Figure S2B). Therefore, a detailed analysis of these differentially accumulated metabolites was carried out in this experiment.

First, differential accumulation metabolites (DAMs) identified after Pst inoculation were classified. Among them, flavonoids (59) and phenolic acids (48) account for 31.94% of the total metabolites (Figure 5A). The results showed that flavonoids and phenolic acids in wheat played a very important role in the resistance to Pst infection. Then, these metabolites were enriched by KEGG function. It turns out that these DAMs are mainly enriched in phenylpropanoid biosynthesis, flavonoid biosynthesis, starch and sucrose metabolism, phenylalanine, tyrosine, and tryptophan biosynthesis (Figure 5B). This result was also significantly related to the types of metabolites, suggesting that these pathways are also the main factors involved in wheat resistance to Pst.

## 3.5 Candidate gene cloning and VIGS verification

In order to further explore the relationship between phenylpropanoid, flavonoid biosynthesis, and lignin accumulation to Pst resistance in wheat, *TaPAL*, an upstream regulatory gene regulating flavonoid and lignin synthesis, was selected as a candidate gene for cloning and functional analysis based on gene expression and metabolite accumulation content. To verify the reliability of RNA-Seq data, we performed RT-PCR analysis of *TaPAL* gene (*TraesCS6B02G258400*). The results showed that CY12 and R88 had the same expression level of *TaPAL* gene before inoculation with Pst. However, the expression level was significantly increased 24 h after inoculation, and R88 was significantly higher than CY12 (Supplementary Figure S2A). Next, we designed primers to clone this candidate gene and to compare with Chinese spring sequences, in which the comparison rate was >97.1% and contains PLN02457 conserved domain (Supplementary Figure S2B).

The phylogenetic analysis revealed a close relationship of *TaPAL* to PAL proteins in *Zea mays*, *Oryza sativa*, *Aegilops tauschii*, and other plant species (Figure 6A). In this experiment, the phenotypes of CY12 and R88 seedlings were observed at 28 days after inoculation with Pst-CYR34. Among them, CY12 seedling leaves were basically covered by stripe rust spores, while R88 was almost immune to stripe rust (Figure 6B). Two specific gene fragments (1 as and 2 as) were designed to specifically silence the *TaPAL* gene in wheat. Compared with CK, the 2-week-old wheat leaves faded after inoculation with BSMV : TaPDS in CY12 and R88 (Figure 6C). The results suggest that *TaPDS* was successfully silenced in wheat seedlings. The virus leaf with empty vector and target gene showed striped chlorotic symptoms (Figure 6C), which indicate that the *TaPAL* gene silencing vector was also successfully constructed in this study. After 20 days of inoculation with Pst-CYR34, CY12 and MX169 were very susceptible to disease and a

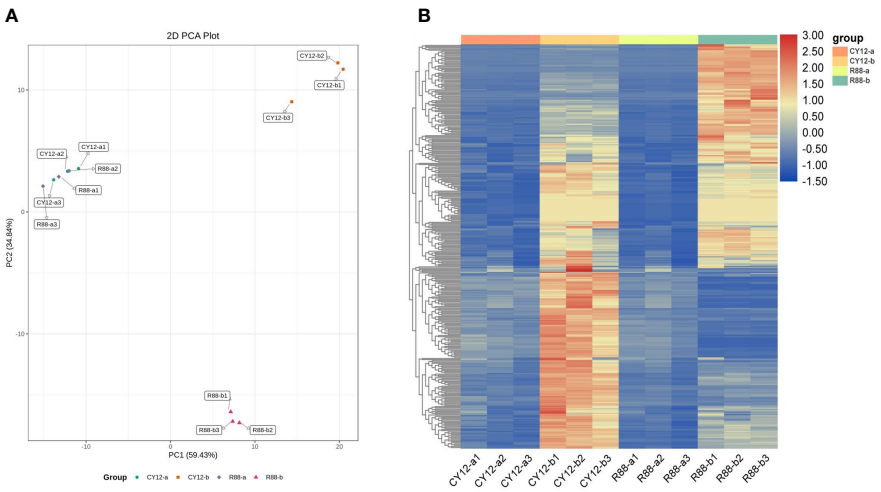



FIGURE 4
Principal component analysis (PCA) of differentially accumulated metabolites (DAMs) during each Pst inoculation periods in CY12 and R88 **(A)** and the heat map of relative accumulated content of DAMs during each Pst inoculation periods in CY12 and R88 **(B)**.

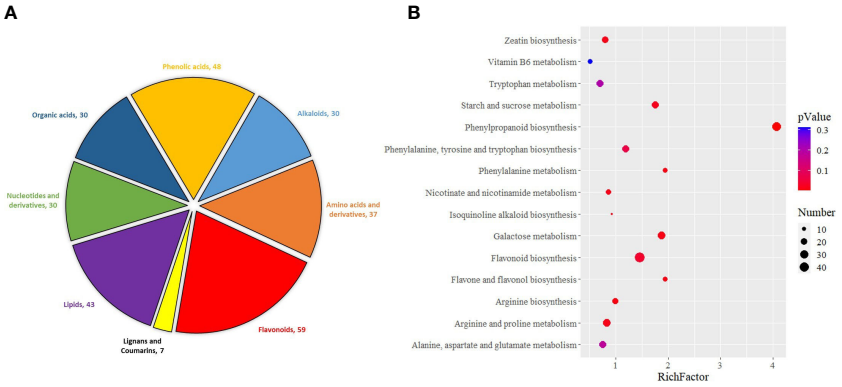



FIGURE 5
Classification of metabolites in Pst-resistant and Pst-susceptible wheat seedlings at different periods **(A)** and KEGG enrichment analysis of differential accumulation metabolites **(B)**.

large number of spore piles appeared in the leaves, while R88 was a near-immune wheat line and almost no spore piles were seen in the leaves (Figure 6C). Compared with WT (R88 without gene silencing), R88 1as and R88 2as inoculated with Pst-CYR34 displayed susceptibility in BSMV : TaPAL-silenced lines, which leaves appeared a small amount of spore piles (Figure 6C). These results suggest that *TaPAL* is associated with Pst-CYR34 resistance in R88 wheat, which can positively regulate the resistance response in wheat by regulated phenylpropanoid biosynthesis and lignin biosynthesis.

## 3.6 Combined transcriptome and metabolome analysis in wheat to resistant Pst-CYR34

Combined analysis of transcriptome and metabolome data showed that the Pst-induced DEGs and metabolites involved in phenylpropanoid biosynthesis were examined in this study. Compared with CY12, most DEGs of phenylpropanoid biosynthesis pathway were upregulated in L58 after Pst inoculation at 24 h (Figure 7). In this study, phenylpropanoid was

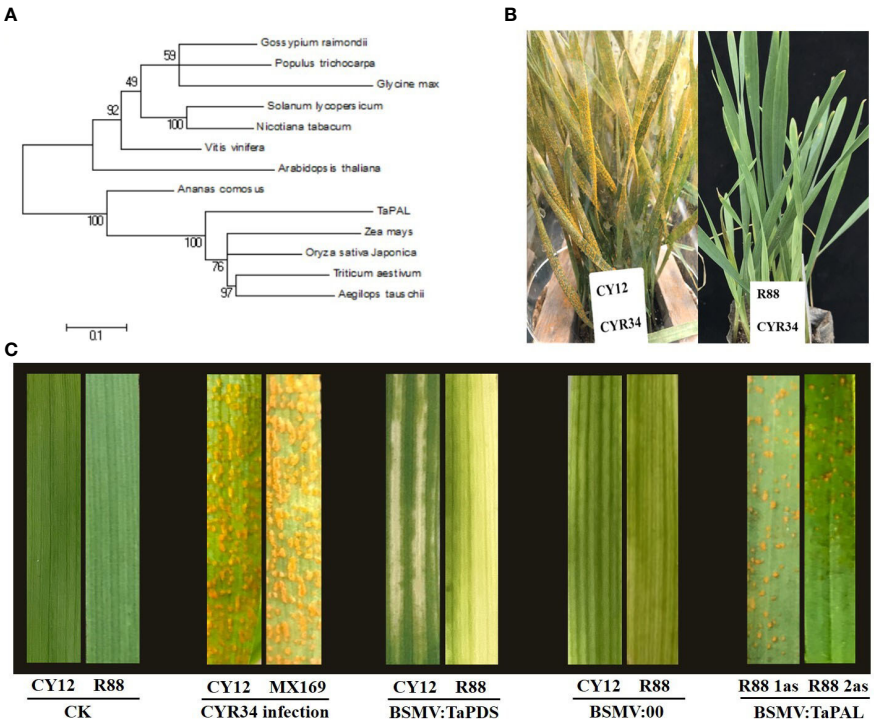



FIGURE 6
Phylogenetic analysis of *TaPAL* in wheat and other plants **(A)**, phenotypes of CY12 and R88 after inoculation with Pst-CYR34 at 28 days **(B)**, and short-time gene silencing of *TaPAL* by VIGS for gene function verification **(C)**. CK, not inoculated Pst-CYR34 (negative control); BSMV : TaPDS, silence TaPDS gene; BSMV:00, unconnected gene fragment (positive control); BSMV : *TaPAL*, silence TaPAL gene.

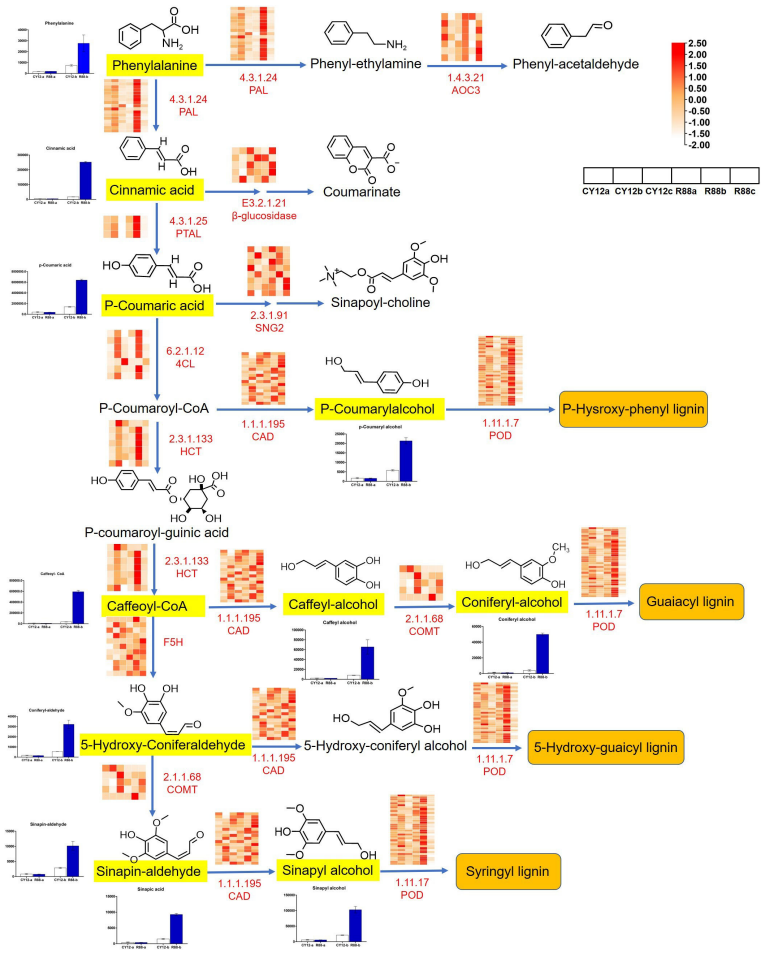



FIGURE 7
Analysis of DEGs and differentially accumulated metabolites (DAMs) in phenylpropanoid biosynthesis pathway flow to lignin synthesis after inoculation of Pst in wheat. The black letters represent metabolites produced in the pathway, with the chemical formula for the metabolite at the top. The bar chart shows the relative content changes of the corresponding metabolites during different periods of Pst inoculation. The small heat map shows the changes in gene expression marked in red letters during different periods of Pst inoculation. In the color block bar, red represents upregulated gene expression, while white represents downregulated gene expression.

biosynthesized from phenylalanine, followed by different steps of the pathway catalyzed by four types of enzymes: ammonia lyase (PAL), hydroxycinnamoyl transferase (HCT), cinnamyl alcohol dehydrogenase (CAD), and E1.11.1.7. Finally, the pathway produced three types of monolignols that polymerize to form lignin: p-hydroxyphenyl (H), guaiacyl (G), 5-hydroxy-guaicyl, and syringyl (S) lignin. The expression levels of DEGs in this pathway were illustrated using a heat map and estimated by log2 (fold change) (Figure 7). Although there were some cases of downregulation or non-significant upregulation for some genes at different time points, most of the four types of genes in this pathway were upregulated in R88 within 24 hpi. Generally, the upregulation of phenylpropanoid biosynthesis genes began at 24 hpi or earlier.

Based on the upregulation of genes related to the phenylpropanoid biosynthesis pathway, we determined lignin biosynthesis-related metabolites by LC-MS/MS in CY12 and R88 at 24 hpi. Metabolome analysis showed that a total of 11 lignin-related metabolite compounds were detected in wheat, including phenylalanine, cinnamic acid, p-coumaric acid, p-coumarylalcohol, caffeoyl-coA, caffeyl-alcohol, coniferyl-alcohol, 5-hydroxy-coniferaldehyde, sinapin-aldehyde, and sinapyl-alcohol (Figure 7). Compared with CY12, the content of these metabolites was significantly increased in R88 at 24 hpi. All of these compounds had a >1.5-fold change in content. This result is consistent with that of gene expression in the transcriptome.

## 4 Discussion

Plant pathogens are major factors affecting crop yield and quality worldwide. Plants have evolved pathogen-associated molecular pattern-triggered immunity (PTI) and damage-associated molecular patterns that trigger immunity and effector-triggered immunity (ETI), which generate resistance gene products to specifically recognize the effectors released from the invader and to confer disease resistance (Tang et al., 2017). Previous studies have shown that proline act as osmo-regulator to biotic or abiotic stress, which can regulate osmolytes accumulation to scavenge reactive

oxygen species in response to these different stresses (Li et al., 2017). In the present study, our results indicated that the proline content were significantly increased in wheat seedlings after Pst infection. Similar to previous studies, proline significantly accumulated in tomato, eggplant, and certain mustard varieties by blight fungi infection (Kumari et al., 2017; Attia et al., 2022). The MDA content increased significantly by Pst infection at 24 and 72 h in both CY12 and R88 wheat seedlings, which can cause intracellular oxidative stress to severe disturbance in the plant cell (Badawy et al., 2021). However, the results of this experiment showed that the increased level of MDA in R88 was significantly lower than that in CY12 after Pst infection. This result indicates that Pst-resistant wheat plants can minimize the damage to plant cells due to lower MDA accumulation content. In addition, this experiment found that CAT, PAL, and POD enzyme activities are also related to plant resistance. The results of gene expression levels corresponding to these physiological indexes were consistent in the transcriptome profile. The results showed that most of the DEGs were highly expressed after inoculation with CYR-34. In addition, we found that those genes together with other genes also participate in the activation and response of disease-resistant pathways. These immune strategies trigger downstream responses, including the accumulation of secondary metabolites to activate defense systems (Wu et al., 2014).

Phenylpropanoid biosynthesis metabolism is the most important secondary metabolic pathway, playing a fundamental role in plant defense against biotic (invading pathogens) and abiotic stresses (La Camera et al., 2004). The main branches of the phenylpropanoid pathway lead to the synthesis of lignin and flavonoids. Although certain studies have provided evidence to suggest that lignin, flavonoid, and phenolic compounds play important roles during the plant defense response to pathogens, few studies have analyzed the overall transcriptome and metabolome changes in phenylpropanoid biosynthesis genes or related metabolites in wheat. The present results indicated that the total gene expression and metabolite accumulation profiles of the phenylpropanoid biosynthesis pathway were responsive to Pst-CYR34, which is consistent with the results of our previous studies (Liu et al., 2022). As a result, the study focused on the phenylpropanoid biosynthesis pathway with an emphasis on the synthesis and deposition of lignin because most of the enriched DEGs and related metabolites were upregulated in this pathway. Previous studies have reported that induced lignification is one of several plant defense responses to wounding and to viral or microbial attack (Bhuiyan et al., 2009). In infected plants, lignification and deposition of lignin are part of plant cell wall reinforcement and are important processes in the response of plants to restrict pathogen invasion (Kauss et al., 1993) and provide a physical barrier to limit pathogen colonization (Bonello and Blodgett, 2003).

Induced lignification around infection centers is generally accompanied by the increased activity of a number of enzymes. Genetic evidence for the role of lignin-associated defense is rare, however, either because of the indispensability of lignin for the plant form. In this study, DEGs encoding the enzymes involved in the lignin pathway showed upregulated expression, which indicated the putative role of lignin in the wheat defense response (Liu et al., 2022). The activities of enzymes involved in the biosynthesis of lignin were measured in this study. POD is involved in the polymerization of monolignols into lignin and reinforcement of the cell wall after pathogen attack and wounding (Marjamaa et al., 2009). Phenylalanine ammonia-lyase (PAL) plays an essential role in the phenylpropanoid pathway and has been reported to be responsive to both biotic and abiotic stresses, including pathogen attack, wounding, cold, UV stress, and other stress conditions (Huang et al., 2010). Consistent with previous reports, PAL and POD activities increased with localized lignin deposition (Nicholson and Hammerschmidt, 2003; Caldo et al., 2004) after Pst-CYR34 inoculation in both susceptible and resistant wheat. However, compared with resistant plants (R88), the inoculated susceptible plants (CY12) showed an obviously slower increase and lower enzyme activities. This result suggests that lignin may be the active form of lignification in HR to protect cells against fungal invasion (Bhuiyan et al., 2009). Together with the gene expression profiles, we speculated that the most significant difference between susceptible and resistant plants was whether the plants could activate the pathway in a timely and efficient manner.

The lignin biosynthetic pathway is initially induced in both resistant and susceptible wheat following rust infection (Moerschbacher et al., 1988). In this study, the metabolites of the lignin biosynthesis pathway were investigated in both resistant and susceptible wheat seedlings after Pst-CYR34 infection. Pst-CYR34 increased the total lignin content in wheat and to a lesser degree for CY12 than R88, which suggested that lignification may be a mechanism through which wheat restricts the pathogen. An increase in the lignin content was determined in the inoculated resistant wheat (R88) compared with the control at both 24 and 72 h after Pst infection. The 11 lignin-related compounds determined by LC-MS/MS also showed upregulation in R88 compared with CY12 at 24 hpi. Overall, the resistant wheat line displayed increased lignin concentrations and was speculated to play a vital role in the wheat defense response. In accordance with these reports, this result suggested that increased lignification or lignin biosynthesis is critically important for wheat to defend against pathogen invasion.

PAL is a key enzyme in the phenylpropanoid pathway, which is involved in the improvement of resistance plants (Zhang and Liu, 2015; Lu et al., 2019). *PAL* gene expression can be induced at the transcriptional level under the stress of exogenous signal compounds (mechanical damage, bacteria, and viruses), and PAL activity increases rapidly to activate phenylpropane metabolism in the defense system (Leyva et al., 1992; Pina and Errea, 2008). In this study, we have found that the *TaPAL* plays an important role involved in resistance response by regulating the production of lignin in wheat. After transient VIGS silencing *TaPAL* gene expression, Pst-resistant wheat line R88 showed symptoms of disease. These results suggest that *TaPAL* gene also plays an important role in wheat resistance to Pst. Previous studies have shown that PAL is also important for pathogen-induced SA formation and phenolic phytoalexins biosynthesis in plants

(Huang et al., 2010; He et al., 2020). Therefore, the results of this experiment speculated that *TaPAL* gene not only regulates the synthesis of lignin in wheat but also participates in the synthesis of phenolic phytoalexins and SA, so as to play a dual defense role in wheat after Pst infection.

## 5 Conclusion

To better understand the possible molecular mechanisms of plant–pathogen interactions and to improve plant disease resistance, molecular biology and bioinformatic technology have recently been developed to achieve these goals and cultivate wheat stripe rust resistance varieties. It is very important to mine resistance genes, which can improve wheat resistance to Pst and combine with other resistance-related genes to activate certain pathways to defend against the invasion of pathogenic microorganisms. In responding to pathogen (Pst-CYR34) invasion, wheat plants sense external signals in a timely fashion, and then, signal transduction is conducted by signaling pathways, plant defense-related gene expression, and cell wall reinforcement in wheat to an immune response. In the resistant wheat line R88, transcripts of lignin biosynthesis genes were significantly expressed at 24 hpi, which led lignin content accumulated in wheat to be a defense against Pst infection. By combining RNA-seq, LC-MS/MS and biochemical data, it was speculated that DEG expression levels, metabolite accumulation, enzyme activities, and total lignin contents were increased after Pst inoculation in wheat. At the same time, the function of *TaPAL*, a key enzyme in the phenylpropanoid biosynthesis pathway, was analyzed in the present study, which plays an important role in the regulation of lignin biosynthesis. The VIGS verification results show that *TaPAL* gene expression positively regulated Pst resistance in wheat. Therefore, a critical role for *TaPAL* expression level and lignin content was deduced to contribute to wheat disease resistance.

## 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

RL and XL contributed to the methodology, analyzed RNA-Seq and LC-MS/MS data, gene clone, draw figures, and wrote the draft-manuscript. RL and XW performed experimental physiological index detection. LY analyzed RNA-Seq and LC-MS/MS data. JC, YD and WW reviewed and edited the manuscript. YW and RL contributed to the conceptualization, resources, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

## Funding

This work is supported by grants from the Ph.D. Fund Project of Yibin University (2022QH07), Natural Science Foundation Project in Sichuan province (2023NSFSC1268), and the "Strategic Priority Research Program" of the Chinese Academy of Sciences (XDA24030401-2).

## Conflict of interest

LY was employed by Wuhan Metware Biotechnology.

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1174450/full#supplementary-material

## References

Attia, M. S., Hashem, A. H., Badawy, A. A., and Abdelaziz, A. M. (2022). Biocontrol of early blight disease of eggplant using endophytic aspergillus terreus: improving plant immunological, physiological and antifungal activities. *Bot. Stud.* 63 (1), 26. doi: 10.1186/s40529-022-00357-6

Badawy, A. A., Alotaibi, M. O., Abdelaziz, A. M., Osman, M. S., Khalil, A. M. A., Saleh, A. M., et al. (2021). Enhancement of seawater stress tolerance in barley by the endophytic fungus aspergillus ochraceus. *Metabolites* 11 (7), 428. doi: 10.3390/metabo11070428

Bhuiyan, N. H., Selvaraj, G., Wei, Y., and King, J. (2009). Gene expression profiling and silencing reveal that monolignol biosynthesis plays a critical role in penetration defence in wheat against powdery mildew invasion. *J. Exp. Bot.* 60 (2), 509–521. doi: 10.1093/jxb/ern290

Bonello, P., and Blodgett, J. T. (2003). Pinus nigra-sphaeropsis sapinea as a model pathosystem to investigate local and systemic effects of fungal infection of pines. *Physiol. Mol. Plant P* 63, 249–261. doi: 10.1016/j.pmpp.2004.02.002

Caldo, R. A., Nettleton, D., and Wise, R. P. (2004). Interaction-dependent gene expression in mla-specified response to barley powdery mildew. *Plant Cell* 16 (9), 2514–2528. doi: 10.1105/tpc.104.023382

Cass, C. L., Peraldi, A., Dowd, P. F., Mottiar, Y., Santoro, N., Karlen, S. D., et al. (2015). Effects of PHENYLALANINE AMMONIA LYASE (PAL) knockdown on cell

wall composition, biomass digestibility, and biotic and abiotic stress responses in brachypodium. *J. Exp. Bot.* 66 (14), 4317–4335. doi: 10.1093/jxb/erv269

Chen, X. M. (2005). Epidemiology and control of stripe rust [Puccinia striiformis f. sp. tritici] on wheat. *Can. J. Plant Pathol.* 27, 314–337. doi: 10.1080/07060660509507230

Chen, Y., Mao, H., Wu, N., Ma, J., Yuan, M., Zhang, Z., et al. (2019). Effects of stripe rust infection on the levels of redox balance and photosynthetic capacities in wheat. *Int. J. Mol. Sci.* 21 (1), 268. doi: 10.3390/ijms21010268

Dias, M. P., Bastos, M. S., Xavier, V. B., Cassel, E., Astarita, L. V., and Santarém, E. R. (2017). Plant growth and resistance promoted by streptomyces spp. *tomato Plant Physiol. Biochem.*, 118, 479–493. doi: 10.1016/j.plaphy.2017.07.017

Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* 11 (8), 539–548. doi: 10.1038/nrg2812

dos Santos, A. B., Bottcher, A., Kiyota, E., Mayer, J. L., Vicentini, R., Brito Mdos, S., et al. (2015). Water stress alters lignin content and related gene expression in two sugarcane genotypes. *J. Agric. Food Chem.* 63 (19), 4708–4720. doi: 10.1021/jf5061858

Francesconi, S., and Balestra, G. M. (2020). The modulation of stomatal conductance and photosynthetic parameters is involved in fusarium head blight resistance in wheat. *PloS One* 15 (6), e0235482. doi: 10.1371/journal.pone.0235482

Han, D. J., Wang, Q. L., Chen, X. M., Zeng, Q. D., Wu, J. H., Xue, W. B., et al. (2015). Emerging Yr26-virulent races of puccinia striiformis f. tritici are threatening wheat production in the sichuan basin, China. *Plant Dis.* 99 (6), 754–760. doi: 10.1094/PDIS-08-14-0865-RE

Harvey, S., Kumari, P., Lapin, D., Griebel, T., Hickman, R., Guo, W., et al. (2020). Downy mildew effector HaRxL21 interacts with the transcriptional repressor TOPLESS to promote pathogen susceptibility. *PloS Pathog.* 16 (8), e1008835. doi: 10.1371/journal.ppat.1008835

He, J., Liu, Y., Yuan, D., Duan, M., Liu, Y., Shen, Z., et al. (2020). An R2R3 MYB transcription factor confers brown planthopper resistance by regulating the phenylalanine ammonia-lyase pathway in rice. *Proc. Natl. Acad. Sci. U.S.A.* 117 (1), 271–277. doi: 10.1073/pnas.1902771116

Huang, J., Gu, M., Lai, Z., Fan, B., Shi, K., Zhou, Y. H., et al. (2010). Functional analysis of the arabidopsis PAL gene family in plant growth, development, and response to environmental stress. *Plant Physiol.* 153 (4), 1526–1538. doi: 10.1104/pp.110.157370

Jamieson, M. A., Burkle, L. A., Manson, J. S., Runyon, J. B., Trowbridge, A. M., and Zientek, J. (2017). Global change effects on plant-insect interactions: the role of phytochemistry. *Curr. Opin. Insect Sci.* 23, 70–80. doi: 10.1016/j.cois.2017.07.009

Jiang, B., Wang, C., Guo, C., Lv, X., Gong, W., Chang, J., et al. (2022). Genetic relationships of puccinia striiformis f. sp. tritici in southwestern and northwestern China. *Microbiol. Spectr.* 10 (4), e0153022. doi: 10.1128/spectrum.01530-22

Jost, T., Heymann, T., and Glomb, M. A. (2019). Efficient analysis of 2-Acetyl-1-pyrroline in foods using a novel derivatization strategy and LC-MS/MS. *J. Agric. Food Chem.* 67 (10), 3046–3054. doi: 10.1021/acs.jafc.9b00220

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882

Kauss, H., Franke, R., Krause, K., Conrath, U., Jeblick, W., Grimmig, B., et al. (1993). Conditioning of parsley (Petroselinum crispum l.) suspension cells increases elicitor-induced incorporation of cell wall phenolics. *Plant Physiol.* 102 (2), 459–466. doi: 10.1104/pp.102.2.459

Kiran, K., Rawal, H. C., Dubey, H., Jaswal, R., Bhardwaj, S. C., Prasad, P., et al. (2017). Dissection of genomic features and variations of three pathotypes of puccinia striiformis through whole genome sequencing. *Sci. Rep.* 7, 42419. doi: 10.1038/srep42419

Kumari, M., Pandey, S., Bhattacharya, A., Mishra, A., and Nautiyal, C. S. (2017). Protective role of biosynthesized silver nanoparticles against early blight disease in solanum lycopersicum. *Plant Physiol. Biochem.* 121, 216–225. doi: 10.1016/j.plaphy.2017.11.004

La Camera, S., Gouzerh, G., Dhondt, S., Hoffmann, L., Fritig, B., Legrand, M., et al. (2004). Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. *Immunol. Rev.* 198, 267–284. doi: 10.1111/j.0105-2896.2004.0129.x

Leyva, A., Liang, X., Pintor-Toro, J. A., Dixon, R. A., and Lamb, C. J. (1992). Cis-element combinations determine phenylalanine ammonia-lyase gene tissue-specific expression patterns. *Plant Cell* 4 (3), 263–271. doi: 10.1105/tpc.4.3.263

Li, X., Han, S., Wang, G., Liu, X., Amombo, E., Xie, Y., et al. (2017). The fungus aspergillus aculeatus enhances salt-stress tolerance, metabolite accumulation, and improves forage quality in perennial ryegrass. *Front. Microbiol.*, 8, 1664. doi: 10.3389/fmicb.2017.01664

Liu, R., Lu, J., Zhang, L., and Wu, Y. (2022). Transcriptomic insights into the molecular mechanism of wheat response to stripe rust fungus. *Heliyon* 8 (10), e10951. doi: 10.1016/j.heliyon.2022.e10951

Liu, T. G., Peng, Y. L., Chen, W. Q., and Zhang, Z. Y. (2010). First detection of virulence in puccinia striiformis f. sp. tritici in China to resistance genes Yr24 (=Yr26) present in wheat cultivar chuanmai 42. *Plant Dis.* 94 (9), 1163. doi: 10.1094/PDIS-94-9-1163C

Liu, R., Zhang, Q. N., Lu, J., Zhang, C. H., Zhang, L., and Wu, Y. (2019). The effects of exogenous pyridoxal-5-phosphate on seedling growth and development of wheat under salt stress. *Cereal Res. Commun.* 47 (3), 442–454. doi: 10.1556/0806.47.2019.22

Lu, J., Shi, Y., Li, W., Chen, S., Wang, Y., He, X., et al. (2019). RcPAL, a key gene in lignin biosynthesis in ricinus communis l. *BMC Plant Biol.* 19 (1), 181. doi: 10.1186/s12870-019-1777-z

Marjamaa, K., Kukkola, E. M., and Fagerstedt, K. V. (2009). The role of xylem class III peroxidases in lignification. *J. Exp. Bot.* 60 (2), 367–376. doi: 10.1093/jxb/ern278

Moerschbacher, B. M., Noll, U. M., Flott, B. E., and Reisener, H. J. (1988). Lignin biosynthetic-enzymes in stem rust infected, resistant and susceptible near-isogenic wheat lines. *Physiol. Mol. Plant P* 33, 33–46. doi: 10.1016/0885-5765(88)90041-0

Nicholson, R. L., and Hammerschmidt, R. (2003). Phenolic compounds and their role in disease resistance. *Annu. Rev. Phytopathol.* 30, 369–389. doi: 10.1146/annurev.py.30.090192.002101

Pina, A., and Errea, P. (2008). Differential induction of phenylalanine ammonia-lyase gene expression in response to *in vitro* callus unions of prunus spp. *Plant Physiol.* 165 (7), 705–714. doi: 10.1016/j.jplph.2007.05.015

Singh, R. P., and Jha, P. N. (2017). The PGPR stenotrophomonas maltophilia SBP-9 augments resistance against biotic and abiotic stress in wheat plants. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.01945

Tang, D., Wang, G., and Zhou, J. M. (2017). Receptor kinases in plant-pathogen interactions: more than pattern recognition. *Plant Cell* 29 (4), 618–637. doi: 10.1105/tpc.16.00891

Van Eck, L., Schultz, T., Leach, J. E., Scofield, S. R., Peairs, F. B., Botha, A. M., et al. (2010). Virus-induced gene silencing of WRKY53 and an inducible phenylalanine ammonia-lyase in wheat reduces aphid resistance. *Plant Biotechnol. J.* 8 (9), 1023–1032. doi: 10.1111/j.1467-7652.2010.00539.x

Wang, K., Liao, Y., Xiong, Q., Kan, J., Cao, S., and Zheng, Y. (2016). Induction of direct or priming resistance against botrytis cinerea in strawberries by β-aminobutyric acid and their effects on sucrose metabolism. *J. Agric. Food Chem.* 64 (29), 5855–5865. doi: 10.1021/acs.jafc.6b00947

Wang, B., Song, N., Tang, C., Ma, J., Wang, N., Sun, Y., et al. (2019). PsRPs26, a 40S ribosomal protein subunit, regulates the growth and pathogenicity of puccinia striiformis f. sp. tritici. *Front. Microbiol.* 10. doi: 10.3389/fmicb.2019.00968

Wu, N., Ozketen, A. C., Cheng, Y., Jiang, W., Zhou, X., Zhao, X., et al. (2022). Puccinia striiformis f. sp. tritici effectors in wheat immune responses. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1012216

Wu, S., Shan, L., and He, P. (2014). Microbial signature-triggered plant defense responses and early signaling mechanisms. *Plant Sci.* 228, 118–126. doi: 10.1016/j.plantsci.2014.03.001

Xu, X., Ma, L., and Hu, X. (2019). Overwintering of wheat stripe rust under field conditions in the northwestern regions of China. *Plant Dis.* 103 (4), 638–644. doi: 10.1094/PDIS-06-18-1053-RE

Zhang, W., Gao, T., Li, P., Tian, C., Song, A., Jiang, J., et al. (2020). Chrysanthemum CmWRKY53 negatively regulates the resistance of chrysanthemum to the aphid macrosiphoniella sanborni. *Hortic. Res.* 7 (1), 109. doi: 10.1038/s41438-020-0334-0

Zhang, X., and Liu, C. J. (2015). Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol. Plant* 8 (1), 17–27. doi: 10.1016/j.molp.2014.11.001

Zheng, L., Zhao, J., Liang, X., Zhan, G., Jiang, S., and Kang, Z. (2017). Identification of a novel alternaria alternata strain able to hyperparasitize puccinia striiformis f. sp. tritici, the causal agent of wheat stripe rust. *Front. Microbiol.* 8. doi: 10.3389/fmicb.2017.00071

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 July 2023
DOI 10.3389/fpls.2023.1204016



OPEN ACCESS

EDITED BY
Milen I. Georgiev,
Bulgarian Academy of Sciences, Bulgaria

REVIEWED BY
Michael Deyholos,
University of British Columbia, Canada
Tatyana Gorshkova,
Kazan Institute of Biochemistry and Biophysics (RAS), Russia

*CORRESPONDENCE
Estelle Goulas
estelle.goulas@univ-lille.fr

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

RECEIVED 11 April 2023
ACCEPTED 20 June 2023
PUBLISHED 17 July 2023

CITATION
Chabi M, Goulas E, Galinousky D, Blervacq A-S, Lucau-Danila A, Neutelings G, Grec S, Day A, Chabbert B, Haag K, Müssig J, Arribat S, Planchon S, Renaut J and Hawkins S (2023) Identification of new potential molecular actors related to fiber quality in flax through Omics.
*Front. Plant Sci.* 14:1204016.
doi: 10.3389/fpls.2023.1204016



# Identification of new potential molecular actors related to fiber quality in flax through Omics

Malika Chabi[1†], Estelle Goulas[1*†], Dmitry Galinousky[1†], Anne-Sophie Blervacq[1], Anca Lucau-Danila[2], Godfrey Neutelings[1], Sébastien Grec[1], Arnaud Day[1,3], Brigitte Chabbert[4], Katharina Haag[5], Jörg Müssig[6], Sandrine Arribat[1], Sébastien Planchon[7], Jenny Renaut[7] and Simon Hawkins[1]

[1]Université de Lille, CNRS, UMR 8576 - UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, Lille, France, [2]Université de Lille, UMRT 1158 BioEcoAgro, Institut Charles Viollette, Lille, France, [3]Fibres Recherche Développement, Technopole de l'Aube en Champagne – Hôtel de Bureaux 2, 2 rue Gustave Eiffel, CS 90601, Troyes, France, [4]Université de Reims Champagne-Ardenne, INRAE, FARE, UMR A 614, Reims, France, [5]Fraunhofer-Institute for Manufacturing Technology and Advanced Materials IFAM, Bremen, Germany, [6]The Biological Materials Group, HSB – City University of Applied Sciences, Bremen, Germany, [7]Department of Environmental Research and Innovation, Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg

One of the biggest challenges for a more widespread utilization of plant fibers is to better understand the different molecular factors underlying the variability in fineness and mechanical properties of both elementary and scutched fibers. Accordingly, we analyzed genome-wide transcription profiling from bast fiber bearing tissues of seven different flax varieties (4 spring, 2 winter fiber varieties and 1 winter linseed) and identified 1041 differentially expressed genes between varieties, of which 97 were related to cell wall metabolism. KEGG analysis highlighted a number of different enriched pathways. Subsequent statistical analysis using Partial Least-Squares Discriminant Analysis showed that 73% of the total variance was explained by the first 3 X-variates corresponding to 56 differentially expressed genes. Calculation of Pearson correlations identified 5 genes showing a strong correlation between expression and morphometric data. Two-dimensional gel proteomic analysis on the two varieties showing the most discriminant and significant differences in morphometrics revealed 1490 protein spots of which 108 showed significant differential abundance. Mass spectrometry analysis successfully identified 46 proteins representing 32 non-redundant proteins. Statistical clusterization based on the expression level of genes corresponding to the 32 proteins showed clear discrimination into three separate clusters, reflecting the variety type (spring-/winter-fiber/oil). Four of the 32 proteins were also highly correlated with morphometric features. Examination of predicted functions for the 9 (5 + 4) identified genes highlighted lipid metabolism and senescence process. Calculation of Pearson correlation coefficients between expression data and retted fiber mechanical measurements (strength and maximum force) identified 3 significantly correlated genes. The genes were predicted to be connected to cell wall dynamics, either directly (Expansin-like protein), or indirectly (NAD(P)-binding Rossmann-fold superfamily protein). Taken together, our results have allowed

the identification of molecular actors potentially associated with the determination of both *in-planta* fiber morphometrics, as well as *ex-planta* fiber mechanical properties, both of which are key parameters for elementary fiber and scutched fiber quality in flax.

KEYWORDS

**flax (*Linum usitatissimum* L.), proteomics, transcriptomics, cell wall, fiber, retting, quality**

## 1 Introduction

Climate change is becoming a major concern, and strong decisions will have to be taken in the coming years to reduce global $CO_2$ production. In this context, a major challenge will be finding environmentally-friendly replacements for the fossil-fuel-based composites used in everyday life. Renewable materials containing plant fibers (natural fiber composites: NFCs) offer the double advantage of having a low environmental impact during production and being compatible with material end-of-life issues. Beyond the environmental arguments, natural fibers from plants have comparable and/or superior properties to synthetic fibers leading to increased use in several different industries, including building, transportation and sports (Bourmaud et al., 2018).

Flax (*Linum usitatissinum* L.) is an economically important fiber species grown for its long cellulose-rich fibers that are grouped together in fiber bundles located in the outer stem tissues, as well as for its seeds (Esau, 1943). It is one of the oldest cultivated plants and the current *L. usitatissinum* species is probably derived from pale flax *L. bienne* (Diederichsen and Hammer, 1995; Rahman and Hoque, 2023). There is evidence of flax fiber use in Neolithic cultures, and plants were later grown in Egypt about 6,000 years ago (Zohary and Hopf, 2004). However, whether flax selection was driven mainly for seed oil content or for fiber yield in the stems is still a matter of debate.

Previous work on flax-based NFCs has shown that their properties depend on several factors, including processing technique, polymer type and fiber quality (Haag et al., 2017). Defining "fiber quality" is difficult since it is related to what is really meant by the term "fiber" (Wrobél-Kwiatowska et al., 2010). The word fiber can refer to individual fiber cells - the botanical "elementary fibers" - or to groups of several elementary fibers glued together in which case the term "scutched fiber" is sometimes used. The fineness and the mechanical properties of both elementary and scutched fibers are key parameters used to define fiber quality, especially for textiles and/or NFCs. These parameters are linked to the plant genotype, environmental conditions, and the degree of elementary fiber separation obtained during the retting, scutching, and cleaning processes (Chabbert et al., 2020).

At the cellular level, both fineness and mechanical properties are related to the structural composition and the architecture of the fiber composite cell wall consisting of the middle lamella, the primary and gelatinous secondary cell wall characterized by high amounts of crystalline cellulose and very low lignin levels (Morvan et al., 2003; Day et al., 2005; Bourmaud et al., 2013). Although some broad correlations can currently be made between cell wall structure and fiber quality (e.g., negative relationship between lignin content and textile fiber quality), it is likely that many other more or less subtle variations in the cell wall structure and metabolism could contribute to differences in fiber properties. At the molecular level, approximately 10% of the genome is linked to cell wall construction, dynamic architecture, and sensing functions (Carpita, 2011), and a major challenge is to identify which ones are involved in this process and to understand their roles. Globally, flax is a well-studied plant from a molecular genetics point of view – its genome has been sequenced (Wang et al., 2012; You and Fofana, 2023; You et al., 2023) and chromosome pseudomolecules have been assigned (You et al., 2018), an integrated consensus genetic and physical map was constructed (Cloutier et al., 2012), and the karyotypes of common flax (*Linum usitatissimum* L.) (Rachinskaya et al., 2011) and wild species were also investigated (Muravenko et al., 2001). Moreover, the flax plant is a convenient object to study cell wall development because the stem contains both cell walls rich in cellulose (bast fibers) and lignin (xylem cells) (Day et al., 2005; Gorshkova et al., 2012; Huis et al., 2012; Petrova et al., 2022). Several previous studies on flax using Omics together with biochemical and imaging approaches have identified changes in the expression of numerous genes (Hotte and Deyholos, 2008; Fenart et al., 2010; Huis et al., 2012; Zhang and Deyholos, 2016; Chabi et al., 2017; Gorshkov et al., 2017; Gorshkova et al., 2018; Simon et al., 2018; Gorshkova et al., 2023; Somalraju and Fofana, 2023). In many cases, the interpretation of the results was based on the hypothesis that changes in the expression of genes related to the biosynthesis of major cell wall polymers (e.g., cellulose, hemicellulose, pectins and lignin) could be related to differences in cell wall structure and/or fiber properties. However, while such an hypothesis is no doubt basically correct, it is also possible that changes in the expression of other, less obvious genes (i.e., not apparently related to cell wall metabolism) could affect key fiber parameters such as fineness and mechanical properties.

In an attempt to obtain further information about underlying molecular processes potentially related to flax fiber quality, we

combined Omics and morphometric/mechanical analyses in bast fiber bearing tissues (a complex mixture of tissues that mainly contains bast fibers, but also epidermis, cortical parenchyma and phloem sieve tubes) of seven flax varieties belonging to three contrasted morphotypes: fiber spring (Belinka, Diane, Drakkar, Hermes), fiber winter (Adelie and Violin) and oil winter (Oliver) grown under field conditions. A panel of statistical analyses allowed us to highlight a limited number of relevant central actors of *in-planta* and *ex-planta* fiber quality. Their role in the context of elementary fiber development and fiber bundle quality is discussed.

# 2 Materials and methods

## 2.1 Plant growth and sample collection

The seven flax cultivars (Diane, Hermes, Drakkar, Belinka, Adelie, Violin, and Oliver) were grown under field conditions in plots at Terre de Lin (TdL), Saint-Pierre-le-Viger, France, except for the cultivar Violin, cultivated by LINEA, Grandviliers, France. All retting and scutching were performed by Terre de Lin. For Omics and morphometrics, plant samples were col2lected from three randomly distributed areas of the field at the budding stage. Five-cm-long fragments were removed from stems at 5 cm above the cotyledons and outer tissues (peeled off fiber-enriched outer part of stem) were manually separated from the inner tissues and frozen in liquid nitrogen before storing at -80° C until used for transcriptomic and proteomic analyses. The peeled off tissues are therefore a complex mixture of tissues that contain, besides bast fibers, at least epidermis, cortical parenchyma and phloem sieve tubes. A 1-cm long fragment was removed from the stem at 10 cm above the cotyledons from the same plants for morphometric analyses. For analyses of mechanical properties, plants were pulled (harvested) at maturity, field-retted, and scutched to produce fiber bundles corresponding to the length of the flax stems.

## 2.2 Morphometric analyses

Measurements were performed on one-cm stem fragments, collected 10 cm above the cotyledon from the same plants used for Omics approaches. Eighty-one fiber bundles [3 fiber bundles x 3 cross-sections x 3 individuals x 3 field areas] per cultivar were examined. Two hundred and seventy measurements [10 measures x 3 cross-sections x 3 individuals x 3 field areas] per variety were made to determine the fiber cell surface, including the G-layer of secondary cell wall also called tertiary cell wall (but excluding the common middle lamella with the adjacent fiber).

Preliminary observations were made on freehand cross-sections, without fixation in ethanol, and double stained with Alcian Blue (0.1%, w/v in water, 5 min), followed by Safranin O (0.1%, w/v in EtOH, 30 secs), mounted in water, and observed on an Olympus BH-2 light microscope. Remaining stem fragments were fixed in 80% ethanol, then dehydrated using a graded ethanol series. According to the manufacturer's protocol, three individual stem fragments per field area were embedded in Technovit 7100 (Hereaus). Semi-thin sections (5 µm) were obtained with a Leica RM2065 microtome with glass knives. Three spatially separated cross-sections (every 200 µm) were selected for each individual for measurements. Toluidine-Blue O (TBO) coloration (0.1%, w/v in water) was applied on semi-thin sections to facilitate visual separation between the inner and outer tissues, allowing discrimination of fiber bundles and cell types. Measures were acquired with FIJI free software[1]. Cell surface and number of elementary fibers per bundle were further referred to as "*in-planta*" features.

## 2.3 Mechanical analyses of scutched fiber bundles

The following settings were used to characterize the tensile mechanical properties of flax fiber bundles from the 7 flax varieties (55 fiber bundles per variety were tested). The force-displacement curves were measured with Fafegraph M (Textechno GmbH, Mönchengladbach, DE; clamping length: 20 mm; test speed: 20 mm/min; pretension: 250 mg) (Haag and Müssig, 2016).

The thickness of the scutched fiber bundles of the 7 studied flax varieties were measured with at least 15 000 individual values by the scanner-based image analysis system FibreShape 5.1 (IST AG, Vilters, CH). The fiber bundles were cut to a length of approx. 20 mm and distributed on the scanner surface. The images were taken at a resolution of 2400 dpi (Epson Perfection V700, Epson GmbH, Meerbusch, DE). Six images per fiber batch were measured with the measuring mask Long_fibres_2400dpi_Epson_Bast_uncalib in the FibreShape software (minimum element width: 11 µm). The cross-sectional area was calculated from the thickness value of the fiber bundle using a SEM based correction model (Haag and Müssig, 2016):

Area = $d^{1.7256}$ * 0.21113, where d – the thickness of the scutched fiber bundle.

Strength, Young's modulus, strain, maximum force, and thickness measurements were further refered as "*ex-planta*" features.

## 2.4 Omics analyses

### 2.4.1 Transcriptomics

#### 2.4.1.1 RNA extraction and quality verification

Total RNA was extracted from three pools of outer stem tissues for each of the 7 varieties using TRI-Reagent (MRC, Inc., Dundee, Scotland, UK). RNA integrity and concentration were evaluated with RNA StdSens Chips using the Experion automated electrophoresis system (Bio-Rad, Marnes-la-Coquette, France).

#### 2.4.1.2 Hybridization and data analyses

RNA processing, Cy3-labeling and hybridization were made following the manufacturer's instruction for One-Color Microarray base gene Expression Analysis (Agilent Technologies). Hybridization was performed on Agilent microarrays Agilent-

1 https://imagej.net/software/fiji/

045382 UGSF flax 45K v1.0 array based upon flax genome coding sequence (*Lus-names*) (Wang et al., 2012) available at Phytozome[2]. Some genes that have genolin-names (Fenart et al., 2010) were added according to the EST database. The 4-plex array contains 45,220 60-mer *in situ* oligonucleotides per block. Hybridization and washing were performed following Agilent manufacturer's instruction, and slides were immediately scanned at 5 mm pixel$^{-1}$ resolution using an Axon GenePix 4000B scanner driven by GenePix Pro 6.0 software (Molecular Devices Corporation, Sunnyvale, CA, USA).

Scanned images (TIFF format) were then submitted to grid alignment and expression data analyses. For each slide, a lowest and print-tip median normalization was performed using R packages[3] as implemented in CLC bio software[4] followed by an inter-slide normalization. The 3 control samples were filtered for p-value < 0.05, and the average was calculated for each gene. A fold change (FC) value was calculated between individual treated samples and the mean of corresponding controls. Differentially expressed genes (DEGs) were selected for a threshold >2.0 or ≤0.5. Functional annotation of DEGs was based on NCBI GenBank, and related genes' physiological processes were assigned with NCBI, AmiGO 2 Gene Ontology and UniProt. KEGG pathway analysis was also used to identify relevant biological pathways of selected genes.

### 2.4.2 Proteomics

#### 2.4.2.1 Soluble protein extraction

Four grams of each of the three pools of frozen outer stem material were ground in liquid nitrogen to a fine powder, followed by 5 min grinding in 10 ml of Tris HCl buffer 50 mM, 0.06% IPC, pH 7.5, before centrifuging 10 min at 4°C/16,000 g. The supernatant was incubated for 15 min at room temperature with protamine sulfate under a low agitation, then centrifuged for 10 min at 18,000 g. The pellet was discarded, and the proteins were precipitated with 10% TCA (mass/v), 1 hour at -20°C. The pellet was washed once with cold acetone, and then dried for 5 min at room temperature (Goulas et al., 2001).

#### 2.4.2.2 Protein labeling

Dry protein pellets were resolved in 2D-DIGE labeling buffer (7 M urea, 2 M thiourea, 4% (mass/v) CHAPS and 30 mM Tris) at room temperature with agitation for 1 hours. The pH of the solution was adjusted to 8.5, and protein concentration was determined by 2D Quant kit (GE Healthcare). The protein samples were labeled with CyDye minimal dyes (GE Healthcare) according to the manufacturer's instructions and as described in Grimaud et al. (2013). 30 µg of proteins were labeled by adding 240 pmol of fluorochromes (Cy3 or Cy5). A pooled internal standard was performed by mixing 15 µg of each sample and labeled with Cy2 dye and included in all gel runs (Goulas et al., 2006).

#### 2.4.2.3 Electrophoresis

For each 2-D gel a mixture with a final volume of 120 µl containing two samples labeled Cy3 and Cy5, internal standard labeled with Cy2, and labeling buffer [7 M urea, 2 M thiourea, 4% (mass/v) CHAPS, 6 µl.ml$^{-1}$ DeStreak reagent (GE Healthcare)] and 1.5% (v/v) Biolyte, were loaded onto a 24 cm IPG non-linear strip with a pH range 3–10 (Bio-Rad) and separated using the IPGphor3 system with a rapid voltage slope to reach the maximum of 10 000 V and a total of 72 000 Vh. Second dimension resolution was carried out by SDS-PAGE on a 12.5% (v/v) resolving gel (HPE, Gel company) according to 2D HPE Large Gels manufacturer's instructions.

#### 2.4.2.4 Image analysis and protein identifications

Scanning of the 2D-DIGE gels was carried out using a Typhoon Imager 9400 (GE Healthcare) at three different wavelengths corresponding to the different CyDyes as described in Grimaud et al. (2013). The generated gel images were analyzed with DeCyder 7.0 software, using the Cy2 channel as a standard. The multiple 2-D DIGE gels were matched, and the significance of changes in the abundance of specific proteins from different genotypes was established. By including the internal standard on each gel in the experiment, together with the individual biological samples, it is possible to measure the abundance of each protein spot on a gel relative to its corresponding spot in the internal standard present on the same gel (i.e. as a ratio). All matched spots were manually verified on all gels.

#### 2.4.2.5 Enzymatic digestion and protein identification

Differentially abundant spots (based on the statistical analyses described below) were excised from the gels and digested using an EVO2 Workstation (TECAN). All MS and MS/MS analyses were carried out by using a 5800 MALDI TOF/TOF (AB Sciex, Foster City CA, USA) (Chabi et al., 2017), and Protein identification was performed using MASCOT server with a *Linum* database downloaded from phytozome (Lusitatissimum_200_V1.0 with 43484 sequences; 17356610 residues) according to (Wang et al., 2012).

## 2.5 Bioinformatics and statistical analyses

### 2.5.1 Data deposition

Transcriptomic data have been deposited to the NCBI online repository with accession number GSE222066. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[5] partner repository (Perez-Riverol et al., 2022) with the dataset identifier PXD039389.

### 2.5.2 Transcriptomic analysis

A script utilizing library functions in R with a Bonferroni-corrected p-value of less than 0.05 and a cut-off of +/- 1.5 log2 ratio was used for all experimental conditions to identify genes displaying a significant change in expression (differently expressed gene DEG) over the repetitions. Only genes with smooth expression profiles were retained. Differential analyses were performed for each variety vs the other 6 varieties giving 21 possible combinations. The

2 https://phytozome-next.jgi.doe.gov

3 http://cran.r-project.org/

4 https://digitalinsights.qiagen.com/

5 https://www.ebi.ac.uk/pride/archive

differentially expressed genes and differently abundant proteins were annotated according to KEGG database using ShinyGO online tool[6] (Ge et al., 2020). The MapMan framework for functional classification (Schwacke et al., 2019) with Mercator4 v.2 online tool[7] was also used to assign the protein sequence annotation.

### 2.5.3 Proteomic analysis

Statistical analyses were performed with the Extended Data Analysis (EDA) module in DeCyder 2D Differential Analysis Software v.7.0 (Goulas et al., 2006). For every set of matched spots, the average intensity was compared to the standard deviation on these intensities for a given spot between the two compared varieties. Only spots present on all gel images from each set of samples were considered. One-way ANOVA analysis was performed for every set of matched spots, comparing the average and standard deviation for a given spot between Diane and Oliver groups to find significant differences (ANOVA, p-value ≤ 0.05; average ratio >1.5) above the experimental variation.

### 2.5.4 Global statistical analyses

Sparse Partial Least-Squares Discriminant Analysis (sPLS-DA) analyses were conducted by using RStudio software (2022.02.03 build 492) and the R-package mixOmics (v. 6.1.2). The PLS method searches for the highest covariance between variates, which are a linear combination of initial variables, taking into account all variables in the input matrices. sPLS combines this approach with a Lasso penalization to enable the selection of variables the most explanatory of varieties discrepancy. The arithmetic means of the gene expression value and their log2 transformation were used for statistical analysis. Morphometric and mechanical data were tested for normality by the Shapiro-Wilk test. We used the ANOVA test, Pearson correlation coefficient, Student's t-test for normally distributed data, Spearman correlation, and the Kruskal-Wallis test if the normality test was not met. We used the R lm-function to generate the regression plots. The calculated determination coefficient ($R^2$) was used to estimate the proportion of variation of the dependent variance that could be explained by the variation of the independent variance.

## 3 Results

### 3.1 Morphometric characterization of the 7 varieties, in-planta features

We determined the number of elementary fiber cells per bundle and fiber surface in four flax spring fiber varieties (Belinka, Diane, Drakkar, Hermes), two winter fiber varieties (Adelie, Violin) and one winter oil variety (Oliver) (Figure 1). The results show a large scale of variation for both parameters with all fiber spring varieties showing the highest values and fiber winter varieties the lowest. The oil variety Oliver showed intermediate values.

### 3.2 Transcriptomic analysis of 7 flax varieties and correlation with morphometric parameters

We evaluated the abundance of 42 963 transcripts (99.3%) among the 43 384 annotated genes in the flax genome[8]. Comparative whole transcriptomics of the fiber-bearing outer-stem tissues from the 7 varieties allowed us to identify 1041 differently expressed genes (DEGs) that showed significant differences in transcript accumulation (Figure 2; Supplementary Table 1). According to MapMan annotation, only 40 of them were classified into cell wall organization (Supplementary Table 1, bold characters). However, further expert curation allowed us to highlight 57 more genes related to cell wall metabolism (Supplementary Table 2). These 1041 DEGs were also annotated according to KEGG categories (Figure 2). Our results show that glucosinolate-, diterpenoid- and zeatin-biosynthesis are pathways in which the genes are the most overrepresented (up to 8-fold). Circadian rhythm, tryptophan, starch, sucrose metabolism, and phenyl biosynthesis also showed significant enrichment (approx. 4-fold). Finally, biosynthesis of secondary metabolites and metabolic pathways showed less fold-enrichment (approx. 2-fold), but with a much higher number of significant DEGs.

All 1041 significant DEGs were processed with mixOmics R-package to Principal Component Analysis (PCA, Figure 3A-i), and Sparse Partial Least-Squares Discriminant Analysis (sPLS-DA, Figure 3A-ii). As shown in Figure 3A-i, the PCA scree plot showed that 75% of the total variance was explained with the first three principal components (43, 18, and 14% of the total variation, respectively). Thus, the 3 principal components were selected to further sPLS-DA (Figure 3A-ii) which showed that 73% of the total variance was explained by the first three X-variates (37, 19 and 17%, respectively), corresponding to 56 DEGs. When considering the X-variate 1, a clear discrimination was noticed between the first cluster, composed of the two winter varieties, one of them is fiber flax (Violin), and the other one is oil (Oliver), and all the other varieties. The X-variate 2 showed discrimination between this first cluster and a second one composed of two spring (Drakkar, Hermes) and one winter (Adelie) fiber varieties, a third composed of one spring fiber variety (Belinka), and a fourth one composed of another spring fiber variety (Diane). Such clusterization was further confirmed by hierarchical clusterization on Euclidian distance (Figure 3A-iii).

The contribution of each variable as a component of sPLS-DA indicate the importance of that variable for discrimination. We identified a total of 56 DEGs transcripts that have higher contribution values than others of the studied transcripts (Supplementary Table 3), 17 transcripts for X-variate 1 (Figure 3B-i), 19 transcripts for X-variate 2 (Figure 3B-ii), and 20 transcripts for X-

6 http://bioinformatics.sdstate.edu/go/

7 https://www.plabipd.de/portal/web/guest/mercator4

8 https://phytozome-next.jgi.doe.gov

10.3389/fpls.2023.1204016

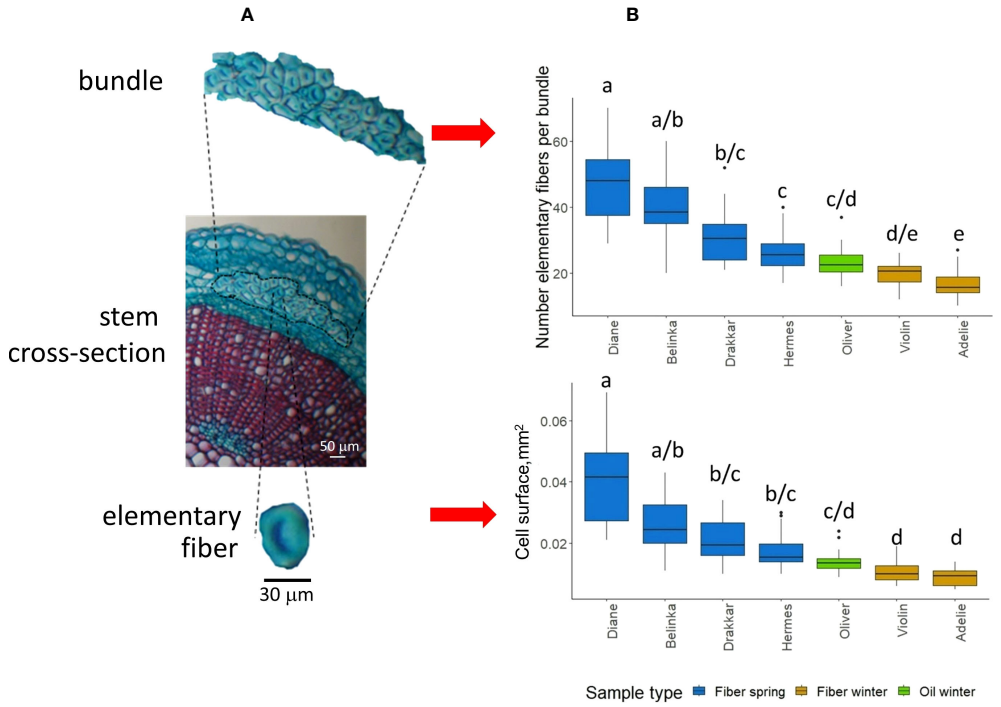



FIGURE 1
Flax fiber morphometric properties. **(A)** Light microscope image showing flax stem cross-section (middle), fiber bundle (top), and a single fiber cell (bottom). Sample stained in safranin O/alcian blue. **(B)** Box-plot of two morphometric features, the number of elementary fibers per bundle (upper), and the cell surface of single fibers within fiber bundles (lower) for the seven flax varieties (Diane, Hermes, Drakkar, Belinka, Adelie, Violin, and Oliver). Letters shows statistical differences between flax varieties, p-value ≤ 0.05 (Dunn's test).

variate 3 (Figure 3Biii; Supplementary Table 3). The DEGs that were the most explanatory for sPLS-DA X-variate 1 were down-regulated only in Oliver and Violin, whereas others were up-regulated in Drakkar, Belinka, Hermes, and Adelie. For X-variate 2, the explanatory DEGs are almost all up-regulated in Belinka only, with the exception of a single down-regulated gene in Adelie. For X-variate 3, the most explanatory DEGs were down-regulated, especially in Belinka, or up-regulated only in Diane.

The variation of morphometric parameters was calculated within each of the four clusters (Figure 3C) highlighted after analysis of the 1041 DEGs (Figures 3Aii, iii). Based on the transcriptome analysis, we discriminate groups of varieties that differ significantly in morphometric traits. They showed significant differences in the studied phenotypic traits between the four clusters. The varieties within the 4th cluster have more individual fibers in bundles, and their cell diameter was the highest. And *vice versa*, the varieties joined in the first cluster carried fewer single cells in fiber bundles, and the cells were smaller.

The Pearson correlation coefficient between the expression level of the 56 genes highlighted after sPLS-DA analysis of bast fiber transcriptome and values of morphometric features (the number of elementary fiber cells in the bundle and the single cell surface) of the same varieties were calculated (Figure 4). We identified 5 genes whose expression level had a strong statistically significant correlation with studied morphometric features (maximum value 0.84; minimum value -0.86) (Figure 4). One gene was identified as the *LusWRKY85* (*Lus10022736*) gene, which is a transcription factor of the WRKY family, and it is orthologous to the *Arabidopsis AtWRKY62* (*AT5G01900.1*) (Yuan et al., 2021). Another gene is *LusS40-6* (*Lus10002073*), and its *Arabidopsis*'s orthologue *AtS40-6* (*AT1G61930*) codes senescence regulator (Fischer-Kilbienski et al., 2010). The next gene in the list is *LusGLYI6* (*Lus10038612*), and it is also negatively correlated with cell surface (Figure 4). The *Lus10038612* gene has homology with *AtGLYI6 (AT1G67280.1)*, which encodes a $Ni^{2+}$-dependent glyoxylase (Batth et al., 2020). Finally, two genes (*Lus10039365* and *Lus10017734*) with unknown functions were also highlighted (Figure 4).

## 3.3 Proteomic analysis of bast fibers from a winter oil and a spring fiber variety and correlation with morphometric parameters

A comparative proteomic analysis was performed on the peeled off fiber-enriched outer stem parts of Diane and Oliver since these varieties belonged to the two transcriptomics clusters showing the most discriminant and significant differences regarding morphometric measurements (Figure 3). A total of 1490 protein spots were identified on the 2D gel. Among these, 108 showed significant differential abundance (p.value ≤ 0.05) with at least a 1.5-fold change between the two varieties (7.25% of protein spots in the gel). Subsequent excision, trypsin digestion, and mass spectrometry allowed the successful identification of 46 proteins representing 32 non-redundant proteins (Table 1), among which 15 of them were more abundant in Diane and 17 in Oliver (Supplementary Figure 1). One peptide spot (Master number 1046) that was more

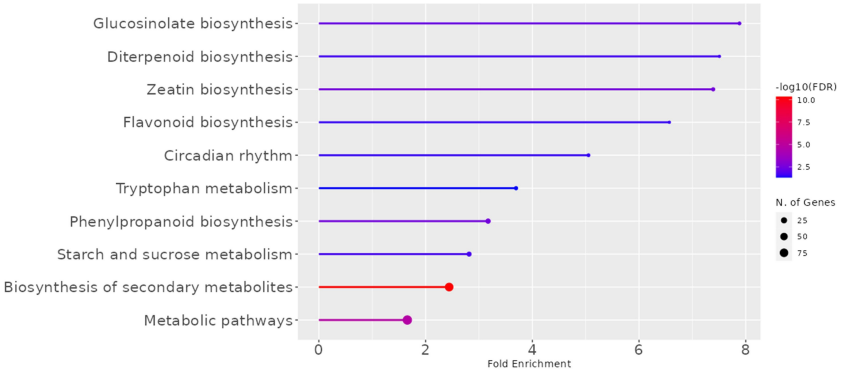



FIGURE 2
KEGG Pathways enriched in the transcriptome of fiber-bearing outer-stem tissues. The 1041 DEGs of the 7 varieties (Diane, Hermes, Drakkar, Belinka, Adelie, Violin, and Oliver) were selected, and their *Arabidopsis thaliana* orthologues were compared to the gene list of the complete *Arabidopsis* genome. Line length corresponds to the fold of enrichment. The Line color indicates the FDR value. The size of the end spot shows the number of genes from the 1041 *LusID* DEGs list that are identified in the corresponding *A. thaliana* KEGG-category.[10]

abundant in the Diane proteome represents a mixed sample, which included two peptides identified like ascorbate peroxidase 1 and ferretin 1 (Table 1; Supplementary Figure 1). Both of them were included in further analysis. The transcript annotations of genes coding the 32 identified proteins were classified according to MapMan annotation, and only 1 of them could be classified directly into cell wall organization, *Lus10037377*, a xyloglucan endotransglucosylase/hydrolase 32 [homologous to ELF4-like in *Arabidopsis thaliana* genome (*AT2G36870.1*)] (Supplementary Tables 1, 2).

Interestingly, both in Diane and in Oliver, proteins with highest variations in abundancy ratio (Table 1) were two different protein spots (Master number 1269 and 1284, respectively) identified as the same polyketide cyclase/dehydrase and lipid transport superfamily protein (Lus10008932), roughly twice less in Diane (4.07) than in Oliver (9.78). This fact may indicate an alternative modification of the Lus10008932 protein in the proteome of two studied flax varieties. Such 2-fold range differences in ratio abundancy were also observed for a methyl esterase 4 (Lus10009205), a xyloglucan endotransglucosylase/hydrolase 32 (Lus10037377), a NAD(P)-binding Rossmann-fold superfamily protein (Lus10010846) and a cold, circadian rhythm, and RNA binding 2 protein (Lus10023569) whose ratio variation was around 2.3 in Diane whereas a unassigned protein (Lus10014003), a rotamase CYP 3 (Lus10012167), a beta glucosidase 32 (Lus10030577) and a cyclase family protein (Lus10001508) showed a ratio variation of around 4.5 in Diane. Nine proteins in Diane and twelve in Oliver showed variation in abundance ratio of roughly 1.5. It has to be noticed that the four proteins directly or indirectly linked to cell wall development were differently abundant in the studied varieties. Two of them (Lus10009205 – methyl esterase 4 and Lus10037377 – xyloglucan endotransglucosylase/hydrolase 32) were more abundant in Diane (2.54 and 2.38 times, respectively), and the other two (Lus10030577 – beta glucosidase 32 and gi|171462119 – Expansine-related protein 3 precursor) – in Oliver (4.03 and 1.80 times, respectively) (Table 1).

Since proteomic data was only available for two varieties it was not possible to evaluate the relative weight of the 32 identified proteins as potential markers for fiber-related parameters in the complete panel of 7 varieties. The expression profiles of the genes corresponding to the 32 candidate proteins IDs to gene IDs were therefore further established for the 7 flax varieties studied. Clusterization based on their expression level showed clear discrimination into three separate clusters (Figure 5A). The first cluster differentiated spring fiber varieties Diane, Belinka, Drakkar, and Hermes from winter fiber varieties Adelie and Violin (the second cluster). The winter oil flax Oliver formed the most distant cluster (the third one) from the other flax varieties. It should be noted that the cluster of the spring fiber varieties combines samples with the highest amount of fiber cells in the bundle and the biggest cell surface (Figure 5B).

Pearson correlation coefficients were further calculated between the expression levels of 32 genes based on the results of the proteomic study and the morphometric features of the same varieties (Figure 6). We identified 4 genes whose expression level strongly correlated with studied morphometric features (maximum value -0.72; minimum value -0.93) (Figure 6). Two genes (*Lus10010846* and *Lus10024383*) encoding putative 3-beta-hydroxysteroid dehydratase/isomerases were identified as putative lipid-binding proteins in phloem exudates of *Arabidopsis* plants (Guelette et al., 2012). Another gene (*Lus10008932*) potentially related to lipid metabolism was also identified (Figure 6); its product is described as a plant-specific protein (Gutierrez et al., 2004).

## 3.4 *Ex-planta* mechanical features of scutched fibers and correlation with gene candidates and in- planta morphometric parameters

Mechanical (strength, Young's modulus, strain, maximum force) and geometric (thickness) measurements, *ex-planta* features, were made on scutched fiber bundles obtained from the 7 varieties (Supplementary Table 4). Statistically significant

10 http://bioinformatics.sdstate.edu/go74/

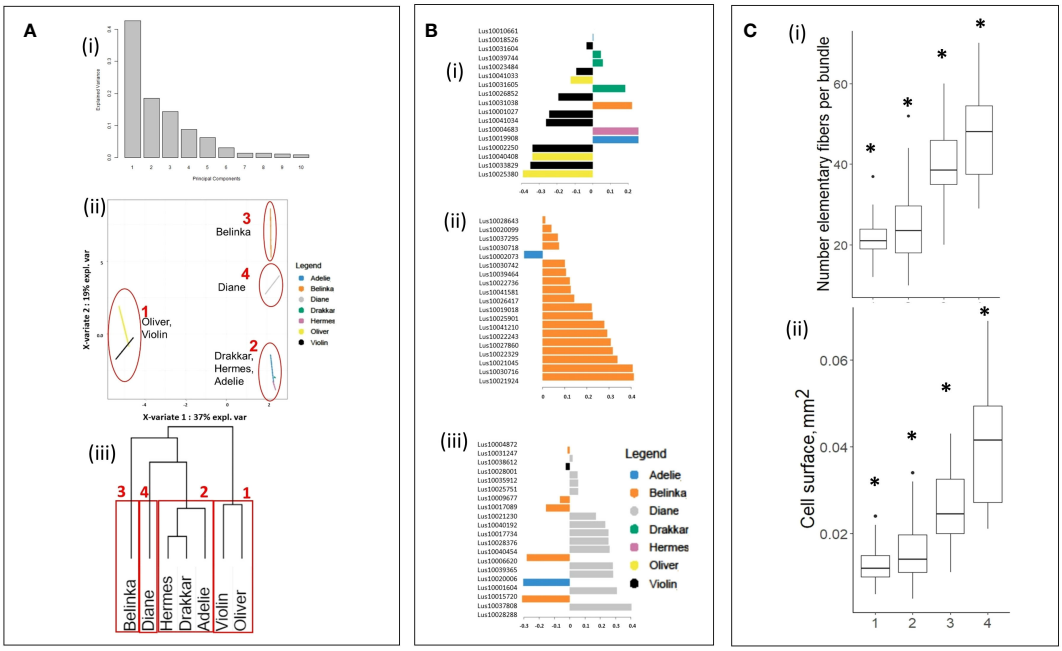



FIGURE 3
Statistical clusterization of the 7 flax varieties based upon DEG values from transcriptomic analysis. **(A)** (i) Scree plot explained variance of PCA, (ii) plot of sparse Partial Least-Squares Discriminant Analysis (sPLS-DA), and (iii) hierarchical clusterization on Euclidian distance[11], **(B)** Contribution plots of the DEGs on X-variates 1 (i), 2 (iii) and 3 (iii) of sPLS-DA, **(C)** Boxplots of (i) number of single fiber cells and (ii) cell surface of fiber cells within clusters highlighted in **(A)**. Asterisks indicate statistically significant differences (p-value ≤ 0.05) between groups..

differences for strength, strain, and maximum force were found at 95% of confident interval (Figures 7A, C, D), as for the thickness of the scutched fiber bundle (Figure 7E). We noticed significant differences of Young's modulus between the seven varieties when we accepted 90% level of the confidence interval (not at 95%). The scattering of mechanical values is a common observation with plant fibers and it is related to the variability of these natural material. There is indeed a large range of factors which can affect variability, from plant growth conditions to fiber sampling, processing and testing (Baley et al., 2020). The average calculated values for each variety are in accordance with the results reported previously (Bourmaud et al., 2018).

We combined the two lists of candidate genes that we identified according to previous transcriptomic (Supplementary Table 3) and proteomic (Table 1) analyses to calculate Pearson correlation coefficients between expression levels and mechanical values measured for the same varieties. We identified 3 genes whose expression level was strongly correlated with mechanical properties (Table 2). One of the candidate genes is an expansin-like protein (*gi|1714621119*) which is connected to cell wall metabolism (Roach and Deyholos, 2008). The two other genes (*Lus10011125* and *Lus10010846*) encode an enzyme that catalyzes the oxidative decarboxylation of L-malate, the NADP-malic enzyme 3, and an NAD(P)-binding Rossmann-fold superfamily protein, also identified as a putative 3-β-hydroxysteroid dehydratase/isomerase in *Arabidopsis* phloem exudates (Guelette et al., 2012).

To see whether those *ex-planta* mechanical values (Figure 7) of scutched fibers could be related to *in-planta* fiber morphometric data (Figure 1), we calculated Pearson correlation coefficients with cell surface and number of fibers within bundles. The bivariate correlations between the 2 sets of morphological data (Figure 1) and the 5 sets of mechanical data (Figure 8) are shown on regression plots (Figure 8). The regression lines reflect positive correlations between both morphometric features and Young's modulus (PC of 0.8 and 0.77, respectively) and both morphometric features and strength (PC of 0.61 and 0.60, respectively), whereas no correlation could be observed with strain, maximum force or with the area of scutched fibers. We also did not find any significant correlation between fiber bundle surface *in-planta* (calculated as the multiplication of cell surface and number elementary fibers per bundle) and scutched fiber area *ex-planta* (data not shown).

## 4 Discussion

The fineness and mechanical properties of fibers are key quality parameters for textiles and NFCs that are determined by two main processes in addition to scutching: i) fiber formation during the growth of the flax plant and ii) fiber separation and extraction during the post-harvest retting process (Djemiel et al., 2020). Both of these processes are, in turn, determined by the flax genotype and environmental conditions (including the composition of the retting microbiome). In this paper, we combined Omics performed on bast fiber bearing tissues and morphometric/mechanical analyses in seven flax varieties grown under field conditions to provide

11 https://software.broadinstitute.org/morpheus/

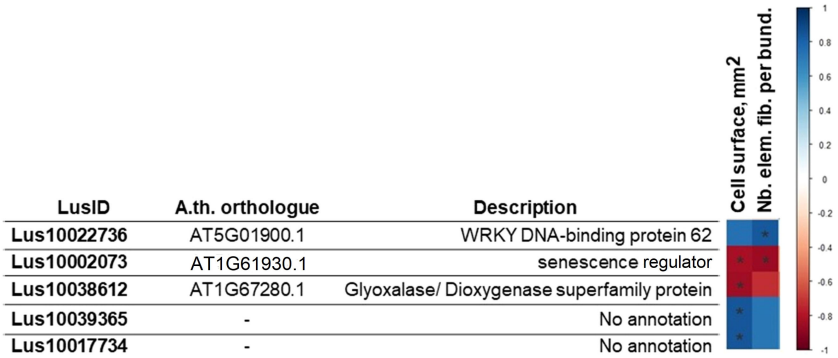



FIGURE 4
Pearson correlation coefficient (PC) between the expression level of genes selected on transcriptomic data analysis and morphometric (the surface of a fiber cell, $mm^2$; the number of fibers in the bundle). The p-value for each PC was calculated, and the PC with p-value ≤ 0.05 are given in the table. The red cell color marks a negative correlation, the blue one a positive one. * - p-value of PC ≤ 0.05.

**TABLE 1** Proteins and their gene IDs differently abundant in Diane and Oliver flax varieties.

| N | Transcript | Abundance ratio | P-value | Master numb. | Arabidopsis ortholog | Gene abbreviation | Pfam domen (s) | Definition according to *At* genome |
|---|---|---|---|---|---|---|---|---|
| | | | *Proteins (and their genes) that are more abundant in Diane* | | | | | |
| 1 | Lus10008932 | 4.07 | 2.6E-05 | 1269 | AT4G14060.1 | | PF00407 | Polyketide cyclase/dehydrase and lipid transport superfamily protein |
| 2 | Lus10009205 | 2.54 | 5.2E-05 | 953 | AT2G23580.1 | ABE4, ATMES4, MES4 | | methyl esterase 4 |
| 3 | Lus10037377 | 2.38 | 1.2E-02 | 859 | AT2G36870.1 | XTH32 | PF00722, PF06955 | xyloglucan endotransglucosylase/hydrolase 32 |
| 4 | Lus10010846 | 2.09 | 6.4E-04 | 890 | AT2G37660.1 | | PF01073 | NAD(P)-binding Rossmann-fold superfamily protein |
| 5 | Lus10023569 | 2.03 | 1.5E-03 | 1350 | AT2G21660.1 | ATGRP7, CCR2, GR-RBP7, GRP7 | PF00076 | cold, circadian rhythm, and rna binding 2 |
| 6 | Lus10015049 | 1.99 | 2.8E-06 | 941 | AT5G02790.1 | GSTL3 | PF00043 | Glutathione S-transferase family protein |
| 7 | Lus10041602 | 1.75 | 1.1E-02 | 1125 | AT3G52150.1 | | PF00076 | RNA-binding (RRM/RBD/RNP motifs) family protein |
| 8 | Lus10023180 | 1.67 | 3.6E-06 | 1307 | AT1G65980.1 | TPX1 | PF08534 | thioredoxin-dependent peroxidase 1 |
| 9 | Lus10026241 | 1.63 | 7.2E-03 | 797 | AT5G36700.1 | ATPGLP1, PGLP1 | | 2-phosphoglycolate phosphatase 1 |
| 10 | Lus10010340 | 1.58 | 4.8E-06 | 639 | AT1G04420.1 | | PF00248 | NAD(P)-linked oxidoreductase superfamily protein |
| 11 | Lus10031708 | 1.56 | 4.3E-05 | 183 | AT3G60750.1 | | PF02779, PF02780, PF00456 | Transketolase |
| 12 | Lus10013537/ Lus10012499, Lus10022673 | 1.55 | 4.7E-03 | 1046 | AT1G07890.1/ AT5G01600.1 | APX1, ATAPX01, ATAPX1, CS1, EE6/ ATFER1,FER1 | PF00141/ PF00210 | ascorbate peroxidase 1/ ferretin 1 |
| 13 | Lus10017158 | 1.53 | 1.3E-02 | 1354 | AT5G57040.1 | | PF00903 | Lactoylglutathione lyase/ glyoxalase I family protein |

*(Continued)*

TABLE 1 Continued

| N | Transcript | Abundance ratio | P-value | Master numb. | Arabidopsis ortholog | Gene abbreviation | Pfam domen (s) | Definition according to *At* genome |
|---|---|---|---|---|---|---|---|---|
| 14 | Lus10024383 | 1.50 | 3.2E-02 | 870 | AT2G37660.1 | | PF01073 | NAD(P)-binding Rossmann-fold superfamily protein |
| | | | *Proteins (and their genes) that are more abundant in Oliver* | | | | | |
| 1 | Lus10008932 | 9.78 | 1.0E-05 | 1284 | AT4G14060.1 | | PF00407 | Polyketide cyclase/dehydrase and lipid transport superfamily protein |
| 2 | Lus10014003 | 5.81 | 5.0E-06 | 1449 | | | | |
| 3 | Lus10012167 | 4.52 | 2.6E-05 | 1227 | AT2G16600.1 | ROC3 | PF00160 | rotamase CYP 3 |
| 4 | Lus10030577 | 4.03 | 1.5E-04 | 198 | AT5G24550.1 | BGLU32 | PF00232 | beta glucosidase 32 |
| 5 | Lus10001508 | 3.63 | 6.1E-06 | 878 | AT4G34180.1 | | PF04199 | Cyclase family protein |
| 6 | gi\|171462119 | 1.80 | 3.4E-04 | 1420 | | | | Expansin-related protein 3 precursor |
| 7 | Lus10043025 | 1.77 | 6.6E-04 | 249 | AT5G25880.1 | ATNADP-ME3, NADP-ME3 | PF00390, PF03949 | NADP-malic enzyme 3 |
| 8 | Lus10042468 | 1.66 | 4.7E-06 | 1078 | AT1G78380.1 | ATGSTU19, GST8, GSTU19 | PF02798, PF00043 | glutathione S-transferase TAU 19 |
| 9 | Lus10025889 | 1.60 | 5.3E-03 | 1391 | AT5G48580.1 | FKBP15-2 | PF00254 | FK506- and rapamycin-binding protein 15 kD-2 |
| 10 | Lus10013078 | 1.58 | 4.9E-03 | 1526 | AT3G27830.1 | RPL12, RPL12-A | PF00542 | ribosomal protein L12-A |
| 11 | Lus10026370 | 1.54 | 2.9E-02 | 223 | AT3G08590.1 | | PF01676, PF06415 | Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent |
| 12 | Lus10035263 | 1.52 | 3.0E-02 | 402 | AT5G08680.1 | | PF00006, PF11421, PF02874, PF00306 | ATP synthase alpha/beta family protein |
| 13 | Lus10011125 | 1.51 | 5.4E-04 | 232 | AT5G25880.1 | ATNADP-ME3, NADP-ME3 | PF00390, PF03949 | NADP-malic enzyme 3 |
| 14 | Lus10020717 | 1.51 | 1.6E-02 | 1165 | AT1G06680.1 | OE23, OEE2, PSBP-1, PSII-P | PF01789 | photosystem II subunit P-1 |
| 15 | Lus10002620 | 1.50 | 7.1E-03 | 69 | AT5G17920.1 | ATCIMS, ATMETS, ATMS1 | PF01717, PF08267 | Cobalamin-independent synthase family protein |
| 16 | Lus10018156 | 1.50 | 1.0E-04 | 291 | AT1G21750.1 | ATPDI5, ATPDIL1-1, PDI5, PDIL1-1 | PF00085 | PDI-like 1-1 |
| 17 | Lus10026643 | 1.50 | 3.5E-04 | 1094 | AT2G47730.1 | ATGSTF5, ATGSTF8, GST6, GSTF8 | PF02798, PF00043 | glutathione S-transferase phi 8 |

*Arabidopsis* orthologs of the *Lus* DEGs of the 32 differently abundant proteins are presented as well as gene abbreviation and Pfam domains.

further information about putative key molecular actors of elementary (single) fiber constitution and fiber bundle properties.

The morphometric data revealed a relatively high degree of phenotypical variability between the different replicates for the seven varieties analyzed, most likely reflecting the combined effects of genetic and environmental factors as shown in previous studies (Stenström et al., 2002; Clarkson, 2005; Richter et al., 2012). Indeed, clear differences were observed both for elementary fiber number per bundle and fiber cross-sectional surface between the different varieties, with the highest values for the four spring fiber varieties and lowest for the two winter fiber varieties, the oil variety (Oliver) showing intermediate values. During retting and decortication process, fiber bundles are separated from each other but the individual fibers can still be linked within the bundles. The mechanical properties are dependent on the fineness of scutched fibers, i.e. the number of individual fibers in the bundles.

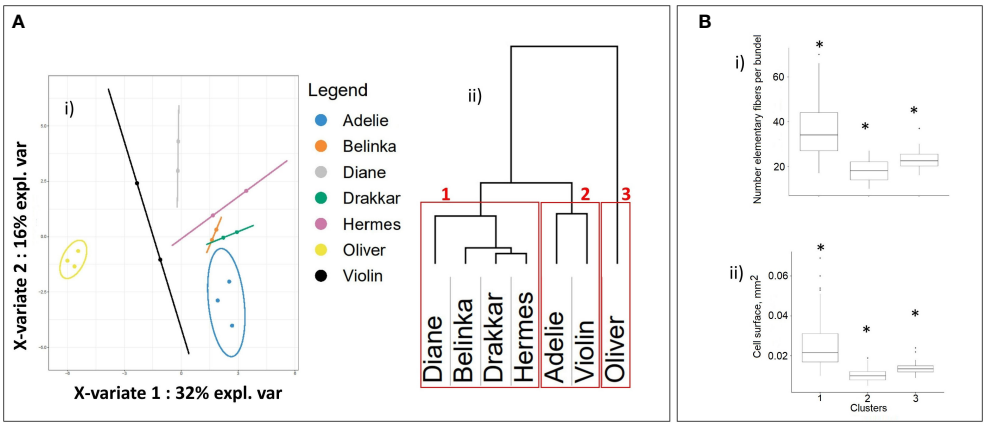
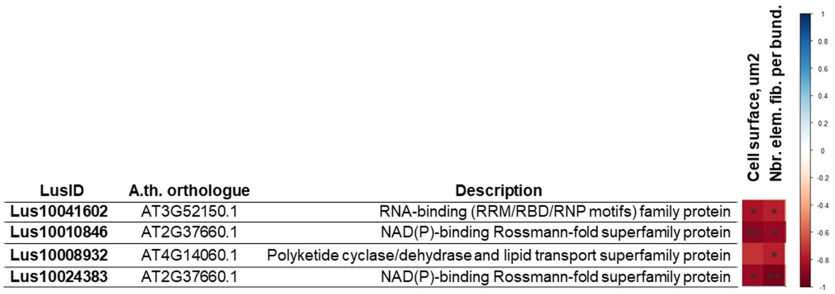



FIGURE 5
Statistical clusterization of the 7 flax varieties based upon expression levels of 32 genes corresponding to non-redundant proteins selected after statistical proteomic analysis. **(A)** (i) Partial Least-Squares Discriminant Analysis (PLS-DA), (ii) hierarchical dendrogram[12], the red frames show three groups of flax varieties that corresponded to spring fiber (1), winter fiber (2), and winter oil (3) varieties. **(B)** Boxplots of (i) individual fiber cells and (ii) surface of fiber cells within clusters highlighted in Figure 5A after conversion of the 32 candidate proteins protein IDs to gene *LusIDs* and then gene sequences. Asterisks indicate statistically significant differences (p-value ≤ 0.05) between groups.




| LusID | A.th. orthologue | Description |
|---|---|---|
| Lus10041602 | AT3G52150.1 | RNA-binding (RRM/RBD/RNP motifs) family protein |
| Lus10010846 | AT2G37660.1 | NAD(P)-binding Rossmann-fold superfamily protein |
| Lus10008932 | AT4G14060.1 | Polyketide cyclase/dehydrase and lipid transport superfamily protein |
| Lus10024383 | AT2G37660.1 | NAD(P)-binding Rossmann-fold superfamily protein |

FIGURE 6
Pearson correlation coefficient (PC) between the expression level of genes selected on proteomic data analysis and morphometric (the surface of a fiber cell, $mm^2$; the number of fibers in the bundle). The p-value for each PC was calculated, and PC with p-value ≤ 0.05 are given in the table. The red cell color marks a negative correlation, * - p-value of PC ≤ 0.05, ** - p-value of PC ≤ 0.01.

Comparative transcriptomics of fiber-enriched tissues from the seven varieties identified a total of 1041 DEGs of which 9.3% were related to cell wall metabolism (Supplementary Tables 1, 2). The majority of these genes is implicated in cell wall polymer modification including beta-glucanases, beta-galactosidases, glycosyl hydrolases, pectin lyase-like and xyloglucan endotransglucosylase/hydrolase. This is in the agreement with previous findings showing that the proteins identified in the outer stem tissues compared to inner stem tissues are mostly associated with pectin and hemicellulose remodeling (Roach et al., 2011; Chabi et al., 2017). Taking into account that the bast fibers contain up to 78% of cellulose (Thygesen et al., 2011), and tertiary cell wall of flax is in fact a composite containing mostly cellulose and pectin and hemicellulose (Gurjanov et al., 2008; Mohnen, 2008; Mikshina et al., 2013), we could hypothesize that pectin and hemicellulose remodeling could represent a biochemical basis for the variation of morphometric and mechanical parameters in different flax varieties.

Subsequent sPLS-DA analysis of the transcriptomic data identified 4 distinct variety clusters and showed that 56 genes corresponded to the first 3 X-variates and explained 73% total variance. Calculations of the average morphometric data for the four clusters revealed statistically significant differences between them, suggesting a possible link between the identified DEGs and morphometric data. Somewhat surprisingly, only 1 out of the 56 candidate genes, *Lus10037377* encoding a xyloglucan endotransglucosylase/hydrolase 32, could be directly related to cell wall organization supporting our hypothesis that other molecular processes may be important in determining key fiber parameters.

The calculation of Pearson correlation coefficients enabled the identification of 5 (out of 56) genes showing a strong correlation between expression levels and morphometric features. One of the five identified genes – *Lus10022736* (*LusWRKY85*) – showed a strong positive correlation with the number of elementary fibers per bundle. There are 102 *LusWRKY* genes predicted in the flax genome, and the *LusWRKY85* (*Lus10022736*) gene, together with its near paralogues *LusWRKY84 (Lus10014177)*, *LusWRKY87 (Lus10025133)* and *LusWRKY88 (Lus10025216)* are included in the same group of

12 https://software.broadinstitute.org/morpheus/

10.3389/fpls.2023.1204016

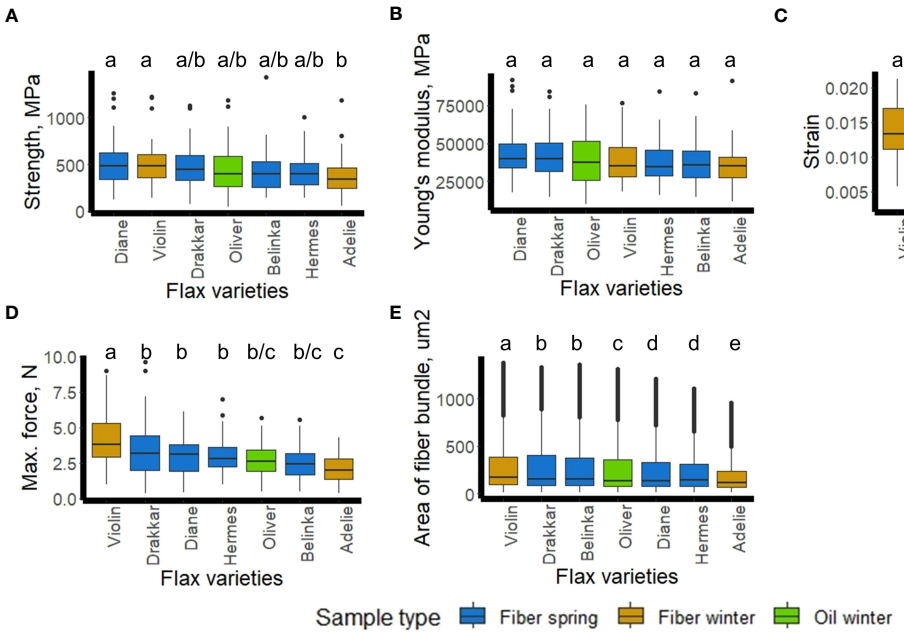



FIGURE 7
Flax fiber mechanical properties (**A**-Strength, **B**-Young's modulus, **C**-Strain, **D**-Maximum force) and geometric properties (**E**-Area of scotched fiber bundles) of the 7 flax varieties (Diane, Hermes, Drakkar, Belinka, Adelie, Violin, and Oliver). Letters shows statistical differences between flax varieties, p-value ≤ 0.05 (Tukey's test).

transcription factors (group III), which is distinguished on the distribution of conserved motifs in *LusWRKY*-genes (Yuan et al., 2021). Developmental studies have shown that *LusWRKY85* (*Lus10022736*) is expressed mostly at the seedling and maturity stage (Yuan et al., 2021). Since the number of elementary fibers per bundle is largely determined by their intrusive growth (Mokshina et al., 2018), it is tempting to hypothesize that the observed expression profile of this gene at the early stages of development could explain the correlation with this trait. Another gene *LusS40-6* (*Lus10002073*) encodes a senescence regulator and shows a strong statistically significant negative correlation with both of the measured morphometric features. Members of the *S40* gene group were described in barley and *Arabidopsis* plants as playing a role in regulation of leaf senescens (Krupinska et al., 2002; Fischer-Kilbienski et al., 2010), and it is possible that its low expression may lead to a delay in senescence, thereby allowing for increased accumulation of cellulose-rich cell wall material and enhanced cell surface.

In order to identify additional actors of fiber development process and/or confirm transcripts identified via transcriptomics analyses, comparative proteomics were performed on the two varieties Diane (Spring fiber) and Oliver (Winter oil) that belonged to the most widely separated clusters identified both during transcriptomic and morphometric analyses. After analysis, 32 non-redundant proteins were identified, of which 15 and 17 were more abundant in Diane and Oliver, respectively compared to each other. None of the identified proteins were common with the 56 genes identified in the first 3 X-variates of the transcriptomic data. Such an observation is not unusual since protein abundance is not only determined by transcript abundance but also by other factors, including translation efficiency and protein turnover, as well as by a wide range of post-translational modifications (Gygi et al., 1999; Haider and Pal, 2013). Nevertheless, 5 of the proteins were also present in the 1041 DEGs identified by transcriptomics (*Lus10023569*, *Lus10014003*, *Lus10037377*, *Lus10001508*, and *Lus10030577*).

TABLE 2 Pearson correlation coefficient (PC) between the expression level of selected genes and mechanical features (strength, Young's modulus, strain, maximum force, and area of a fiber bundle).

| LusID | *A. thaliana* orthologue | Description | Strength, MPa | Young's modulus, MPa | Strain | Max. force, N | Area, $mm^2$ |
|---|---|---|---|---|---|---|---|
| Lus10011125 | AT5G25880.1 | NADP-malic enzyme 3 | 0.75* | 0.45 | 0.56 | 0.59 | 0.61 |
| Lus10010846 | AT2G37660.1 | NAD(P)-binding Rossmann-fold superfamily protein | -0.80* | -0.72 | -0.51 | -0.25 | -0.38 |
| gi\|171462119 | | Expansin-like protein | 0.28 | -0.07 | 0.52 | 0.78* | 0.54 |

The p-value for each PC was calculated, and PC with p-value ≤ 0.05 are given in the table. Red highlighting indicates a high correlation, * - p-value of PC ≤0.05.

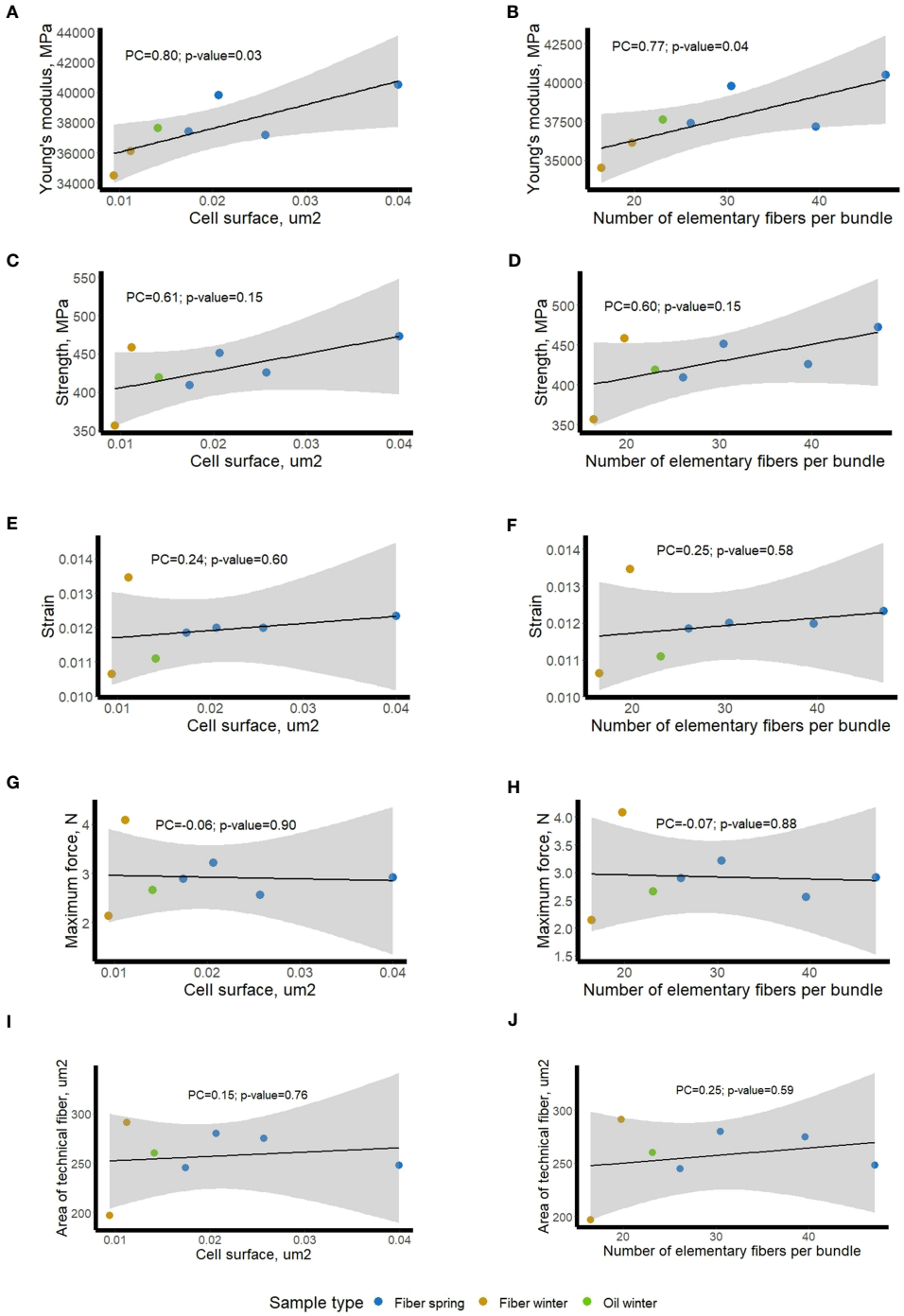



FIGURE 8
Relationship between mechanical data (*ex-planta*) of fiber bundles and morphometric data (*in-planta*) of stem fibers/fiber bundles in different flax varieties. Blue dots – fiber spring, bronze dots – fiber winter, and green dot – oil winter; mechanical data *vs* cell surface **(A, C, E, G, I)**; mechanical data vs bundle fiber number **(B, D, F, H, J)**. The regression line is plotted with a black line, a grey shaded area shows 95% confidence interval. Pearson Correlation coefficients (PC) and their p-values are given on the plot.

To evaluate the relative weight of the 32 identified proteins as potential markers for fiber-related parameters in the complete panel of 7 varieties, we used PLS-DA to generate clusters based on the transcript levels of the corresponding genes. This was done since insufficient proteomic data was available for sPLS-DA. Hierarchical clustering identified 3 clusters that more accurately reflected the variety morphotype (spring-/winter-fiber and oil) than similar clustering based on the 1041 DEGs. A similar clusterization was obtained for the same 7 flax varieties based on retrotransposon-derived polymorphism (Galindo-González et al., 2016). This study also revealed that the Ty1-copia retrotransposons have been active since breeding started in flax, and that they decreased the expression level of some genes.

Calculation of Pearson correlation coefficients between the expression levels of the genes corresponding to the 32 identified

proteins and morphometric data led to the identification of 4 genes showing significant negative correlations. Two of the genes (*Lus10010846* and *Lus10024383*) corresponded to the same *Arabidopsis thaliana* orthologue (*AT2G37660*), which encodes a NAD(P)-binding Rossmann-fold superfamily protein, also referred to a putative 3-β-hydroxysteroid dehydratase/isomerase. This multifunctional steroid metabolism enzyme is suspected to be involved in antioxidative defense, protein folding, and repair in various subcellular locations such as the apoplast, chloroplast, chloroplast stroma, endoplasmic reticulum and thylakoid. Interestingly, certain *Arabidopsis* sterol biosynthesis mutants show crucial alterations in plant organization at the cellular level in globular stage embryos, including incomplete cell wall formation and aberrant cell wall thickening (Schrick et al., 2004). The authors hypothesize that sterols could be crucial for cellulose synthesis during cell wall construction. Another example of a sterol-related enzyme is GAME25 that plays a key role in the formation of steroidal specialized metabolites (Sonawane et al., 2018). Its activity not only affects the enormous diversity of steroidal glycoalkaloids and steroidal saponins, which are produced in hundreds of plant species but also modulates the molecules' toxic effects. It is interesting to note that this protein has been previously referred to as a lipid-binding protein in phloem exudates of *Arabidopsis* (Guelette et al., 2012) and that its gene is co-expressed with *Lus10006497* coding a Lipase/lipooxygenase, involved in mediating response to stresses such as pathogen infection (Ascencio-Ibáñez et al., 2008). Another highlighted gene is *Lus10041602* encoding an RNA-binding (RRM/RBD/RNP motifs) family protein such as PLASTID-SPECIFIC RIBOSOMAL PROTEIN 2, a protein unique to plastid ribosomes (Yamaguchi and Subramanian, 2003). Transgenic *Arabidopsis* plants overexpressing PSRP2 showed delayed germination compared with that of wild-type plants under salt, dehydration, or low-temperature stress conditions (Tiller et al., 2012). The PSRP2 protein shows RNA chaperone activity, and taken together these results could suggest that PSRP2 may play a role as a negative regulator in seed germination (Xu et al., 2013). The fourth gene (*Lus10008932*) encodes a cytosolic or nuclear polyketide cyclase/dehydrase and lipid transport superfamily protein, also known as MLP-like protein 328 (Major Latex protein-). A previous study showed that MLP-like protein 328 transcripts were the most abundant in flax stem tissues above the snap point (Guo et al., 2021), which is the transition zone in development process of fiber cells (Gorshkova et al., 2003).

Our attempts to link the Omics results with morphometric data were based upon analyses of living plant material harvested at the budding stage. As such, we considered the morphometric data as *in-planta* factors. However, as explained above, key fiber parameters (fineness, mechanical properties) are also affected by fiber separation and extraction during the post-harvest retting process (Djemiel et al., 2020). Since these actions occur after the plant has been harvested, we decided to refer to the obtained data as *ex-planta*. In order to see whether these two sets of data were related, linear regressions were made between each of the four fiber bundle quality properties (dependent variable) with the two morphometric parameters (independent variables). The results showed that the studied morphometric features were able to explain close to 50% of the Young's modulus variation and about 30% of the strength variation. This fact suggests that the measurements of the cell surfaces and the number of elementary fibers per bundle before retting could be considered as relevant indicators of the biological background that influence the mechanical properties of the fiber bundle after retting and scutching. It also suggests that it is necessary to consider both *in-planta* and *ex-planta* properties to be able to explain fiber variability especially when considering the Young's modulus, which is a relevant indicator of the biological quality of a fiber bundle (Haag and Müssig, 2016).

Calculation of Pearson correlation coefficients led to the identification of 3 genes whose expression levels within bast fiber bearing tissues showed a strong correlation with *ex-planta* mechanical parameters of scutched fibers obtained after retting. Similar approach using transcriptomic data and *ex-planta* parameters was recently described for flax (Galinousky et al., 2020) which shown comparable explanatory potential of individual molecular markers for mechanical features of the scutched fibers (up to 54% described by Galinousky et al., 2020; and 48-54% in the current study). In our study we highlighted the *Lus10011125* gene that encodes an enzyme that catalyzes the oxidative decarboxylation of (S)-malate to pyruvate using NADP+ as a cofactor. According to the FIBexDB[9] data, the expression level of this gene increases in flax roots treated with *Fusarium oxysporum* or aluminum, and also in the reduced fiber flax mutant (the rdf-mutant) (Dmitriev et al., 2016; Dmitriev et al., 2017; Mokshina et al., 2021; Petrova et al., 2022). The expansin-like protein encoding gene (*gi|171462119*) whose expression strongly correlates with the maximum force, may regulate the mechanical properties of flax fibers by affecting the degree of cellulose microfibril alignment and the interactions between cellulose and other cell wall components (Cosgrove, 2000). At last, *Lus10010846*, coding a binding Rossman-fold superfamily protein involved in biosynthesis and maintenance of plant cell walls, was already previously shown to be highly correlated with morphometric features (Figure 6), and could be of great interest for both *in-planta* and *ex-planta* fiber features management and prediction.

## 5 Conclusion and future perspectives

Our study indicates that Omics combined with appropriate statistical analyses can be a powerful tool for identifying genes potentially involved in determining *in-planta* and *ex-planta* properties related to flax fiber quality. The results suggest that the observed differences in morphometric properties between the studied varieties have a background in differentially-expressed genes involved in different biochemical pathways. The genetic variability of the studied varieties may be reflected in the functioning of a sub-set of key genes present in the bast fiber bearing tissues, including those involved in lipid metabolism in the phloem and senescence process. The expression of specific lipid transfer proteins was previously correlated with either the elongation or thickening stage of flax bast fiber development (Roach and Deyholos, 2007). Similarly, previous expression data showed that about 11% of over-expressed genes in the bast of the stem were

9 https://ssl.cres-t.org/fibex/

potentially associated with lipid and wax metabolism (Fenart et al., 2010). The association of delayed senescence processes and intensive formation of bast fibers could be related to the later transition of flax fiber varieties to budding and flowering compared to oil seed, as well as to their more prolonged rapid growth stage and higher stem length (Diederichsen and Richards, 2003). There is also increasing evidence that remodeling of cell wall polymers is an important process during the formation and maturation of flax fibers. Molecular actors involved in cell wall remodeling may therefore represent a potential target for improving flax fiber properties.

Genomic breeding approaches commonly include the development of an optimal genomic model which utilizes a set of genetic markers favorable for the desired phenotypic trait (He et al., 2019; Khan et al., 2023; Rahman and Hoque, 2023). In our current study, we show how individual protein and RNA markers have a higher explanatory potential for phenotypic features and potentially could be used as effective predictors of flax fiber quality parameters. In the context of the global climate change trend, these genes we have highlighted could be tested over several years and under different climatic conditions.

Overall, our results suggest that the range of possible molecular processes that could be targeted for genetic improvement of flax fiber quality is far larger than previously imagined above and beyond the different metabolisms associated with the biosynthesis of cell wall polymers. We hypothesize the development of a new generation of breading approaches based on the combined use of Omic data and appropriate bio-informatics tools.

## Data availability statement

The transcriptomic data presented in the study are deposited in the NCBI repository, accession number GSE222066. The proteomics data presented in the study are deposited in the ProteomeXchange repository, accession number PXD039389.

## Author contributions

MC, SH, GN, AD, A-SB, BC, and SG contributed to conception and design of the study. SH secured the funding. MC conducted the experiment. KH and JM conducted the mechanical analyses. MC, JR, and SP conducted the proteomic analysis. MC and AL-D conducted the transcriptomic analyses. EG and DG performed the Omics statistical analyses. MC and SH wrote the first draft of the manuscript. EG, DG, GN, SH, AD and BC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## Funding

This work was carried out within the framework of the FIBRAGEN project financed by the national funding agencies, namely the German Federal Ministry of Education and Research [BMBF, Project Fibragen; FKZ: 0315911]; the French National Research Agency [ANR-10- KBBE-0003, Fibragen]; the Spanish Ministry of Economy and Competitiveness – General Management of Scientific and Technical Research [PIM2010PKB-00672]; and Genome Canada and the Natural Sciences and Engineering Research Council of Canada.

## Acknowledgments

DG gratefully acknowledges the College de France PAUSE program (https://www.college-de-france.fr/programme-pause) and the University of Lille for financial support. The authors thank Terre de Lin (Saint-Pierre-le-Viger, France) and LINEA (Grandviliers, France) for flax field growing and for flax fibers extraction.

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1204016/full#supplementary-material

## References

Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T. J., Chu, T. M., Wolfinger, R. D., Cella, R., et al. (2008). Global analysis of arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol.* 148, 436–454. doi: 10.1104/pp.108.121038

Baley, C., Gomina, M., Breard, J., Bourmaud, A., and Davies, P.. (2020). Variability of mechanical properties of flax fibres for composite reinforcement. a review. *Ind. Crops Prod.* 145, 111984. doi: 10.1016/j.indcrop.2019.111984

Batth, R., Jain, M., Kumar, A., Nagar, P., Kumari, S., and Mustafiz, A. (2020). Zn2+ dependent glyoxalase i plays the major role in methylglyoxal detoxification and salinity stress tolerance in plants. *PloS One* 15, 1–14. doi: 10.1371/journal.pone.0233493

Bourmaud, A., Beaugrand, J., Shah, D. U., Placet, V., and Baley, C. (2018). Towards high-performance biocomposites for targeted applications: how can we best define the diversity and specificities of reinforcement plant fibre cell walls?

Bourmaud, A., Morvan, C., Bouali, A., Placet, V., Perré, P., and Baley, C. (2013). Relationships between micro-fibrillar angle, mechanical properties and biochemical composition of flax fibers. *Ind. Crops Prod.* 44, 343–351. doi: 10.1016/j.indcrop.2012.11.031

Carpita, N. C. (2011). Update on mechanisms of plant cell wall biosynthesis: how plants make cellulose and other (1→4)-β-d-Glycans1. *Plant Physiol.* 155, 171–184. doi: 10.1104/pp.110.163360

Chabbert, B., Padovani, J., Djemiel, C., Ossemond, J., Lemaître, A., Yoshinaga, A., et al. (2020). Multimodal assessment of flax dew retting and its functional impact on fibers and natural fiber composites. *Ind. Crops Prod.* 148, 112255. doi: 10.1016/j.indcrop.2020.112255

Chabi, M., Goulas, E., Leclercq, C. C., De Waele, I., Rihouey, C., Cenci, U., et al. (2017). A cell wall proteome and targeted cell wall analyses provide novel information on hemicellulose metabolism in flax. *Mol. Cell. Proteomics* 16, 1634–1651. doi: 10.1074/mcp.M116.063727

Clarkson, D. (2005). Roots: the dynamic interface between plants and the earth. * Abe J, ed. 2003. * dordrecht: kluwer academic publishers. * 165 (hardback). 460 pp. *Ann. Bot.* 95, 704–704. doi: 10.1093/aob/mci074

Cloutier, S., Ragupathy, R., Miranda, E., Radovanovic, N., Reimer, E., Walichnowski, A., et al. (2012). Integrated consensus genetic and physical maps of flax (Linum usitatissimum l.). *Theor. Appl. Genet.* 125, 1783–1795. doi: 10.1007/s00122-012-1953-0

Cosgrove, D. J. (2000). Expansive growth of plant cell walls cell wall growth / endo-1,4-P-D-glucanase ! expansin / xyloglucan endotransglycosylase EG, endo-1,4-P-D-glucanase / XET, xyloglucan endotransglycosylase. *PIam Ph~siol Bioclzern* 38, 109–124. doi: 10.1016/s0981-9428(00)00164-9

Day, A., Ruel, K., Neutelings, G., Crônier, D., David, H., Hawkins, S., et al. (2005). Lignification in the flax stem: evidence for an unusual lignin in bast fibers. *Planta* 222, 234–245. doi: 10.1007/s00425-005-1537-1

Diederichsen, A., and Hammer, K. (1995). Variation of cultivated flax (Linum usitatissimum l. subsp.usitatissimum) and its wild progenitor pale flax (subsp.angustifolium (Huds.) thell.). *Genet. Resour. Crop Evol.* 42, 263–272. doi: 10.1007/BF02431261

Diederichsen, A., and Richards, K. (2003). "Cultivated flax and the genus linum l.: taxonomy and germplasm conservation," in *Flax the genus linum edited*. Eds. A. D. Muir and N. D. Westcott (London and New York: CRC Press), 9–15.

Djemiel, C., Goulas, E., Badalato, N., Chabbert, B., Hawkins, S., and Grec, S. (2020). Targeted metagenomics of retting in flax: the beginning of the quest to harness the secret powers of the microbiota. *Front. Genet.* 11. doi: 10.3389/fgene.2020.581664

Dmitriev, A. A., Krasnov, G. S., Rozhmina, T. A., Kishlyan, N. V., Zyablitsin, A. V., Sadritdinova, A. F., et al. (2016). Glutathione s-transferases and UDP-glycosyltransferases are involved in response to aluminum stress in flax. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01920

Dmitriev, A. A., Krasnov, G. S., Rozhmina, T. A., Novakovskiy, R. O., Snezhkina, A. V., Fedorova, M. S., et al. (2017). Differential gene expression in response to fusarium oxysporum infection in resistant and susceptible genotypes of flax (Linum usitatissimum l.). *BMC Plant Biol.* 17 (Suppl 2), 253. doi: 10.1186/s12870-017-1192-2

Esau, K. (1943). Vascular differentiation in the vegetative shoot of linum. III. the origin of the bast fibers. *Am. J. Bot.* 30, 579–586. doi: 10.2307/2437468

Fenart, S., Ndong, Y.-P., Duarte, J., Rivière, N., Wilmer, J., van Wuytswinkel, O., et al. (2010). Development and validation of a flax (Linum usitatissimum l.) gene expression oligo microarray. *BMC Genomics* 11, 592. doi: 10.1186/1471-2164-11-592

Fischer-Kilbienski, I., Miao, Y., Roitsch, T., Zschiesche, W., Humbeck, K., and Krupinska, K. (2010). Nuclear targeted AtS40 modulates senescence associated gene expression in arabidopsis thaliana during natural development and in darkness. *Plant Mol. Biol.* 73, 379–390. doi: 10.1007/s11103-010-9618-3

Galindo-González, L., Mhiri, C., Grandbastien, M. A., and Deyholos, M. K. (2016). Ty1-copia elements reveal diverse insertion sites linked to polymorphisms among flax (Linum usitatissimum l.) accessions. *BMC Genomics* 17, 1–19. doi: 10.1186/s12864-016-3337-3

Galinousky, D., Mokshina, N., Padvitski, T., Ageeva, M., Bogdan, V., Kilchevsky, A., et al. (2020). The toolbox for fiber flax breeding: a pipeline from gene expression to fiber quality. *Front. Genet.* 11. doi: 10.3389/fgene.2020.589881

Ge, S. X., Jung, D., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. doi: 10.1093/bioinformatics/btz931

Gorshkov, O., Mokshina, N., Gorshkov, V., Chemikosova, S., Gogolev, Y., and Gorshkova, T. (2017). Transcriptome portrait of cellulose-enriched flax fibres at advanced stage of specialization. *Plant Mol. Biol.* 93, 431–449. doi: 10.1007/s11103-016-0571-7

Gorshkova, T., Brutch, N., Chabbert, B., Deyholos, M., Hayashi, T., Lev-Yadun, S., et al. (2012). Plant fiber formation: state of the art, recent and expected progress, and open questions. *CRC Crit. Rev. Plant Sci.* 31, 201–228. doi: 10.1080/07352689.2011.616096

Gorshkova, T., Chernova, T., Mokshina, N., Gorshkov, V., Kozlova, L., and Gorshkov, O. (2018). Transcriptome analysis of intrusively growing flax fibers isolated by laser microdissection. *Sci. Rep.* 8, 1–17. doi: 10.1038/s41598-018-32869-2

Gorshkova, T., Mokshina, N., Mitsuda, N., and Gorshkov, O. (2023). "Key stages of flax bast fiber development through the prism of transcriptomics," in *The flax genome*. Eds. F. M. You and B. Fofana (Canada: Springer Cham), 149–198. doi: 10.1016/S0926-6690(03)00043-8

Gorshkova, T. A., Sal'nikov, V. V., Chemikosova, S. B., Ageeva, M. V., Pavlencheva, N. V., and Van Dam, J. E. G. (2003). The snap point: a transition point in linum usitatissimum bast fiber development. *Ind. Crops Prod.* 18, 213–221. doi: 10.1016/S0926-6690(03)00043-8

Goulas, E., Le Dily, F., Teissedre, L., Corbel, G., Robin, C., and Ourry, A. (2001). Vegetative storage proteins in white clover (Trifolium repens l.): quantitative and qualitative features. *Ann. Bot.* 88, 789–795. doi: 10.1006/anbo.2001.1501

Goulas, E., Schubert, M., Kieselbach, T., Kleczkowski, L. A., Gardeström, P., Schröder, W., et al. (2006). The chloroplast lumen and stromal proteomes of arabidopsis thaliana show differential sensitivity to short- and long-term exposure to low temperature. *Plant J.* 47, 720–734. doi: 10.1111/j.1365-313X.2006.02821.x

Grimaud, F., Renaut, J., Dumont, E., Sergeant, K., Lucau-Danila, A., Blervacq, A. S., et al. (2013). Exploring chloroplastic changes related to chilling and freezing tolerance during cold acclimation of pea (Pisum sativum l.). *J. Proteomics* 80, 145–159. doi: 10.1016/j.jprot.2012.12.030

Guelette, B. S., Benning, U., and Hoffmann-Benning, S. (2012). Identification of lipids and lipid-binding proteins in phloem exudates of arabidopsis thaliana. *J. Exp. Bot.* 63, 3603–3616. doi: 10.1093/jxb/ers028

Guo, Y., Wen, L., Chen, J., Pan, G., Wu, Z., Li, Z., et al. (2021). Comparative transcriptomic analysis identifies key cellulose synthase genes (CESA) and cellulose synthase-like genes (CSL) in fast growth period of flax stem (Linum usitatissimum l.). *J. Nat. Fibers* 19 (9), 1–16. doi: 10.1080/15440478.2021.1993510

Gurjanov, O. P., Ibragimova, N. N., Gnezdilov, O. I., and Gorshkova, T. A. (2008). Polysaccharides, tightly bound to cellulose in cell wall of flax bast fibre: isolation and identification. *Carbohydr. Polym.* 72, 719–729. doi: 10.1016/j.carbpol.2007.10.017

Gutiérrez, R. A., Green, P. J., Keegstra, K., and Ohlrogge, J. B. (2004). Phylogenetic profiling of the Arabidopsis thaliana proteome: what proteins distinguish plants from other organisms? *Genome Biol.* 5 (8), R53. doi: 10.1186/gb-2004-5-8-r53

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. (Nature America Inc.), 17, 1–6. doi: 10.1038/13690

Haag, K., and Müssig, J. (2016). Scatter in tensile properties of flax fibre bundles: influence of determination and calculation of the cross-sectional area. *J. Mater. Sci.* 51, 7907–7917. doi: 10.1007/s10853-016-0052-z

Haag, K., Padovani, J., Fita, S., Trouvé, J. P., Pineau, C., Hawkins, S., et al. (2017). Influence of flax fibre variety and year-to-year variability on composite properties. *Ind. Crops Prod.* 98, 1–9. doi: 10.1016/j.indcrop.2016.12.028

Haider, S., and Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* 14, 91–110. doi: 10.2174/1389202911314020003

He, L., Xiao, J., Rashid, K. Y., Yao, Z., Li, P., Jia, G., et al. (2019). Genome-wide association studies for pasmo resistance in flax (Linum usitatissimum l.). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01982

Hotte, N. S. C., and Deyholos, M. K. (2008). A flax fibre proteome: identification of proteins enriched in bast fibres. *BMC Plant Biol.* 8, 52. doi: 10.1186/1471-2229-8-52

Huis, R., Morreel, K., Fliniaux, O., Lucau-Danila, A., Fénart, S., Grec, S., et al. (2012). Natural hypolignification is associated with extensive oligolignol accumulation in flax stems. *Plant Physiol.* 158, 1893–1915. doi: 10.1104/pp.111.192328

Khan, N., Shazadee, H., Cloutier, S., and You, F. M. (2023). "Genomics assisted breeding strategy in flax," in *The flax genome*. Eds. F. M. You and B. Fofana (Canada: Springer Cham), 253–272.

Krupinska, K., Haussühl, K., Schäfer, A., van der Kooij, T. A. W., Leckband, G., Lörz, H., et al. (2002). A novel nucleus-targeted protein is expressed in barley leaves during senescence and pathogen infection. *Plant Physiol.* 130, 1172–1180. doi: 10.1104/pp.008565

Mikshina, P., Chernova, T., Chemikosova, S., Ibragimova, N., Mokshina, N., and Gorshkova, T. (2013). "Cellulosic fibers: role of matrix polysaccharides in structure and function.," in. *Cellulose - Fundam. Aspects* 91–112. doi: 10.5772/2705

Mohnen, D. (2008). Pectin structure and biosynthesis. *Curr. Opin. Plant Biol.* 11, 266–277. doi: 10.1016/j.pbi.2008.03.006

Mokshina, N., Chernova, T., Galinousky, D., Gorshkov, O., and Gorshkova, T. (2018). Key stages of fiber development as determinants of bast fiber yield and quality. *Fibers* 6, 1–11. doi: 10.3390/fib6020020

Mokshina, N., Gorshkov, O., Takasaki, H., Onodera, H., Sakamoto, S., Gorshkova, T., et al. (2021). FIBexDB: a new online transcriptome platform to analyze development of plant cellulosic fibers. *New Phytol.* 231, 512–515. doi: 10.1111/nph.17405

Morvan, C., Andème-Onzighi, C., Girault, R., Himmelsbach, D. S., Driouich, A., and Akin, D. E. (2003). Building flax fibres: more than one brick in the walls. *Plant Physiol. Biochem.* 41, 935–944. doi: 10.1016/j.plaphy.2003.07.001

Muravenko, O. V., Samatadze, T. E., Popov, K. V., Amosova, A. V., and Zelenin, A. V. (2001). Comparative genome analysis in two flax species by c-banding patterns. *Russ. J. Genet.* 37, 253–256. doi: 10.1023/A:1009052909215

Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50, D543–D552. doi: 10.1093/nar/gkab1038

Petrova, A., Zakharov, M., Ageeva, M., McKenzie, R., Gorshkova, T., Deyholos, M., et al. (2022). A flax mutant with impaired intrusive growth of phloem and xylem fibres demonstrates constitutive gravitropic response. *Plant Sci.* 323, 111399. doi: 10.1016/j.plantsci.2022.111399

Rachinskaya, O. A., Lemesh, V. A., Muravenko, O. V., Yurkevich, O. Y., Guzenko, E. V., Bol'sheva, N. L., et al. (2011). Genetic polymorphism of flax linum usitatissimum based on the use of molecular cytogenetic markers. *Russ. J. Genet.* 47, 56–65. doi: 10.1134/S1022795411010108

Rahman, M., and Hoque, A. (2023). "Flax breeding," in *The flax genome*. Eds. F. M. You and B. Fofana (Canada: Springer Cham), 55–68.

Richter, S., Kipfer, T., Wohlgemuth, T., Guerrero, C. C., Ghazoul, J., and Moser, B. (2012). Phenotypic plasticity facilitates resistance to climate change in a highly variable environment. *Oecologia* 169, 269–279. doi: 10.1007/s00442-011-2191-x

Roach, M. J., and Deyholos, M. K. (2007). Microarray analysis of flax (Linum usitatissimum l.) stems identifies transcripts enriched in fibre-bearing phloem tissues. *Mol. Genet. Genomics* 278, 149–165. doi: 10.1007/s00438-007-0241-1

Roach, M. J., and Deyholos, M. K. (2008). Microarray analysis of developing flax hypocotyls identifies novel transcripts correlated with specific stages of phloem fibre differentiation. *Ann. Bot.* 102, 317–330. doi: 10.1093/aob/mcn110

Roach, M. J., Mokshina, N. Y., Badhan, A., Snegireva, A. V., Hobson, N., Deyholos, M. K., et al. (2011). Development of cellulosic secondary walls in flax fibers requires β-galactosidase. *Plant Physiol.* 156, 1351–1363. doi: 10.1104/pp.111.172676

Schrick, K., Fujioka, S., Takatsuto, S., Stierhof, Y. D., Stransky, H., Yoshida, S., et al. (2004). A link between sterol biosynthesis, the cell wall, and cellulose in arabidopsis. *Plant J.* 38, 227–243. doi: 10.1111/j.1365-313X.2004.02039.x

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., et al. (2019). MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* 12, 879–892. doi: 10.1016/j.molp.2019.01.003

Simon, C., Lion, C., Spriet, C., Baldacci-cresp, F., and Hawkins, S. (2018). One , two , Three : a bioorthogonal triple labelling strategy for studying the dynamics of plant cell wall formation *In vivo*. Angewandte Chemie International Edition - Wiley Online Library, 130 1–8. doi: 10.1002/anie.201808493

Somalraju, A., and Fofana, B. (2023). "Metabolomics and transcriptomics-based tools for linseed improvement," in *The flax genome*. Eds. F. M. You and B. Fofana (Canada: Springer Cham), 199–216.

Sonawane, P. D., Heinig, U., Panda, S., Gilboa, N. S., Yona, M., Pradeep Kumar, S., et al. (2018). Short-chain dehydrogenase/reductase governs steroidal specialized metabolites structural diversity and toxicity in the genus solanum. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5419–E5428. doi: 10.1073/pnas.1804835115

Stenström, A., Jónsdóttir, I. S., and Augner, M. (2002). Genetic and environmental effects on morphology in clonal sedges in the Eurasian arctic. *Am. J. Bot.* 89, 1410–1421. doi: 10.3732/ajb.89.9.1410

Thygesen, A., Madsen, B., Bjerre, A. B., and Lilholt, H. (2011). Cellulosic fibers: effect of processing on fiber bundle strength. *J. Nat. Fibers* 8, 161–175. doi: 10.1080/15440478.2011.602236

Tiller, N., Weingartner, M., Thiele, W., Maximova, E., Schöttler, M. A., and Bock, R. (2012). The plastid-specific ribosomal proteins of arabidopsis thaliana can be divided into non-essential proteins and genuine ribosomal proteins. *Plant J.* 69, 302–316. doi: 10.1111/j.1365-313X.2011.04791.x

Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., et al. (2012). The genome of flax (Linum usitatissimum) assembled *de novo* from short shotgun sequence reads. *Plant J.* 72, 461–473. doi: 10.1111/j.1365-313X.2012.05093.x

Wrobél-Kwiatowska, M., Hawkins, S., and Szopa, J. (2010). *Engineering of plants for improved fiber qualities. advances in textile biotechnology* (Cambridge: Woodhead Publishing Series in Textiles). Woodhead p. doi: 10.1533/9780857090232.2.150

Xu, T., Lee, K., Gu, L., Kim, J., and Kang, H. (2013). Functional characterization of a plastid-specific ribosomal protein PSRP2 in arabidopsis thaliana under abiotic stress conditions. *Plant Physiol. Biochem.* 73, 405–411. doi: 10.1016/j.plaphy.2013.10.027

Yamaguchi, K., and Subramanian, A. R. (2003). Proteomic identification of all plastid-specific ribosomal proteins in higher plant chloroplast 30s ribosomal subunit: PSRP-2 (U1A-type domains), PSRP-3α/β (ycf65 homologue) and PSRP-4 (Thx homologue). *Eur. J. Biochem.* 270, 190–205. doi: 10.1046/j.1432-1033.2003.03359.x

You, F. M., and Fofana, B. (Eds.) (2023). *The flax genome* (Canada: Springer Cham). doi: 10.1007/978-3-031-16061-5

You, F. M., Moumen, I., Khan, N., and Cloutier, S. (2023). "Reference genome sequence of flax," in *The flax genome*. Eds. F. M. You and B. Fofana (Canada: Springer Cham), 1–17.

You, F. M., Xiao, J., Li, P., Yao, Z., Jia, G., He, L., et al. (2018). Chromosome-scale pseudomolecules refined by optical, physical and genetic maps in flax. *Plant J.* 95, 371–384. doi: 10.1111/tpj.13944

Yuan, H., Guo, W., Zhao, L., Yu, Y., Chen, S., Tao, L., et al. (2021). Genome-wide identification and expression analysis of the WRKY transcription factor family in flax (Linum usitatissimum l.). *BMC Genomics* 22, 1–15. doi: 10.1186/s12864-021-07697-w

Zhang, N., and Deyholos, M. K. (2016). RNASeq analysis of the shoot apex of flax ( linum usitatissimum ) to identify phloem fiber specification genes. *Front. Plant. Sci.* 7, 1–5. doi: 10.3389/fpls.2016.00950

Zohary, D., and Hopf, M. (2004). *Domestication of plants in the old world* (Oxford, UK: Oxford University Press), 125–132.

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 11 August 2023
DOI 10.3389/fpls.2023.1185377



OPEN ACCESS

EDITED BY
Kashif Ali,
Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Pakistan

REVIEWED BY
Congting Ye,
Xiamen University, China
Tri Rini Nuringtyas,
Gadjah Mada University, Indonesia

*CORRESPONDENCE
Lingli Lu
lulingli@zju.edu.cn

RECEIVED 13 March 2023
ACCEPTED 26 July 2023
PUBLISHED 11 August 2023

CITATION
Chen C, Ge Y and Lu L (2023) Opportunities and challenges in the application of single-cell and spatial transcriptomics in plants.
*Front. Plant Sci.* 14:1185377.
doi: 10.3389/fpls.2023.1185377



# Opportunities and challenges in the application of single-cell and spatial transcriptomics in plants

Ce Chen[1], Yining Ge[1] and Lingli Lu[1,2]*

[1]Ministry of Education Key Laboratory of Environment Remediation and Ecological Health, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China, [2]Key Laboratory of Agricultural Resource and Environment of Zhejiang Province, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China

Single-cell and spatial transcriptomics have diverted researchers' attention from the multicellular level to the single-cell level and spatial information. Single-cell transcriptomes provide insights into the transcriptome at the single-cell level, whereas spatial transcriptomes help preserve spatial information. Although these two omics technologies are helpful and mature, further research is needed to ensure their widespread applicability in plant studies. Reviewing recent research on plant single-cell or spatial transcriptomics, we compared the different experimental methods used in various plants. The limitations and challenges are clear for both single-cell and spatial transcriptomic analyses, such as the lack of applicability, spatial information, or high resolution. Subsequently, we put forth further applications, such as cross-species analysis of roots at the single-cell level and the idea that single-cell transcriptome analysis needs to be combined with other omics analyses to achieve superiority over individual omics analyses. Overall, the results of this review suggest that combining single-cell transcriptomics, spatial transcriptomics, and spatial element distribution can provide a promising research direction, particularly for plant research.

KEYWORDS

**single-cell transcriptomics, spatial transcriptomics, single-nucleus RNA-seq, single-cell RNA-seq, protoplast**

## 1 Introduction

A transcriptome is the collection of all transcripts in a cell under specific physiological conditions. Transcriptomics can reveal differences in gene expression under different conditions (Shojaee et al., 2021). From the beginning of transcriptome analysis of various organs to the combination of laser capture microdissection (LCM) and high-throughput sequencing for single-cell transcriptome analysis, the demand for the resolution of transcriptome analysis is increasing. Traditional simple transcriptome analysis cannot meet this demand; however, single-cell transcriptome analysis can. As its name implies, a single-cell transcriptome is the transcriptome of a single cell. Since scientists first reported single-cell transcriptome technology, it has undergone considerable development (Tang

et al., 2009). In 2015, two research groups from Harvard developed the Drop-seq and InDrop technologies and applied them to study mouse cells (Klein et al., 2015; Macosko et al., 2015; Klein and Macosko, 2017). In 2016, the 10× Genomics Chromium system, which is a droplet-based technique, was developed. This single-cell sequencing system allows high throughput and can detect rare cell types (See et al., 2018; Svensson et al., 2018; Wang et al., 2021b). Since then, the rapid and low-cost gene expression analyses of many single cells have become a reality. Related single-cell sequencing technologies, such as GemCode single-cell technology (Zheng et al., 2017), MARS-seq2.0 (Keren-Shaul et al., 2019), and Paired-seq (Zhang et al., 2020), are constantly improving in many aspects, such as high cell capture capacity, high gene detection capability, low technical cell-to-cell contamination rate, and low cost, making it possible to use single-cell sequencing technology more conveniently and accurately. Currently, there are three main single-cell sequencing approaches: plate-based, combinatorial indexing-based, and bead-based. Plate-based approaches, e.g., Smart-seq2 and CEL-Seq2, sort a single cell into a well of a multi-well plate; however, these approaches have low throughput and are time consuming. Bead-based methods, such as 10× Chromium, Drop-seq, inDrop, and Seq-Well, use tiny droplets or wells to distribute cells, and are high-throughput and low-cost. Combinatorial indexing-based approaches, such as sciRNA-seq, can reverse transcribe and barcode mRNAs without physically isolating the cells (Ding et al., 2020). Single-cell transcriptomics technology opens a new avenue for molecular studies of tissues and organs from animals and plants, and avoids the lack of information on a particular cell type and heterogeneity from average data (Shulse et al., 2019; Mo and Jiao, 2022).

In plant research, single-cell transcriptomics has developed relatively late, but today it plays an essential and decisive role. Compared to traditional transcriptomics, the sequencing of single plant cells can resolve their heterogeneity. For example, when utilising traditional transcriptomics to profile drought-induced transcriptomic changes, researchers have observed the downregulation of growth-related, energy-consuming processes and the upregulation of stress defence genes at the organ level. However, single-cell transcriptome results suggested that, under mild drought stress, the mesophyll showed significant downregulation of genes, whereas most genes were upregulated in the epidermis. The response of plants to environmental stimuli is complex, and the key to understanding it depends on the study of networks at the single-cell level. The use of single-cell transcriptomics can help pinpoint tissue-specific pathways that respond to stress, and ultimately, these pathways can be specifically engineered in truly important tissues without causing unexpected side effects in other tissues or organs (Tenorio Berrio et al., 2022). There are two sequencing methods: one for sequencing protoplasts and the other for sequencing nuclei. However, the large and uncertain size of plant cells complicates single-cell sequencing. Problems persist even after removing the cell walls to obtain protoplasts, in addition to existing issues with the methods of protoplast extraction (Shulse et al., 2019). These problems can be solved using single nuclei for sequencing instead of whole cells, and the technique has a gene detection sensitivity similar to that of protoplast-based sequencing (van den Brink et al., 2017; Denisenko et al., 2020; Ding et al., 2020; Farmer et al., 2021). Both methods have the disadvantage of losing the spatial information of the cells when isolating them from tissues. Therefore, identifying ways to solve this problem is a widely discussed topic in single-cell transcriptomic research.

Spatial transcriptome technology has undergone considerable development in recent years. This provides a solution that compensates for the shortcomings of single-cell transcription analysis. Spatial transcriptome analysis differs from the method used for single-cell transcriptome analysis and faces specific difficulties. Compared to single-cell and traditional transcriptomics, spatial transcriptomics preserves spatial information while profiling the transcriptome, which provides more precise results for studying regulatory networks in plants. The data obtained from single-cell or traditional transcriptomic analyses are mapped back to the plant through cell sorting instead of using spatial data (Gurazada et al., 2021). During growth, differences in gene expression patterns at different locations in plant tissues lead to different phenotypes. Therefore, spatial transcriptomics provide a new perspective on plant growth and development.

Thus far, there has been a lack of studies to compare the existing methods for plant single-cell transcriptomics, specifically for isolating single cells and nuclei. Studies on spatial transcriptomics in plants are also rare. Here, we introduce and compare in detail the techniques used in related studies in recent years, explain the distinctions among different methods for extracting sequencing materials, and describe the differences between protoplast- and nucleus-based single-cell and spatial transcriptomics. Then, we put forth the challenges and future development directions and provide an idea for combining single-cell and spatial transcriptomics and spatial distribution analysis of elements in plant tissues to promote plant research.

## 2 Development of plant single-cell transcriptomics

Single-cell-related research has been conducted on animals, microorganisms, and plants. Microbiologists have used single-cell-related technologies to explore the differences in RNA accumulation between different parts of microorganisms (de Bekker et al., 2011). Zoologists have used this technology to study tumour cells (Ramsköld et al., 2012) and immune cells (Shalek et al., 2013) or to explore new methods for RNA sequencing (Grindberg et al., 2013). In recent years, plant single-cell transcriptomics has been developed. Researchers have conducted extensive plant studies using single-cell transcriptomic analyses. We reviewed the relevant literature in recent years and tabulated the results (Table 1). In 2015, Yan et al. published the first study on whole-genome sequencing of *Zea mays* using single plant cells (Li et al., 2015). Since then, the development of single-cell isolation and sequencing methods has made single-cell transcriptomics a popular topic in botanical

TABLE 1 An overview of plant single-cell transcriptomics related research in recent years.

| Plant Species | Positions | Research direction | Publication Year | Reference |
|---|---|---|---|---|
| *Oryza sativa* L. | middle part of the second leaf | monoallelic gene expression | 2017 | (Han et al., 2017) |
| | leaves | a simple system for predicting transcription factor targets | 2020 | (Xie et al., 2020) |
| | tips (5 mm) of crown roots | root cell atlas | 2021 | (Liu et al., 2021b) |
| | proximal shoots, roots | atlas | 2021 | (Wang et al., 2021c) |
| | root tips of rice radicles | atlas (differentiation trajectory) | 2021 | (Zhang et al., 2021b) |
| | young inflorescences, leaves | inflorescence development and atlas | 2022 | (Zong et al., 2022) |
| *Arabidopsis thaliana* | whole roots | atlas | 2019 | (Jean-Baptiste et al., 2019) |
| | primary root tips | atlas (gene expression map) | 2019 | (Ryu et al., 2019) |
| | roots | atlas, cell type identification | 2019 | (Shulse et al., 2019) |
| | sperm cells | protocol | 2019 | (Misra et al., 2019) |
| | root tips | atlas | 2019 | (Denyer et al., 2019) |
| | root tips | atlas | 2019 | (Zhang et al., 2019) |
| | cotyledons | stomatal lineage cell development | 2020 | (Liu et al., 2020) |
| | female gametic cells | polyploid | 2020 | (Song et al., 2020) |
| | root tips | properties of the cell cycle | 2020 | (Torii et al., 2020) |
| | root tips | atlas, limiting phosphate condition | 2020 | (Wendrich et al., 2020) |
| | roots | lateral root development | 2021 | (Gala et al., 2021) |
| | root tips | phloem development | 2021 | (Roszak et al., 2021) |
| | roots | brassinosteroid signaling | 2021 | (Graeff et al., 2021) |
| | roots | optimizing sample size | 2021 | (Chen et al., 2021a) |
| | roots | vascular development | 2021 | (Yang et al., 2021) |
| | ovule | the female germline differentiation trajectory | 2021 | (Hou et al., 2021) |
| | primary roots | atlas | 2021 | (Lhamo and Luan, 2021) |
| | roots | regulatory landscape | 2021 | (Dorrity et al., 2021) |
| | vascular cells from leaf | atlas | 2021 | (Kim et al., 2021) |
| | shoot apexes | atlas | 2021 | (Zhang et al., 2021c) |
| | seedlings, developing flowers | nuclei isolation method, atlas | 2021 | (Sunaga-Franze et al., 2021) |
| | whole aerial tissue, first true leaves | development | 2021 | (Lopez-Anido et al., 2021) |
| | cauline leaves | spatial transcriptome profiles | 2022 | (Xia et al., 2022) |
| | leaf explants | *de novo* root regeneration | 2022 | (Liu et al., 2022c) |
| | leaves | plant immune system | 2022 | (Salguero-Linares et al., 2022) |
| | seedling cotyledons | development of leaf veins | 2022 | (Liu et al., 2022d) |
| | roots | root atlas | 2022 | (Shahan et al., 2022) |
| | primary roots, above-ground tissues | transcriptome profiles (atlas) | 2022 | (Apelt et al., 2022) |

*(Continued)*

TABLE 1 Continued

| Plant Species | Positions | Research direction | Publication Year | Reference |
|---|---|---|---|---|
| | the third leaf | atlas | 2022 | (Tenorio Berrio et al., 2022) |
| *Zea mays* L. | root hairs, primary roots without root hairs | development, gene identification | 2017 | (Hey et al., 2017) |
| | anthers | development | 2019 | (Nelms and Walbot, 2019) |
| | shoot apical meristem | developmental genetic organization | 2020 | (Satterlee et al., 2020) |
| | ears | developmental atlas | 2021 | (Xu et al., 2021a) |
| | anthers | cytoplasmic male sterility | 2021 | (Zhang et al., 2021a) |
| | seedlings, tassel or ear primordia, root tips, crown roots, axillary buds, whole roots | cis-regulatory atlas | 2021 | (Marand et al., 2021) |
| | anthers, pollens | haploid induction key gene molecular mechanism | 2022 | (Jiang et al., 2022) |
| *Gossypium hirsutum* | fiber | photoinduced fiber color formation, molecular regulatory | 2021 | (Tang et al., 2021) |
| | root tips | method, atlas, salt stress | 2022 | (Liu et al., 2022b) |
| *C. sinensis* var. *sinensis* | leaves | atlas | 2022 | (Wang et al., 2022) |
| *Populus alba* | stems | vascular development | 2021 | (Chen et al., 2021b) |
| *Triticum aestivum* L. | coleoptile | atlas, the turgor alteration of guard cells | 2021 | (Wang et al., 2021a) |
| *Arachis hypogaea* L. | leaves | method, atlas, developmental trajectory and interaction network | 2021 | (Liu et al., 2021a) |

Published studies on plant single-cell transcriptomics are listed in this table. The plant species, the position of plant tissues used in these studies, research directions, and publication years are listed in the table.

research. In 2017, to explore the establishment of monoallelic gene expression in single plant cells, Jiao et al. developed a single-cell transcriptome sequencing protocol and applied it to *Oryza sativa*, which was the first report on the application of single-cell transcriptomics to plants (Han et al., 2017). To date, there have been four single-cell transcriptome studies of plant species: *Arabidopsis thaliana*, *Oryza sativa*, *Solanum lycopersicum*, and *Zea mays*, in the Single Cell Expression Atlas of the European Molecular Biology Laboratory (Conde et al., 2021).

From the published studies listed in Table 1, it is clear that most experiments were based on *A. thaliana*, and subsequently on *O. sativa* and *Z. mays*. Scientists typically use whole plant roots and root tips as experimental materials. A few scientists have also used other plant parts, such as leaves, stems, or germ cells, based on their research direction and purpose. Table 1 shows that most scientists have mainly constructed a single-cell atlas to explore the cell types of particular parts of plants, specifically roots, and examined developmental or differentiation trajectories combined with trajectory analysis, as transcriptomics can deliver information about gene expression.

Single-cell transcriptome technology has been used in numerous plant studies. In 2019, a group studied cell types from the primary root tissues of *A. thaliana* using a high-throughput Drop-seq approach. Their findings confirmed that RNA sequencing helps describe the developmental processes of plants (Shulse et al., 2019). It is worth mentioning that they also revealed that external stimuli, such as sucrose, can influence the developmental process, causing changes in sucrose response-related cells. In a similar case, in 2022, a group studied another external stimulus, salt stress, using cotton as their object and observed the influence of external stimuli on the profile and dynamic changes in gene expression (Liu et al., 2022b). For further research on the exploration of cellular heterogeneity, it is essential to compare gene expression profiles before and after external environmental stimuli by studying single-cell transcriptome sequencing results and constructing an atlas. In addition to investigating the environmental influences on plant cell gene expression, single-cell transcriptomics is a hot topic in the study of plant development and differentiation. One group has built models of cell differentiation within leaf tissues, studying stomatal lineage, and thereby determined a series of cellular programs related to tissue flexibility (Lopez-Anido et al., 2021).

With the help of single-cell transcriptomics, scientists have explored gene expression and signal transduction pathways while studying cell developmental trajectories. However, no research has

been conducted on this topic. For example, researchers have observed how cytokinin signalling creates a link between the vascular perception of limited phosphate availability and epidermal responses (Wendrich et al., 2020). They drew an intersection between a single-cell atlas and target genes. They observed the distribution of transcription factor complexes in specific tissues, revealing the significance of the cytokinin signalling pathway. Comparing single-cell transcriptomic atlas and target genes in a critical signalling pathway, such as brassinosteroid signalling (Graeff et al., 2021), helps to understand the functions or effects of particular gene products or signal complexes in plants and their delivery or distribution in different tissues. Many scientists have used single-cell transcriptome techniques to create an atlas of plant species, which still have some genes that need to be explored, and have compared them with those of *A. thaliana* to study the functions of a few genes. This comparison reduces the difficulties and increases the hope of constructing a cell atlas for different plant species. However, studies involving target genes are limited, and even fewer studies have examined gene functions. In future studies, combining the analysis of an atlas with the target genes will be an analytical method that will attract increasing attention.

In addition to single-cell transcriptomics, there are many other omics and single-cell omics studies, and a few scientists have combined single-cell transcriptomic analysis with other omics analyses. For example, plant cells can be isolated for single-cell transcriptomics, which results in the loss of spatial information. Therefore, other omics techniques, such as spatial transcriptomics, are required to supplement missing data. In 2022, scientists applied scStereo-seq technology to plant research and constructed a single-cell spatial atlas (Xia et al., 2022). Owing to the combination of two omics technologies, they avoided a few disadvantages of a single omics technology and observed that the expression levels of related genes showed a gradient change trend in space. Another group combined single-cell transcriptomics technology with multi omics technology to study the molecular basis of developing *Z. mays* ears (Xu et al., 2021a). In addition to combining single-cell transcriptomics with spatial transcriptomics, some scientists have applied single-nucleus RNA sequencing and ATAC sequencing to plant roots, compared transcriptomic and epigenomic data at the single-cell level, and revealed how chromatin accessibility influences gene expression (Farmer et al., 2021). Using different omics technologies in combination with single-cell transcriptomics helps researchers solve problems associated with using only one technology, such as gene redundancy and spatial information.

As mentioned above, plant cells have cell walls, which make it challenging to obtain a single cell. Isolating single cells is affected by many factors, such as the parts or tissues of the plant used and the environmental conditions or genotypes of the plants (Shaw et al., 2021). Therefore, few researchers have been concerned with the methods used to perform single-cell transcriptomics. They introduced a commonly used method for isolating plant nuclei and demonstrated its universal applicability (Sunaga-Franze et al., 2021). A few researchers have introduced protoplast isolation methods for cotton (Liu et al., 2022b). New methods, not only for single-cell extraction but also for sequencing, have significance in plant research, reducing costs and obtaining enough single cells.

## 3 Standard methods for single-cell transcriptomics in plants

Sequencing of plant single-cell transcriptomes can be divided into sequencing of protoplasts and sequencing of nuclei. More accurately, the latter should be referred to as the single-nucleus transcriptome. Comparing the sequencing processes of the two, overall, the sequencing materials are all obtained from plant tissues, and then purified and separated for sequencing. Specifically, to protect the activity and integrity of protoplasts, the methods used are gentler and require higher freshness of plant samples during the process of obtaining protoplasts, whereas frozen or dried plant samples can be used to obtain nuclei. The following is a detailed introduction to both methods.

Based on recent protoplast-based single-cell transcriptomics studies (Table 1), it is clear that these procedures are similar (Figures 1, 2). First, the plants were grown based on the research goals. For example, to explore the single-cell transcriptome of rice roots under Cd stress, a small amount of rice was planted in both standard and Cd-stress environments. After a period of growth, the plant material is harvested from the required parts, including the root tips, whole roots, leaves, or other parts of the plant, and cut or crushed into small pieces to expand the area of the plant material exposed to the enzyme solution and shorten the time required for enzyme treatment. Plant materials are then mixed with the enzyme solution, called the digesting enzyme solution, and incubated under appropriate conditions to release protoplasts or plant cell nuclei. Plant materials must be mixed with an enzyme solution as soon as possible after being cut or crushed to maintain cell activity and improve the quality of protoplasts. This step is followed by further steps that involve filtration, centrifugation, and washing to obtain sufficient material for sequencing, relatively fewer undigested plant cells, and fewer broken cell organelles. Digestion helps remove the cell wall to isolate protoplasts, filtration helps remove substances that are not protoplasts or are larger in size, centrifugation helps concentrate protoplasts, and washing helps remove impurities and residual reagents. The quality and concentration of sequencing materials must be confirmed, as high quality and quantity are the basis for further sequencing to obtain sufficiently precise transcriptomic data (Liu et al., 2022b). A single-cell RNA sequencing (scRNA-seq) library is prepared and sequenced. Finally, the data are analysed to obtain the experimental results. This is the entire process of using single-cell transcriptomics; however, when it comes to each specific experiment, there still exist differences, particularly in the methods used to obtain ideal protoplast materials. A critical step in protoplast-based single-cell transcriptomics in plants is obtaining high-quality sequencing objects. Culture conditions, genotype, plant age, and other factors affect extraction methods to varying degrees, implying that when preparing protoplasts, the procedures must be adjusted based on the experimental conditions. In Table 2, the methods of partial studies are compared, that is, from studies listed in Table 1 or studies the

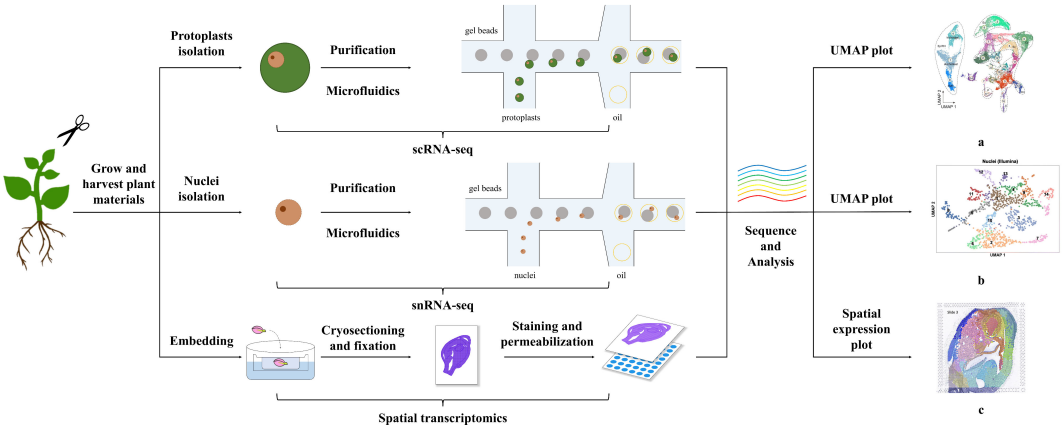



FIGURE 1
Methods of Single-Cell Transcriptomics and Spatial Transcriptomics in Plants. The figure shows the general process for performing single-cell and spatial transcriptomics. Single-cell transcriptome technologies are divided mainly into protoplast-based and nucleus-based methods. Both approaches include collecting plant materials, separating protoplasts/nuclei through microfluidics after purification steps, i.e., filtration, centrifugation, and washing, confirming the viability and concentration of obtained protoplasts or cell nuclei, and sequencing and analysis. Spatial transcriptome methods differ from single-cell transcriptome methods and include embedding, cryosectioning, fixation, staining, and permeabilization steps. The plots a, b, c shown in the figure are visual diagrams of analysis results of protoplast-based sequencing, nucleus-based sequencing, and spatial transcriptomics sequencing respectively from published studies (Long et al., 2021; Zhang et al., 2021b; Liu et al., 2022a). The plots of single-cell transcriptomics, such as uniform manifold approximation and projection (UMAP) plots, present the cell classification results of sequencing, and those of spatial transcriptomics demonstrate the combination of transcriptome and spatial information.

authors referred to if they did not write the procedure for obtaining protoplasts.

First, the enzyme solution used during protoplasting differs among the different methods, but the types of reagents and enzymes used are similar in some methods. Enzymes are used to digest cell walls, whereas other reagents primarily serve to maintain the state of protoplasts or the efficiency of enzymes. Plant cell walls are mainly composed of cellulose, hemicellulose, and pectin, and the corresponding enzymes, including cellulase, pectolyase, and hemicellulose, are used to digest the cell walls (Somssich et al., 2016). However, most studies have used only cellulase and macerozymes, as macerozymes contain pectolyase and hemicellulase activities. A few scientists have added all four types of enzymes, that is cellulase, macerozyme, hemicellulose, and pectolyase, to the solution to obtain protoplasts from rice (Liu et al., 2021b). A few scientists have also added another enzyme,

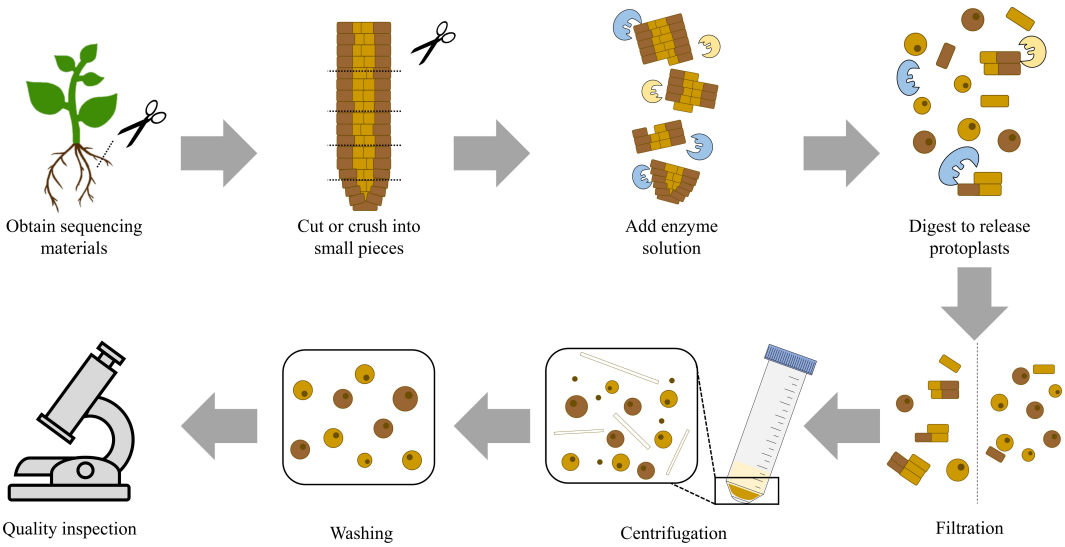



FIGURE 2
Flow Chart for Protoplast Isolation. The figure shows the general process for protoplast isolation. First, obtain sequencing materials and cut or crush them into small pieces. Then, add the prepared enzyme solution. The amount and proportion of enzymes need to be appropriate. The reaction is carried out under appropriate conditions and enzymes digest plant cell walls to release protoplasts. This step is followed by further steps that involve filtration, centrifugation, and washing. Finally, the quality and quantity of the obtained protoplast solution are tested. If both meet the expectations, the preparation of the protoplast solution is completed.

TABLE 2 Comparisons between methods used for protoplast-based single-cell transcriptome analysis in partially published studies.

| Species | Position | Composition of enzyme solution (generally used) | Composition of enzyme solution (optionally used) | Optional incubation condition | Optional centrifugation condition | Optional washing solution | Optional sequencing protocol | Reference |
|---|---|---|---|---|---|---|---|---|
| *Oryza sativa* L. | roots, proximal shoots | 1. BSA<br>2. $CaCl_2$<br>3. Cellulase<br>4. Macerozyme<br>5. Mannitol<br>6. MES | 1. Hemicellulase<br>2. KCl<br>3. NaCl<br>4. Mercapto-ethanol<br>5. Pectolyase | 1. 10–30 min at vacuum pump at room temperature and then 2–2.5 h at room temperature (about 28 °C) with shaking at 50–70 rpm | 1. 130 g for 5 min<br>2. 100 g for 3 min | 1. 8% mannitol | 1. 10× Genomics<br>2. BD Rhapsody system<br>3. Cel-Seq2<br>4. SMART-seq | (Han et al., 2017; Xie et al., 2020; Liu et al., 2021b; Wang et al., 2021c; Zhang et al., 2021b; Zong et al., 2022) |
| | leaves, inflorescences | | | 1. 2-3 h at room temperature | 1. 200 g for 3 min | 1. 8% mannitol<br>2. PBS-BSA | | |
| | leaves | | | 1. 3 h at 28 °C with shaking at 70 rpm in the dark | 1. 300 g | 1. W5 solution | | |
| *Arabidopsis thaliana* | roots | 1. BSA<br>2. $CaCl_2$<br>3. Cellulase<br>4. KCl<br>5. Macerozyme<br>6. Mannitol<br>7. MES | 1. Actinomycin D<br>2. Cordycepin<br>3. DTT<br>4. HCl<br>5. Mercapto-ethanol<br>6. $MgC_{l2}$<br>7. Pectolyase | 1. 1 h at room temperature with shaking at 75–85 rpm<br>2. 1 h at 75 rpm (1 g whole roots and 10 mL enzyme solution)<br>3. 2 h at 20 °C at 200 rpm<br>4. 1 h at room temperature at 75 rpm (1,500 seedling roots and 10 mL enzyme solution) | 1. 500 g for 5 –10 min<br>2. 200 g for 6 min | 1. 8% mannitol<br>2. Protoplast solution without enzymes<br>3. W5 solution | | (Yoo et al., 2007; Bargmann and Birnbaum, 2010; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Wendrich et al., 2020; Gala et al., 2021; Hou et al., 2021; Kim et al., 2021; Lopez-Anido et al., 2021; Yang et al., 2021; Zhang et al., 2021c; Apelt et al., 2022; Liu et al., 2022c; Liu et al., 2022d; Shahan et al., 2022; Tenorio Berrio et al., 2022) |
| | leaves | | | 1. 30 min at room temperature<br>2. 2 h with gentle shaking (whole aerial tissue or first true leaf and 15 mL enzyme solution)<br>3. 30 min in the dark with vacuum treatment and then 3 h at room temperature (10–20 leaves in 5–10 ml enzyme solution)<br>4. 2 h at 30 rpm | 1. 500 g for 5–10 min<br>2. 100 g for 1–7 min | 1. W5 solution<br>2. protoplast solution without enzymes | | |
| | shoot apexes, leaves | | | 1. 2 h at room temperature | 1. 500 g for 5–10 min | 1. 8% mannitol | | |

*(Continued)*





10.3389/fpls.2023.1185377

TABLE 2 Continued

| Species | Position | Composition of enzyme solution (generally used) | Composition of enzyme solution (optionally used) | Optional incubation condition | Optional centrifugation condition | Optional washing solution | Optional sequencing protocol | Reference |
|---|---|---|---|---|---|---|---|---|
| | cotyledons | | | 1. 10 min for vacuum infiltration and then 4 h for incubation | | | | |
| | ovule | | | 1. 3 h at 100 rpm | | 1. protoplast solution without enzymes | | |
| *Zea mays* L. | shoot apical meristem | 1. BSA<br>2. $CaCl_2$<br>3. Cellulase (Cellulase Onozuka R-10, Cellulase Onozuka RS)<br>4. Hemicellulose<br>5. HCl<br>6. KCl<br>7. Macerozyme<br>8. Mannitol<br>9. MES<br>10. Pectolyase | 1. MOPS<br>2. β-mercaptoethanol | 1. 2 h at 29 °C with gentle shaking | 1. 250 g for 3 min | 1. Washing buffer (0.65 M mannitol, 10 mM MOPS pH7.5 and 10 mM L-Arginine, at pH 7.5) | | (Ortiz-Ramirez et al., 2018; Satterlee et al., 2020; Xu et al., 2021a) |
| | ears, roots | | | 1. 45 minutes at room temperature with gentle shaking | 1. 500 g for 3 min | 1. Washing solution (1.82 g mannitol, 0.097 g MES, 1 M KCl, 1M $CaCl_2$, 0.025 g BSA, adjust pH solution to 5.7 with 1M TRIS) | | |
| *Gossypium hirsutum* | roots | 1. BSA<br>2. $CaC_{l2}$<br>3. Cellulase<br>4. KCl<br>5. Mannitol<br>6. MES<br>7. Pectolyase | | 1. 1 h under dark with vacuum treatment and then 6 h at 25 °C under dark at 80 rpm | | 1. WB buffer | | (Liu et al., 2022b) |
| *C. sinensis* var. *sinensis* | leaves | 1. BSA<br>2. $CaCl_2$<br>3. Cellulase<br>4. KCl<br>5. Macerozyme<br>6. Mannitol<br>7. MES | 1. Snailase<br>2. β-mercaptoethanol | 1. 5 min with vacuum treatment and then 4 h at 25 °C under dark with gentle shaking | 1. 100 g for 2 min | 1. W5 solution | | (Xu et al., 2021b; Wang et al., 2022) |
| | fresh samples | | | 1. 30 min under the pressure of −0.1 MPa and then at 25 °C with gentle shaking at 45 rpm under dark (1 g fresh weight samples and 10 mL enzyme solution) | 1. 200 g for 3 min | | | |

(Continued)





10.3389/fpls.2023.1185377

TABLE 2 Continued

| Species | Position | Composition of enzyme solution (generally used) | Composition of enzyme solution (optionally used) | Optional incubation condition | Optional centrifugation condition | Optional washing solution | Optional sequencing protocol | Reference |
|---|---|---|---|---|---|---|---|---|
| *Populus alba* | stems | 1. BSA<br>2. $CaCl_2$<br>3. Cellulase<br>4. KCl<br>5. Macerozyme<br>6. Mannitol | 1. MES | 1. 2.5 h in the dark at room temperature with shaking at 50–55rpm<br>2. 20 min in the dark at room temperature and then gently shake by hand for 30 s (four 10 cm stem segments and 40 mL enzyme solution) | 1. 100 g for 4 min | 1. W5 solution | | (Lin et al., 2014; Chen et al., 2021b) |
| *Triticum aestivum* L. | guard cells | 1. Actinomycin D<br>2. $CaCl_2$<br>3. Cellulase<br>4. Cordycepin<br>5. $KH_2PO_4$<br>6. Macerozyme<br>7. MES<br>8. $MgCl_2$<br>9. RNase inhibitor | | 1. 1.5 h at 25 °C | | | | (Wang et al., 2021a) |
| *Arachis hypogaea* L. | leaves | 1. BSA<br>2. Cellulase<br>3. Macerozyme<br>4. Mannitol (without $Ca^{2+}$ and $Mg^{2+}$)<br>5. MES<br>6. Pectinase | | 1. 2 h at 25 °C with shaking at 40 rpm | | 1. 8% mannitol | | (Liu et al., 2021a) |

The compositions of enzyme solutions, incubation conditions, centrifugation conditions, the types of washing solutions, and sequencing protocols used in partially published studies are listed in this table.
1. If the enzyme solution composition is used in more than half of the references, we consider it as generally-used composition.
2. The solution needs to be heated in a warm bath at approximately 55 °C after adding part of materials (enzymes, etc.) and then, after cooling to room temperature (approximately 25 °C), the solution is added in the rest materials.
3. If the article has mentioned the ratio of plant materials and enzyme solution, we will mark it in brackets.



10.3389/fpls.2023.1185377

snailase, to reduce digestion time (Wang et al., 2022). Adding more enzymes is conducive to shortening the enzymatic hydrolysis time but can cause enzyme waste. The composition of cell walls differs among different plant species and organs. Therefore, determining the appropriate enzyme content and ratio is conducive for efficient enzymatic digestion. Bovine serum albumin (BSA), $CaCl_2$, mannitol, and 2-(N-morpholino)ethanesulfonic acid (MES) are generally used to protect protoplasts. An enzyme solution with added BSA has the benefit of protecting the protoplasts, and a few researchers have used dithiothreitol (DTT) to perform the same function as BSA (Kim et al., 2021). Although BSA seems necessary to stabilise the enzyme solution, few researchers have used it (Xie et al., 2020). Salts such as $CaCl_2$ are used to maintain the osmotic pressure of protoplasts *in vivo* and *in vitro* and prevent protoplasts from absorbing water and breaking. There do exist differences—for example, when studying *A. thaliana* and *Z. mays*, researchers are more likely to add KCl into the enzyme solution than when studying *O. sativa* as shown in Table 2. There are still differences in details within the same species, even when using the same parts of the materials in different studies. For example, when preparing rice protoplasts, a few studies added KCl to the digestion solution, whereas several other studies did not; however, mercaptoethanol was added (Han et al., 2017; Liu et al., 2021b; Wang et al., 2021c). NaCl has been added to enzyme solutions in a few studies (Xie et al., 2020). In addition, Table 2 shows differences among the types of mercaptoethanol used in solution: a few used 2-mercaptoethanol, others used 4-mercaptoethanol acid, and a few others used β-mercaptoethanol.

After determining the composition of the enzyme solution, reagents and enzymes were mixed. Usually, scientists dissolve cellulase, macerozyme, mannitol, MES, and certain enzyme solution reagents in water and heat the mix to activate enzymes in a warm bath at approximately 55 °C for nearly 10 min. Upon heating, the turbid enzyme solution becomes clear. After heating, it is cooled to room temperature, i.e., approximately 25 °C, and the rest of the materials, including BSA, $CaCl_2$, and mercaptoethanol, are added before adding plant materials. Unlike the general methods of mixing plant materials and the entire digestion solution, a few techniques involve mixing plant materials and the enzyme solution and incubating the mixture in the dark for the reaction. The remaining chemicals, including NaCl, $CaCl_2$, KCl, and MES, are added for further enzymatic reactions to release protoplasts (Yoo et al., 2007; Xie et al., 2020; Gala et al., 2021).

After mixing the enzyme solution with plant materials, it is essential to provide sufficient time and appropriate reaction conditions to ensure the release of protoplasts. Most studies listed in Table 2 did not specify the proportions of plant materials and enzyme solutions. Table 2 lists the incubation conditions. A few groups used a vacuum pump to isolate protoplasts in experiments to reduce the digestion reaction time, whereas others did not. Therefore, the reaction times of others were longer than reaction times of those that used the vacuum pump. This is because negative pressure promotes the infiltration of the enzyme solution into plant materials, such that the contact surface between the enzyme and plant cells increases, thereby reducing the reaction time. In addition, it is helpful to incubate a mixture of enzyme solution and plant materials in the dark or to shake this mixture using a machine or manually to reduce the reaction time. When shaking the enzyme solution, it is necessary to shake it gently at a speed of 40–80 rpm to avoid cell fragmentation caused by violent shaking. However, a few groups used 200 rpm shaking speed (Denyer et al., 2019). The reaction times used in different studies to release protoplasts are similar, but there are still differences in the details. For example, when digesting the leaves of *A. thaliana*, the digestion time used by a few groups was 30 min, whereas others chose to digest for 2 h, probably because the leaves they used were different; that is, leaves from the leaf base regions or the third leaf (Liu et al., 2022c; Tenorio Berrio et al., 2022). Although different digestion times are listed in Table 2, it is only for reference—digestion times can differ based on differences in the digestion conditions, species, and parts of plants used, and it is necessary to check whether most of the plant cell walls are digested at least every half an hour. If measures are taken to shorten the time, the enzyme content in the enzyme solution will be higher, and if the plant tissues used are young or have a short growth time, the digestion time will be shorter.

After protoplast release, cells are filtered, centrifuged, and washed. Filtration aims to remove undigested plant cells and obtain protoplasts; therefore, the filter must be of the proper size. The size of plant protoplasts is small, e.g., *A. thaliana* mesophyll protoplast diameter ranges from 30 to 50 μm, and therefore, most groups used 30–80 μm nylon mesh or cell strainer to filter protoplasts (Yoo et al., 2007). A few groups filtered the protoplasts only once, whereas other groups filtered them twice or more times, and other groups filtered the protoplasts through two different sizes of filters to obtain more protoplasts and less undigested plant cells (Wang et al., 2021c; Zhang et al., 2021b; Liu et al., 2022c). Centrifugation helps concentrate protoplasts, and the centrifugation conditions differ among different plant species. Table 2 lists the centrifugation conditions used in various studies. Centrifugation speed and time depend on the characteristics of the plant materials, such as their fragility, type, and the number of cells produced during digestion (Bargmann and Birnbaum, 2010). High centrifugation speed can increase both the recovery rate of protoplasts and the risk of cell disruption (Yoo et al., 2007). As shown in Table 2, when centrifuging rice protoplasts, the centrifugation speed (e.g., 100 g, 130 g, or 200 g) is slower and the centrifugation time (e.g., 3 min or 5 min) is shorter than that used for centrifuging *A. thaliana* (e.g., 500 g for 5–10 min). The differences in centrifugation speeds may result from the different characteristics of the protoplasts of these two plants or from differences in the isolation steps used in published studies. During centrifugation, a few groups maintained the solution at constant low-temperature conditions, such as 4 °C or −4 °C, to protect protoplasts from damage (Denyer et al., 2019; Satterlee et al., 2020; Xu et al., 2021a). It is also possible to collect small protoplasts by centrifugation at low speeds and then increase the speed of collecting larger protoplasts, avoiding the breakage of small protoplasts and the loss of large protoplasts caused by insufficient centrifugation speeds. Washing is also necessary because it can help remove specific ions or tissue residues that affect the downstream steps, such as undigested plant tissues and $Mg^{2+}$ ions (Liu et al., 2022d). During the preparation of the washing solution, the entry of

impurities is avoided, and an appropriate osmotic pressure amenable to protoplasts is maintained in the washing solution. Scientists have typically used 8% mannitol or a protoplast solution without enzymes for washing; however, a few scientists have also used W5 solution, WB buffer, or other washing buffers (Table 2). It is worth mentioning that the components of W5 solutions prepared by different groups showed little difference (Chen et al., 2021b; Xu et al., 2021b). After washing, the protoplasts were centrifuged again to separate the washing solution from the protoplasts. The centrifugation speed and time can be the same as those used in the centrifugation step, or at a slower rate, shorter time, or both. The protoplasts are resuspended in mannitol, washing solution, or other solutions based on the sequencing protocol, which helps adjust the concentration of protoplasts and avoid influencing reverse transcription reactions (Chen et al., 2021b). The viability and concentration of the protoplasts are then confirmed for better sequencing results. Trypan blue, 4′,6-diamidino-2-phenylindole, fluorescein diacetate, or acridine orange-propidium iodide are used to select protoplasts, and a haemocytometer is used to measure the protoplast concentration. The viability and concentration of protoplasts need to be adjusted to reach the standard of sequencing methods, and the concentration of protoplasts can be adjusted using a resuspension solution. Next, an scRNA-seq library is prepared and sequenced according to different protocols.

The process of isolating plant nuclei is similar to that used to separate plant protoplasts. The plant material used to extract nuclei can be either fresh or frozen. Plant tissues are cut or crushed into small pieces and mixed with a nuclei extraction solution. The solutions used for nuclei extraction in the relevant studies are listed in Table 3; there are similarities, with a few differences. Compared to isolating protoplasts, plant nuclei isolation requires the lysis of plant cell membranes and protection of nuclear membranes; therefore, the solutions are relatively more destructive. The incubation is usually performed on ice. After incubation, filtration, centrifugation, and resuspension are conducted. The speed of centrifugation for isolating plant nuclei is usually faster than 1,000 g; for example, a few groups centrifuge at 2,000 g, whereas others centrifuge at 25,000 g (Giuliano et al., 1988; Dorrity et al., 2021). Plant nuclei are washed, recovered, and resuspended in a buffer solution. Some groups have conducted Percoll gradient centrifugation for impurity removal and nuclear recovery (Dorrity et al., 2021). The remaining steps are performed according to the nucleus-based sequencing protocols of the sequencing platforms.

TABLE 3 Nuclei extraction solutions.

| Species | Composition of solution | Reference |
|---|---|---|
| *Arabidopsis thaliana* | 1. Sucrose<br>2. $MgCl_2$<br>3. Tris·HCl<br>4. Protease Inhibitor Tablet | (Dorrity et al., 2021) |
| | 1. Ficoll 400<br>2. Dextran T40<br>3. Sucrose<br>4. $MgCl_2$<br>5. DTT<br>6. Triton X-100<br>7. cOmplete Protease Inhibitor Cocktail<br>8. RiboLock<br>9. Tris HCl | (Sunaga-Franze et al., 2021) |
| | 1. Ficoll 400<br>2. Dextran T40<br>3. Sucrose<br>4. Tris<br>5. $MgCl_2$<br>6. Triton X-100<br>7. β-mercaptoethanol | (Moreno-Romero et al., 2017) |
| *Zea mays* L. | 1. Tris<br>2. EDTA<br>3. Spermine<br>4. KCl<br>5. NaCl<br>6. 2-ME<br>7. Trixton X-100 | (Marand et al., 2021) |
| *Solanum lycopersicum* | 1. Sucrose<br>2. $MgCl_2$<br>3. Tris·HCl<br>4. Phenylmethylsulphonyl fluoride<br>5. Benzamidine<br>6. 2-mercaptoethanol | (Giuliano et al., 1988) |

Different kinds of nuclei extraction solutions used for isolating nuclei in nucleus-based single-cell transcriptomics are listed in this table.

## 4 Development of plant spatial transcriptomics

Spatial transcriptomic technologies originated in mammalian systems and are widely used in mammals. A few published studies have applied omics technology to plant research. However, more strategies are available for mammalian spatial transcriptomics than for plants, mainly because of the differences between animal and plant cell structures such as cell walls. Currently, research on plant spatial transcriptomics is divided primarily into two categories: finding ways to use and optimise spatial transcriptomic methods, and using spatial technology to solve problems in plant research.

In 2017, Stefania and his group provided a method for generating and studying high-resolution and spatially resolved functional profiles in plants (Giacomello et al., 2017; Giacomello and Lundeberg, 2018). At the method level, research has not been limited to methodology in plants or the application of methods used in mammalian systems to plants, but has also focused on finding ways to improve spatial transcriptome technology in plant and mammalian systems. Spatial transcriptome analysis requires high spatial resolution and high spatial transcriptome data, such as maps in daily life, and provides an extensive range of accurate information on buildings located in an area. A few analytical techniques, such as single-cell transcriptomics, produce high-quality data, implying that they provide relatively high spatial

resolution but lose spatial information. In addition, sorting or isolation technologies, such as fluorescence-activated cell sorting (FACS), isolation of nuclei tagged in specific cell types (INTACT), and LCM, provide spatial resolution or cell-specific expression data. However, these methods have certain limitations. Plant materials must be transgenic using FACS and INTACT; however, this is a challenge for a few plant species. In addition, many factors such as cell type, yield, and purity affect the quality of the LCM and FACS methods. Technologies such as FACS require the digestion of plant cells to isolate protoplasts, which causes loss of spatial information (Giacomello et al., 2017; Gurazada et al., 2021). Fluorescent *in situ* hybridisation (FISH) is an optimal option for obtaining spatial information from plant cells; however, it can only detect one gene at a time, limiting its speed (Tirichine et al., 2009; Gurazada et al., 2021). Therefore, spatial transcriptomics technology needs to be improved. Currently, there are two main types of spatial transcriptome technology: next-generation sequencing (NGS)- and imaging-based approaches. NGS-based methods involve high-throughput sequencing based on scRNA-seq technology and a spatial barcode, whereas imaging-based methods include *in situ* sequencing (ISS) and *in situ* hybridisation (ISH). Recent research on spatial transcriptomic technologies, including ISH-based MERFISH and ISS-based STARmap, has promoted genetic and genomic studies (Moffitt et al., 2018; Wang et al., 2018; Liu et al., 2022a). Spatial indexing approaches, including 10× Visium, use NGS-based methods to quantify the gene expression profiles formed by the local hybridisation of barcodes and RNA molecules, and the spatial information is stored by barcodes. These approaches can detect the entire transcriptome in an unbiased manner without knowing the target gene in advance but must ensure the integrity of mRNA in the tissues. ISS- and ISH-based methods can intuitively analyse data at organisational spatial locations, but mRNA abundance *in situ* is low and often degraded (Elhanani et al., 2023). Although spatial transcriptomic methods face several challenges, including resolution and sensitivity, researchers are constantly attempting to overcome these difficulties (Rao et al., 2021).

Spatial transcriptomics technologies provide new insights into plant growth, development, and molecular biology. A few researchers have mapped and visualised the related gene expression in C4 and crassulacean acid metabolism in *Portulaca* at the spatial level using the 10× Genomics Visium spatial transcriptomics platform (Moreno-Villena et al., 2022). One group combined spatial and temporal transcriptome information to explore the early development of tomatoes (Zhang et al., 2016). In 2022, a Chinese group studied *Phalaenopsis* Big Chili using 10× Visium technology (Liu et al., 2022a). A few studies have analysed transcriptome data at the spatial level; however, in these studies, researchers collected plant materials in spatial order and then sequenced them separately, which preserved the spatial information and provided help in the analysis, but the information lacked high resolution (Guo et al., 2021).

Spatial transcriptome technology is helpful, but there are still few studies using this technology in plant research owing to many reasons and challenges that will be discussed later.

## 5 Methods used in plant spatial transcriptomics

In plant research, technology for preserving spatial information is not widely used. This is partly because appropriate methods have not been used in plants, and the structures of plant cells differ from those of animal cells, such as the presence of cell walls and vacuoles. In a few studies, the processes involved collecting plant materials in spatial order, separating leaves from outer to inner, sorting root and aerial tissues, and sequencing them separately (Guo et al., 2021; Apelt et al., 2022). Such methods are usually not highly resolved because they involve the sequencing of many mixed cells, providing a considerably limited variety of spatial locations. One such process involves dissecting different tissues for transcriptome analysis, such that the spatial information for the entire section is maintained, but the spatial information is distinguished according to the initially selected regions (Stahl et al., 2016).

A few scientists achieved high-resolution gene expression data, even near single-cell resolution, using high-throughput methods. In general, these methods include collecting targeted sequencing materials, embedding, cryosectioning, fixation, staining, imaging, permeabilisation, and sequencing, according to the manufacturer's instructions (Figure 1) (Giacomello et al., 2017; Giacomello and Lundeberg, 2018; Kivivirta et al., 2019; Liu et al., 2022a; Moreno-Villena et al., 2022). Embedding is usually the first step in spatial transcriptomic experiments after harvesting the targeted plant materials. The samples are embedded in cold optimal cutting temperature (OCT) compound to support the plant tissues during cryosectioning. Plant tissues have structures such as hard cell walls and large vacuoles with high water content, which increase the hardness of plant tissues after conventional freezing, making it difficult to obtain high-quality slices. Then, the plant samples are sliced at the proper temperature (−15 to −20 °C). The thickness of the sections must also be appropriate to avoid breakage and overlapping of the different tissue sections during slicing. In addition, the organisational structure of plant samples is complex, the degree of lignification varies, and the relative RNA content of the tissue is relatively low; therefore, the thickness of plant slices generally depends on the type of sample (Giacomello and Lundeberg, 2018; Liu et al., 2022a). Subsequently, the plant sections are fixed on the array at approximately 37 °C. The solutions used in the fixation step differed among different studies; a few studies used neutral formaldehyde, whereas others used methanol. The dyes used in the staining step also differ among various studies, such as toluidine blue or haematoxylin-eosin staining, and the staining time is usually 1 min, followed by imaging of the stained sections. Permeabilisation is a critical step in spatial transcriptomics following imaging. In this step, it is necessary to determine the optimum permeabilisation conditions for plant samples to release sufficient mRNA (Moreno-Villena et al., 2022). A few experiments have also included a pre-permeabilisation step before permeabilisation for partially denaturing proteins to improve the efficiency of permeabilisation (Giacomello and Lundeberg, 2018). After permeabilisation, the samples were

sequenced according to the manufacturer's protocol, including reverse transcription, tissue removal, and probe release. Finally, the water used in these experiments must be nuclease-free in specific steps to avoid degradation of the cDNA-mRNA hybrids.

Although differences exist between different methods and there are only a few relevant studies in plant research, we can learn from the methods used in mammalian systems or follow the manufacturer's instructions to optimise these methods in plant systems (Giacomello and Lundeberg, 2018).

## 6 Comparisons among protoplast-based and nucleus-based single-cell transcriptomics and spatial transcriptomics

Although protoplasts and plant nuclei can be used for single-cell sequencing, there are differences, and both have advantages and disadvantages. Comparing these two sequencing methods and the scope of their application is beneficial for attaining a deeper understanding of single-cell transcriptomics in plant systems and for exploring future research directions.

As shown in Tables 2, 3, enzyme solutions, centrifugation speeds, and other parameters differ between the two methods. For example, because of differences in sequencing materials, these two methods use different solutions to digest or isolate plant cells. When separating protoplasts, researchers usually use enzymes to digest the cell walls of plant cells and add various reagents to protect the cell membrane and maintain protoplast shape. In contrast, when isolating nuclei, researchers usually do not use enzymes, but instead use other reagents, such as Triton X-100, to separate plant nuclei, and the cell membranes need to be destroyed (Sikorskaite et al., 2013).

When separating protoplasts, artefacts affecting the transcriptome are introduced by the addition of enzymes. Scientists have confirmed that protoplasts affect transcriptomics (Birnbaum et al., 2003; Denisenko et al., 2020; Conde et al., 2021; Sunaga-Franze et al., 2021). The addition of enzymes necessitates an appropriate incubation temperature, such that the enzymes are in the most suitable state for digestion. However, cellular machinery is also active at the appropriate temperature for enzymatic action, leading to alterations in gene expression. A few scientists have found that when obtaining single cells from animals, the addition of enzymes at the incubation temperature can lead to a dramatic increase in gene expression after incubation at 37 °C (Adam et al., 2017; van den Brink et al., 2017; Potter, 2018; Denisenko et al., 2020). Similar results were obtained when plant cells were digested to obtain protoplasts. Isolating plant nuclei for single-cell transcriptomics would not face this challenge, as no enzymes are used and incubation is performed on ice, avoiding the activation of gene expression. Therefore, from the perspective of introducing artefacts and anomalous gene expression, plant nucleus-based single-cell transcriptomics sequencing is more precise, convenient, and applicable than protoplast-based single-cell transcriptomics. Researchers have also developed methods to overcome the limitations of protoplast release, such as performing independent bulk RNA sequencing experiments to eliminate the effect of enzyme addition or using transcription inhibitors (Potter, 2018; Denyer et al., 2019; Sunaga-Franze et al., 2021). Specifically, α-amanitin inhibits transcription via RNA polymerase II, thereby preserving gene expression patterns (Potter, 2018). Some researchers have extracted RNA from protoplast and unprotoplast plant tissues, identified genes induced by protoplast, and removed them from analyses (Denyer et al., 2019; Jean-Baptiste et al., 2019; Shulse et al., 2019). Protoplast-based single-cell transcriptome methods are not applicable, whereas nucleus-based techniques can. Protoplasts are used for plant single-cell transcriptomics; however, the methods for isolating protoplasts differ for different plant materials, even for the same plant and their exact position in the plant. As mentioned above, this is mainly because many factors affect the enzyme solution and incubation conditions adopted in the experiments. The main components of cell walls are similar, but differences in the functions of cells result in the production of distinct cell wall polymers surrounding each cell. Therefore, cell walls are considered heterogeneous when comparing plant cells of different types and developmental stages (Somssich et al., 2016). Thus, the methods for obtaining protoplasts lack applicability, leading to limitations in protoplast-based single-cell transcriptomics. In addition, these protoplast-based methods rely heavily on the quality and quantity of protoplasts and their compatibility with sequencing systems (Thibivilliers et al., 2020). Protoplast suspensions need to contain more intact protoplasts and less or no cell debris and damaged cells to avoid low cell numbers and quality, mRNA cell leakage, and obstruction of microfluidic systems, thereby ensuring that the microfluidic systems run correctly (Conde et al., 2021). Nucleus-based single-cell transcriptomics can overcome these problems owing to the introduction of a general method for plant nucleus isolation, which is an alternative to protoplast isolation methods and has been tested for its applicability to several different types of plants (Sunaga-Franze et al., 2021).

The applicability of these two procedures is not only reflected in isolation techniques, but also in the requirements for cell sequencing, such as the size of sequencing materials or freshness of sequencing samples (Hwang et al., 2018; Denisenko et al., 2020; Nadelmann et al., 2021). Generally, to avoid obstruction of sequencing systems, the sequencing platform restricts the cell size. The protoplast was larger than the nucleus, which increased the possibility of obstruction. The freshness of protoplasts is also critical for sequencing, and it is necessary to maintain their shape to avoid membrane breakage. However, nuclei can be obtained from any plant cell, regardless of its freshness and size, and even from frozen plant tissues (Krishnaswami et al., 2016; Lake et al., 2016; Bakken et al., 2018; Wu et al., 2019; Nadelmann et al., 2021). Protoplast-based sequencing is a stringent requirement for sequencing materials.

Nuclei seem to be more appropriate for single-cell transcriptomics than protoplasts; however, isolating nuclei from plant cells results in the loss of cell cytoplasm, chloroplasts, and mitochondria, implying that single-nucleus RNA-seq (snRNA-seq) usually contains less transcription information. Therefore, snRNA-

seq has limitations in obtaining information from particular cell types because it only uses a part of the cell, that is, the nucleus, instead of an entire cell for transcriptional profiling (Hu et al., 2018; Maitra et al., 2021). The cell nucleus contains only 10–20% of all cellular transcripts; therefore, nuclear RNA does not entirely reflect the RNA from the whole cell, which may increase the difficulties in the analysis (Bakken et al., 2018; Potter, 2018; Nadelmann et al., 2021). Moreover, snRNA-seq excludes critical gene information from outside the nucleus and does not completely represent the single-cell transcriptome. From this point of view, protoplasts are more accurate than plant nuclei for single-cell transcriptome analysis. However, in research on frozen samples, precious frozen samples, complex plant tissues that are difficult to digest, and those that are easily influenced by gene expression changes during protoplast digestion, plant nuclei are still more appropriate for sequencing.

The most significant difference between single-cell and spatial transcriptomics is that the former is highly resolved but loses spatial information, whereas the latter preserves spatial information but fails to achieve precision. Therefore, if more detailed research is preferred, such as further exploration of cell types and their molecular characteristics in response to different environmental stimuli, single-cell transcriptomics is often used. If the characteristics of different spatial regions or the molecular characteristics of cells in a certain region of plant organs are preferred, for example, by constructing a spatiotemporal atlas of the development of floral organs, spatial transcriptomics is often used. Microfluidic technology is typically used to separate single cells in single-cell transcriptomics. Microarray technology is typically used to store spatial information in transcriptomics. Each spot on the microarray has millions of oligonucleotide probes. An oligonucleotide probe sequence used in spatial transcriptomics usually has a spatial barcode to store spatial information, whereas the barcode in single-cell transcriptomics is used to distinguish different cells (Gurazada et al., 2021).

## 7 Challenges and perspectives

Single-cell transcriptomics is a mature analytical technology in animal research, but a rising and naive front field in plant research, which still needs further exploration and development. Earlier sections revealed problems in both protoplast-based and plant nucleus-based single-cell transcriptomics, and these difficulties need to be urgently solved. In contrast, this is the era of multi omics, and finding ways to combine single-cell transcriptomics with other omics is noteworthy. Integrating single-cell transcriptome analysis with other single-cell technologies or omics technologies may provide a list of cell types, and more importantly, new insights into the regulatory logic and spatial organisation among cells (Stuart et al., 2019).

Plant materials contain different cell types, a few of which are resistant to digestion owing to their secondary cell walls, implying that, based on currently used methods, such cells pose challenges in protoplast isolation, causing a biased analysis of cell types (Bezrutczyk et al., 2021; Conde et al., 2021; Farmer et al., 2021). The different positions of cells in plant materials also introduce differences in their area exposed to digestive enzymes; for example, a few cells located inside the plant materials and surrounded by other cells are challenging to digest. These cells are readily missed during the analysis because they are removed during filtration. A few groups have discovered that central stele cells are more difficult to capture than epidermal cells; thus, the preparation of plant tissue protoplasts is biased for cells located on the outer surface of plant materials and sufficiently exposed to digestion enzymes (Denyer et al., 2019; Jean-Baptiste et al., 2019; Shulse et al., 2019; Farmer et al., 2021). In addition, a few groups have observed that not only the position of cells, but also their developmental stage influences digestion, as younger cells are more readily digested than mature cells, leading to increased representation in the analysis (Jean-Baptiste et al., 2019; Ryu et al., 2019). Therefore, a few general and effective methods are needed to isolate protoplasts that can avoid errors caused by cell types, relative positions of cells, or developmental stages.

In addition, single-cell sequencing technology requires improvement to achieve more precise sequencing results. For example, droplet-based methods such as the 10× Chromium method help generate an RNA-seq library. In this method, a single cell or nucleus is encapsulated inside a gel bead that can supply a barcoded oligonucleotide for reverse transcription. The droplet-based method can be used to analyse thousands of cells in a single experiment. However, in a few cases, there is more than one cell or nucleus inside a droplet or a suboptimal number of beads inside a droplet, thereby causing problems in the analysis. One group tested the probability of this event and observed that approximately 20% of the droplets contained either no beads or more than one bead in their experiments (Lareau et al., 2020; Sunaga-Franze et al., 2021). Nanowell-based systems, such as Takara iCELL8, can also generate an RNA-seq library and do not suffer from this problem because they capture single cells or nuclei in a nanowell. Researchers can check the number of cells or nuclei by using microscopy. However, the chips used in this method have only a fixed number of nanowells, resulting in limited scalability (Sunaga-Franze et al., 2021). Therefore, it is important to improve the relevant technology to ensure that one droplet contains one bead and one cell or nucleus and provide better scalability.

In single-cell transcriptomics, cells are digested and isolated from their original position in the tissue, and their spatial information is lost. Scientists have successfully mapped scRNA-seq data by inferring cellular sources and locations from scRNA-seq data and *in situ* RNA patterns (Achim et al., 2015; Satija et al., 2015). Scientists have also constructed a 3D atlas in plant research by combining scRNA-seq and microscopy-based 3D spatial reconstruction (Neumann et al., 2022). Recent studies have used spatial transcriptomics to preserve spatial information. Single-cell transcriptome analysis can be strengthened by combining it with a spatial transcriptome analysis. The cell types of whole plant tissues can be classified using single-cell transcriptome analysis, and the spatial information of whole tissues can be obtained using a spatial transcriptome. Researchers have established an *in situ* single-cell spatial transcriptome method in plants based on stereo-seq and applied it to study *A. thaliana* leaves (Chen et al., 2022; Xia et al., 2022). Their method differed from the

single-cell transcriptome methods described, but was similar to spatial transcriptome methods. They sectioned frozen plant tissues and adhered these plant sections to a stereo-seq chip for downstream operations. There are a few other integration strategies, and Longo et al. identified two main types of algorithms: deconvolution and mapping (Longo et al., 2021). Methods based on deconvolution algorithms separate discrete cellular subpopulations from the mRNA transcript mix at each capture point based on single-cell data. SPOTlight is such a method, integrating single-cell and spatial transcriptome, centering on a seeded regression using non-negative matrix factorisation, initialising with cell type marker gene, and non-negative least squares, and deconvoluting capture locations (Elosua-Bayes et al., 2021). There are other algorithms for deconvoluting spatial information, such as dampened weighted least-squares, negative binomial distribution, and Poisson distribution models (Andersson et al., 2020; Dong and Yuan, 2021; Cable et al., 2022). Some scientists believe that single-cell transcriptome data and spatial transcriptome data follow a particular probability distribution, and based on these hypotheses, a few probabilistic models have been proposed to help integrate these two types of data. Mapping methods combine these two types of transcriptome data, mapping the designated scRNA-based cell subtypes to each cell on a high-plex RNA imaging map, and mapping each scRNA-seq cell to a specific niche or region of tissue. Similar to deconvolution methods, a few mapping methods use probabilistic models, such as the variation Bayesian mean-field approximation (Qian et al., 2020). In addition, other methods, such as CellTrek, combine single-cell and spatial transcriptome data using co-embedding and metric learning approaches (Wei et al., 2022).

Similar to single-cell transcriptomics, research on spatial transcriptomics is more prevalent in mammals than in plants, partially because of differences between mammalian and plant cells. Different plant cell structures and organelles, including cell walls, vacuoles, chloroplasts, and the presence of secondary metabolites, make it challenging to obtain spatial transcriptomic data from plant cells or lead to cryosectioning problems (Bourgaud et al., 2001; Cosgrove, 2005; Giacomello et al., 2017; Gurazada et al., 2021). Specifically, owing to the cell walls, plant cell sections can neither be extremely thick nor extremely thin to avoid reducing mRNA release or capture and causing sectioning artefacts or low transcript levels (Gurazada et al., 2021). In addition, because a few plant tissues, such as leaves and roots, have curvatures, many sections are required to obtain complete information. Currently, there are few solutions to the challenges caused by these differences, and it is often necessary to optimise the method for specific plant species or tissues (Giacomello et al., 2017; Giacomello and Lundeberg, 2018; Gurazada et al., 2021). Furthermore, challenges have been introduced by spatial transcriptome technology, including resolution, sensitivity, throughput, and accessibility (Rao et al., 2021).

In addition, the related applications of single-cell transcriptomics and spatial transcriptomics can be explored further, whether using one of these two techniques or in combination with other analytical techniques. For example, although there has been comprehensive research on the root cell types of plants such as *A. thaliana* and *O. sativa*, there is almost no relevant research on the cross-species analysis of roots at the single-cell level. This cross-species analysis may provide a basis for analysing differences in elemental absorption among different plants. Relevant animal studies have identified a new cell subtype, defined signature genes, and revealed differences in drug absorption among species through cross-species analysis (Li et al., 2022). Consideration should be given to the spatial distribution of elements in plants to capture thorough spatial information. Combining single-cell and spatial transcriptome analyses with the distribution patterns and speciation of one or more plant tissue elements offers several opportunities, particularly for studying the effects of stress caused by these elements. In recent years, techniques for analysing elemental distribution and speciation in plant materials have been developed, making it possible to draw an element-based map with high resolution (Lu et al., 2013; Lu et al., 2017). The combination of these techniques will provide new insights into plant research. For example, a combination of these three techniques could be used to explore the effects of different elemental stresses on plant roots. Analysis of the elemental distribution would reveal the characteristics of the absorption of this element and its storage and transportation mechanisms in plant roots. Spatial transcriptome and single-cell transcriptome analyses help analyse the effect of this element on gene expression in plant roots, or the difference in absorption and storage caused by the difference in gene expression, to provide a gene reference for cultivating low-accumulation plant varieties.

## 8 Conclusion

Both single-cell and spatial transcriptome analyses are helpful but still developing technologies in plant research, although they are relatively mature in mammalian systems. Single-cell transcriptomics provides insights into each cell type in plant tissues, such as classifying the cell type and analysing the cell developmental trajectory, and there is likely to be more research investigating functional genes in the future. Some problems in plant single-cell transcriptomics are common to single-cell transcriptomics in all organisms, such as those in sequencing techniques, whereas other difficulties are unique to plant research, including the lack of standard methods for isolating protoplasts. Owing to differences in cellular features and culture conditions, digestion solutions, and incubation conditions for releasing protoplasts differ among studies, leading to a lack of applicability. The snRNA-seq method, which has several advantages over scRNA-seq using protoplasts, can overcome this problem; however, it lacks information outside the nuclei. Spatial transcriptome technology provides spatial information on targeted tissues; however, it fails to reach single-cell resolution. Therefore, in the future, the single-cell transcriptome technique requires more effort to increase the applicability of methods for releasing protoplasts, and the spatial transcriptome technique requires more effort to achieve high resolution. More importantly, the application of single-cell transcriptomics and spatial transcriptomics needs to be explored, such as cross-species analysis using single-cell transcriptomics in the roots of different plants, and the combination of single-cell analysis and other

analyses, such as spatial transcriptome analysis and spatial element distribution analysis, also needs to be emphasised to compensate for their limitations and create ample possibilities for research.

## Author contributions

CC and LL conceived the manuscript. CC wrote the manuscript. YG and LL reviewed and revised the manuscript. All the authors contributed to the manuscript and agreed to the submitted version.

## Funding

This work was supported by the National Natural Science Foundation of China (22276164), the Key Research and Development Project of Science and Technology Department of Zhejiang Province (2023C02020), and the National Natural Science Foundation of Zhejiang Province (LZ22C150004).

## Acknowledgments

We thank Editage for their assistance with English language 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



## References

Achim, K., Pettit, J. B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., et al. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33 (5), 503–509. doi: 10.1038/nbt.3209

Adam, M., Potter, A. S., and Potter, S. S. (2017). Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development* 144 (19), 3625–3632. doi: 10.1242/dev.151142

Andersson, A., Bergenstrahle, J., Asp, M., Bergenstrahle, L., Jurek, A., Fernandez Navarro, J., et al. (2020). Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* 3 (1), 565. doi: 10.1038/s42003-020-01247-y

Apelt, F., Mavrothalassiti, E., Gupta, S., Machin, F., Olas, J. J., Annunziata, M. G., et al. (2022). Shoot and root single cell sequencing reveals tissue- and daytime-specific transcriptome profiles. *Plant Physiol.* 188 (2), 861–878. doi: 10.1093/plphys/kiab537

Bakken, T. E., Hodge, R. D., Miller, J. A., Yao, Z., Nguyen, T. N., Aevermann, B., et al. (2018). Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PloS One* 13 (12), e0209648. doi: 10.1371/journal.pone.0209648

Bargmann, B. O., and Birnbaum, K. D. (2010). Fluorescence activated cell sorting of plant protoplasts. *J. Vis. Exp.* (36), e1673. doi: 10.3791/1673

Bezrutczyk, M., Zollner, N. R., Kruse, C. P. S., Hartwig, T., Lautwein, T., Kohrer, K., et al. (2021). Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. *Plant Cell* 33 (3), 531–547. doi: 10.1093/plcell/koaa055

Birnbaum, K., Shasha, D. E., Wang, J. Y., Jung, J. W., Lambert, G. M., Galbraith, D. W., et al. (2003). A gene expression map of the Arabidopsis root. *Science* 302 (5652), 1956–1960. doi: 10.1126/science.1090022

Bourgaud, F., Gravot, A., Milesi, S., and Gontier, E. (2001). Production of plant secondary metabolites: a historical perspective. *Plant Sci.* 161 (5), 839–851. doi: 10.1016/s0168-9452(01)00490-3

Cable, D. M., Murray, E., Zou, L. S., Goeva, A., Macosko, E. Z., Chen, F., et al. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* 40 (4), 517–526. doi: 10.1038/s41587-021-00830-w

Chen, A., Liao, S., Cheng, M., Ma, K., Wu, L., Lai, Y., et al. (2022). Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 185 (10), 1777–1792.e1721. doi: 10.1016/j.cell.2022.04.003

Chen, H., Lv, Y., Yin, X., Chen, X., Chu, Q., Zhu, Q. H., et al. (2021a). Effects of Sample Size on Plant Single-Cell RNA Profiling. *Curr. Issues Mol. Biol.* 43 (3), 1685–1697. doi: 10.3390/cimb43030119

Chen, Y., Tong, S., Jiang, Y., Ai, F., Feng, Y., Zhang, J., et al. (2021b). Transcriptional landscape of highly lignified poplar stems at single-cell resolution. *Genome Biol.* 22 (1), 319. doi: 10.1186/s13059-021-02537-2

Conde, D., Triozzi, P. M., Balmant, K. M., Doty, A. L., Miranda, M., Boullosa, A., et al. (2021). A robust method of nuclei isolation for single-cell RNA sequencing of solid tissues from the plant genus Populus. *PloS One* 16 (5), e0251149. doi: 10.1371/journal.pone.0251149

Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* 6 (11), 850–861. doi: 10.1038/nrm1746

de Bekker, C., Bruning, O., Jonker, M. J., Breit, T. M., and Wosten, H. A. (2011). Single cell transcriptomics of neighboring hyphae of Aspergillus niger. *Genome Biol.* 12 (8), R71. doi: 10.1186/gb-2011-12-8-r71

Denisenko, E., Guo, B. B., Jones, M., Hou, R., de Kock, L., Lassmann, T., et al. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* 21 (1), 130. doi: 10.1186/s13059-020-02048-6

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M. C. P. (2019). Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev. Cell* 48840–52 (6), e845. doi: 10.1016/j.devcel.2019.02.022

Ding, J., Adiconis, X., Simmons, S. K., Kowalczyk, M. S., Hession, C. C., Marjanovic, N. D., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* 38 (6), 737–746. doi: 10.1038/s41587-020-0465-8

Dong, R., and Yuan, G. C. (2021). SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Genome Biol.* 22 (1), 145. doi: 10.1186/s13059-021-02362-7

Dorrity, M. W., Alexandre, C. M., Hamm, M. O., Vigil, A. L., Fields, S., Queitsch, C., et al. (2021). The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. *Nat. Commun.* 12 (1), 3334. doi: 10.1038/s41467-021-23675-y

Elhanani, O., Ben-Uri, R., and Keren, L. (2023). Spatial profiling technologies illuminate the tumor microenvironment. *Cancer Cell* 41 (3), 404–420. doi: 10.1016/j.ccell.2023.01.010

Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I., and Heyn, H. (2021). SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* 49 (9), e50. doi: 10.1093/nar/gkab043

Farmer, A., Thibivilliers, S., Ryu, K. H., Schiefelbein, J., and Libault, M. (2021). Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level. *Mol. Plant* 14 (3), 372–383. doi: 10.1016/j.molp.2021.01.001

Gala, H. P., Lanctot, A., Jean-Baptiste, K., Guiziou, S., Chu, J. C., Zemke, J. E., et al. (2021). A single-cell view of the transcriptome during lateral root initiation in Arabidopsis thaliana. *Plant Cell* 33 (7), 2197–2220. doi: 10.1093/plcell/koab101

Giacomello, S., and Lundeberg, J. (2018). Preparation of plant tissue to enable Spatial Transcriptomics profiling using barcoded microarrays. *Nat. Protoc.* 13 (11), 2425–2446. doi: 10.1038/s41596-018-0046-1

Giacomello, S., Salmen, F., Terebieniec, B. K., Vickovic, S., Navarro, J. F., Alexeyenko, A., et al. (2017). Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* 3, 17061. doi: 10.1038/nplants.2017.61

Giuliano, G., Pichersky, E., Malik, V. S., Timko, M. P., Scolnik, P. A., and Cashmore, A. R. (1988). An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proc. Natl. Acad. Sci. U.S.A.* 85 (19), 7089–7093. doi: 10.1073/pnas.85.19.7089

Graeff, M., Rana, S., Wendrich, J. R., Dorier, J., Eekhout, T., Aliaga Fandino, A. C., et al. (2021). A single-cell morpho-transcriptomic map of brassinosteroid action in the Arabidopsis root. *Mol. Plant* 14 (12), 1985–1999. doi: 10.1016/j.molp.2021.07.021

Grindberg, R. V., Yee-Greenbaum, J. L., McConnell, M. J., Novotny, M., O'Shaughnessy, A. L., Lambert, G. M., et al. (2013). RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. U.S.A.* 110 (49), 19802–19807. doi: 10.1073/pnas.1319700110

Guo, X., Liang, J., Lin, R., Zhang, L., Wu, J., and Wang, X. (2021). Series-Spatial Transcriptome Profiling of Leafy Head Reveals the Key Transition Leaves for Head Formation in Chinese Cabbage. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.787826

Gurazada, S. G. R., Cox, K. L., Czymmek, K. J., and Meyers, B. C. (2021). Space: the final frontier - achieving single-cell, spatially resolved transcriptomics in plants. *Emerg. Top. Life Sci.* 5 (2), 179–188. doi: 10.1042/ETLS20200274

Han, Y., Chu, X., Yu, H., Ma, Y.-K., Wang, X.-J., Qian, W., et al. (2017). Single-cell transcriptome analysis reveals widespread monoallelic gene expression in individual rice mesophyll cells. *Sci. Bull.* 62 (19), 1304–1314. doi: 10.1016/j.scib.2017.09.011

Hey, S., Baldauf, J., Opitz, N., Lithio, A., Pasha, A., Provart, N., et al. (2017). Complexity and specificity of the maize (Zea mays L.) root hair transcriptome. *J. Exp. Bot.* 68 (9), 2175–2185. doi: 10.1093/jxb/erx104

Hou, Z., Liu, Y., Zhang, M., Zhao, L., Jin, X., Liu, L., et al. (2021). High-throughput single-cell transcriptomics reveals the female germline differentiation trajectory in Arabidopsis thaliana. *Commun. Biol.* 4 (1), 1149. doi: 10.1038/s42003-021-02676-z

Hu, P., Liu, J., Zhao, J., Wilkins, B. J., Lupino, K., Wu, H., et al. (2018). Single-nucleus transcriptomic survey of cell diversity and functional maturation in postnatal mammalian hearts. *Genes Dev.* 32 (19-20), 1344–1357. doi: 10.1101/gad.316802.118

Hwang, B., Lee, J. H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 50 (8), 1–14. doi: 10.1038/s12276-018-0071-8

Jean-Baptiste, K., McFaline-Figueroa, J. L., Alexandre, C. M., Dorrity, M. W., Saunders, L., Bubb, K. L., et al. (2019). Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana. *Plant Cell* 31 (5), 993–1011. doi: 10.1105/tpc.18.00785

Jiang, C., Sun, J., Li, R., Yan, S., Chen, W., Guo, L., et al. (2022). A reactive oxygen species burst causes haploid induction in maize. *Mol. Plant* 15 (6), 943–955. doi: 10.1016/j.molp.2022.04.001

Keren-Shaul, H., Kenigsberg, E., Jaitin, D. A., David, E., Paul, F., Tanay, A., et al. (2019). MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nat. Protoc.* 14 (6), 1841–1862. doi: 10.1038/s41596-019-0164-4

Kim, J. Y., Symeonidi, E., Pang, T. Y., Denyer, T., Weidauer, D., Bezrutczyk, M., et al. (2021). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* 33 (3), 511–530. doi: 10.1093/plcell/koaa060

Kivivirta, K., Herbert, D., Lange, M., Beuerlein, K., Altmuller, J., and Becker, A. (2019). A protocol for laser microdissection (LMD) followed by transcriptome analysis of plant reproductive tissue in phylogenetically distant angiosperms. *Plant Methods* 15, 151. doi: 10.1186/s13007-019-0536-3

Klein, A. M., and Macosko, E. (2017). InDrops and Drop-seq technologies for single-cell sequencing. *Lab. Chip* 17 (15), 2540–2541. doi: 10.1039/c7lc90070h

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., et al. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161 (5), 1187–1201. doi: 10.1016/j.cell.2015.04.044

Krishnaswami, S. R., Grindberg, R. V., Novotny, M., Venepally, P., Lacar, B., Bhutani, K., et al. (2016). Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* 11 (3), 499–524. doi: 10.1038/nprot.2016.015

Lake, B. B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., et al. (2016). Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 352 (6293), 1586–1590. doi: 10.1126/science.aaf1204

Lareau, C. A., Ma, S., Duarte, F. M., and Buenrostro, J. D. (2020). Inference and effects of barcode multiplets in droplet-based single-cell assays. *Nat. Commun.* 11 (1), 866. doi: 10.1038/s41467-020-14667-5

Lhamo, D., and Luan, S. (2021). Potential Networks of Nitrogen-Phosphorus-Potassium Channels and Transporters in Arabidopsis Roots at a Single Cell Resolution. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.689545

Li, H., Wang, X., Wang, Y., Zhang, M., Hong, F., Wang, H., et al. (2022). Cross-species single-cell transcriptomic analysis reveals divergence of cell composition and functions in mammalian ileum epithelium. *Cell Regener.* 11 (1), 19. doi: 10.1186/s13619-022-00118-7

Li, X., Li, L., and Yan, J. (2015). Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. *Nat. Commun.* 6, 6648. doi: 10.1038/ncomms7648

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

Liu, C., Leng, J., Li, Y., Ge, T., Li, J., Chen, Y., et al. (2022a). A spatiotemporal atlas of organogenesis in the development of orchid flowers. *Nucleic Acids Res.* 50 (17), 9724–9737. doi: 10.1093/nar/gkac773

Liu, H., Hu, D., Du, P., Wang, L., Liang, X., Li, H., et al. (2021a). Single-cell RNA-seq describes the transcriptome landscape and identifies critical transcription factors in the leaf blade of the allotetraploid peanut (Arachis hypogaea L.). *Plant Biotechnol. J.* 19 (11), 2261–2276. doi: 10.1111/pbi.13656

Liu, Q., Li, P., Cheng, S., Zhao, Z., Liu, Y., Wei, Y., et al. (2022b). Protoplast dissociation and transcriptome analysis provides insights to salt stress response in cotton. *Int. J. Mol. Sci.* 23 (5), 2845. doi: 10.3390/ijms23052845

Liu, Q., Liang, Z., Feng, D., Jiang, S., Wang, Y., Du, Z., et al. (2021b). Transcriptional landscape of rice roots at the single-cell resolution. *Mol. Plant* 14 (3), 384–394. doi: 10.1016/j.molp.2020.12.014

Liu, W., Zhang, Y., Fang, X., Tran, S., Zhai, N., Yang, Z., et al. (2022c). Transcriptional landscapes of *de novo* root regeneration from detached Arabidopsis leaves revealed by time-lapse and single-cell RNA sequencing analyses. *Plant Commun.* 3 (4), 100306. doi: 10.1016/j.xplc.2022.100306

Liu, Z., Wang, J., Zhou, Y., Zhang, Y., Qin, A., Yu, X., et al. (2022d). Identification of novel regulators required for early development of vein pattern in the cotyledons by single-cell RNA-sequencing. *Plant J.* 110 (1), 7–22. doi: 10.1111/tpj.15719

Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., et al. (2020). Global Dynamic Molecular Profiling of Stomatal Lineage Cell Development by Single-Cell RNA Sequencing. *Mol. Plant* 13 (8), 1178–1193. doi: 10.1016/j.molp.2020.06.010

Long, Y., Liu, Z., Jia, J., Mo, W., Fang, L., Lu, D., et al. (2021). FlsnRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. *Genome Biol.* 22 (1), 66. doi: 10.1186/s13059-021-02288-0

Longo, S. K., Guo, M. G., Ji, A. L., and Khavari, P. A. (2021). Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.* 22 (10), 627–644. doi: 10.1038/s41576-021-00370-8

Lopez-Anido, C. B., Vaten, A., Smoot, N. K., Sharma, N., Guo, V., Gong, Y., et al. (2021). Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. *Dev. Cell* 56 (7), 1043–1055.e1044. doi: 10.1016/j.devcel.2021.03.014

Lu, L., Tian, S., Liao, H., Zhang, J., Yang, X., Labavitch, J. M., et al. (2013). Analysis of metal element distributions in rice (Oryza sativa L.) seeds and relocation during germination based on X-ray fluorescence imaging of Zn, Fe, K, Ca, and Mn. *PloS One* 8 (2), e57360. doi: 10.1371/journal.pone.0057360

Lu, L., Xie, R., Liu, T., Wang, H., Hou, D., Du, Y., et al. (2017). Spatial imaging and speciation of Cu in rice (Oryza sativa L.) roots using synchrotron-based X-ray microfluorescence and X-ray absorption spectroscopy. *Chemosphere* 175, 356–364. doi: 10.1016/j.chemosphere.2017.02.082

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161 (5), 1202–1214. doi: 10.1016/j.cell.2015.05.002

Maitra, M., Nagy, C., Chawla, A., Wang, Y. C., Nascimento, C., Suderman, M., et al. (2021). Extraction of nuclei from archived postmortem tissues for single-nucleus sequencing applications. *Nat. Protoc.* 16 (6), 2788–2801. doi: 10.1038/s41596-021-00514-4

Marand, A. P., Chen, Z., Gallavotti, A., and Schmitz, R. J. (2021). A cis-regulatory atlas in maize at single-cell resolution. *Cell* 184 (11), 3041–3055.e3021. doi: 10.1016/j.cell.2021.04.014

Misra, C. S., Santos, M. R., Rafael-Fernandes, M., Martins, N. P., Monteiro, M., and Becker, J. D. (2019). Transcriptomics of Arabidopsis sperm cells at single-cell resolution. *Plant Reprod.* 32 (1), 29–38. doi: 10.1007/s00497-018-00355-4

Mo, Y., and Jiao, Y. (2022). Advances and applications of single-cell omics technologies in plant research. *Plant J.* 110 (6), 1551–1563. doi: 10.1111/tpj.15772

Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., et al. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362 (6416), eaau5324. doi: 10.1126/science.aau5324

Moreno-Romero, J., Santos-Gonzalez, J., Hennig, L., and Kohler, C. (2017). Applying the INTACT method to purify endosperm nuclei and to generate parental-specific epigenome profiles. *Nat. Protoc.* 12 (2), 238–254. doi: 10.1038/nprot.2016.167

Moreno-Villena, J. J., Zhou, H., Gilman, I. S., Tausta, S. L., Cheung, C. Y. M., and Edwards, E. J. (2022). Spatial resolution of an integrated C4+CAM photosynthetic metabolism. *Sci. Adv.* 8 (31), eabn2349. doi: 10.1126/sciadv.abn2349

Nadelmann, E. R., Gorham, J. M., Reichart, D., Delaughter, D. M., Wakimoto, H., Lindberg, E. L., et al. (2021). Isolation of nuclei from mammalian cells and tissues for single-nucleus molecular profiling. *Curr. Protoc.* 1 (5), e132. doi: 10.1002/cpz1.132

Nelms, B., and Walbot, V. (2019). Defining the developmental program leading to meiosis in maize. *Science* 364 (6435), 52–56. doi: 10.1126/science.aav6428

Neumann, M., Xu, X., Smaczniak, C., Schumacher, J., Yan, W., Bluthgen, N., et al. (2022). A 3D gene expression atlas of the floral meristem based on spatial reconstruction of single nucleus RNA sequencing data. *Nat. Commun.* 13 (1), 2838. doi: 10.1038/s41467-022-30177-y

Ortiz-Ramirez, C., Arevalo, E. D., Xu, X., Jackson, D. P., and Birnbaum, K. D. (2018). An Efficient Cell Sorting Protocol for Maize Protoplasts. *Curr. Protoc. Plant Biol.* 3 (3), e20072. doi: 10.1002/cppb.20072

Potter, S. S. (2018). Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.* 14 (8), 479–492. doi: 10.1038/s41581-018-0021-7

Qian, X., Harris, K. D., Hauling, T., Nicoloutsopoulos, D., Munoz-Manchado, A. B., Skene, N., et al. (2020). Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* 17 (1), 101–106. doi: 10.1038/s41592-019-0631-4

Ramsköld, D., Luo, S., Wang, Y.-C., Li, R., Deng, Q., Faridani, O. R., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30 (8), 777–782. doi: 10.1038/nbt.2282

Rao, A., Barkley, D., Franca, G. S., and Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. *Nature* 596 (7871), 211–220. doi: 10.1038/s41586-021-03634-9

Roszak, P., Heo, J. O., Blob, B., Toyokura, K., Sugiyama, Y., de Luis Balaguer, M. A., et al. (2021). Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization. *Science* 374 (6575), eaba5531. doi: 10.1126/science.aba5531

Ryu, K. H., Huang, L., Kang, H. M., and Schiefelbein, J. (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol.* 179 (4), 1444–1456. doi: 10.1104/pp.18.01482

Salguero-Linares, J., Serrano, I., Ruiz-Solani, N., Salas-Gomez, M., Phukan, U. J., Gonzalez, V. M., et al. (2022). Robust transcriptional indicators of immune cell death revealed by spatiotemporal transcriptome analyses. *Mol. Plant* 15 (6), 1059–1075. doi: 10.1016/j.molp.2022.04.010

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33 (5), 495–502. doi: 10.1038/nbt.3192

Satterlee, J. W., Strable, J., and Scanlon, M. J. (2020). Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. U.S.A.* 117 (52), 33689–33699. doi: 10.1073/pnas.2018788117

See, P., Lum, J., Chen, J., and Ginhoux, F. (2018). A Single-Cell Sequencing Guide for Immunologists. *Front. Immunol.* 9. doi: 10.3389/fimmu.2018.02425

Shahan, R., Hsu, C. W., Nolan, T. M., Cole, B. J., Taylor, I. W., Greenstreet, L., et al. (2022). A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev. Cell* 57 (4), 543–560.e549. doi: 10.1016/j.devcel.2022.01.008

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498 (7453), 236–240. doi: 10.1038/nature12172

Shaw, R., Tian, X., and Xu, J. (2021). Single-cell transcriptome analysis in plants: Advances and challenges. *Mol. Plant* 14 (1), 115–126. doi: 10.1016/j.molp.2020.10.012

Shojaee, A., Saavedra, M., and Huang, S. C. (2021). Potentials of single-cell genomics in deciphering cellular phenotypes. *Curr. Opin. Plant Biol.* 63, 102059. doi: 10.1016/j.pbi.2021.102059

Shulse, C. N., Cole, B. J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., et al. (2019). High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* 27 (7), 2241–2247.e2244. doi: 10.1016/j.celrep.2019.04.054

Sikorskaite, S., Rajamaki, M. L., Baniulis, D., Stanys, V., and Valkonen, J. P. (2013). Protocol: Optimised methodology for isolation of nuclei from leaves of species in the Solanaceae and Rosaceae families. *Plant Methods* 9, 31. doi: 10.1186/1746-4811-9-31

Somssich, M., Khan, G. A., and Persson, S. (2016). Cell wall heterogeneity in root development of arabidopsis. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01242

Song, Q., Ando, A., Jiang, N., Ikeda, Y., and Chen, Z. J. (2020). Single-cell RNA-seq analysis reveals ploidy-dependent and cell-specific transcriptome changes in Arabidopsis female gametophytes. *Genome Biol.* 21 (1), 178. doi: 10.1186/s13059-020-02094-0

Stahl, P. L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353 (6294), 78–82. doi: 10.1126/science.aaf2403

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M.3rd, et al. (2019). Comprehensive integration of single-cell data. *Cell* 177 (7), 1888–1902.e1821. doi: 10.1016/j.cell.2019.05.031

Sunaga-Franze, D. Y., Muino, J. M., Braeuning, C., Xu, X., Zong, M., Smaczniak, C., et al. (2021). Single-nucleus RNA sequencing of plant tissues using a nanowell-based system. *Plant J.* 108 (3), 859–869. doi: 10.1111/tpj.15458

Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* 13 (4), 599–604. doi: 10.1038/nprot.2017.149

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6 (5), 377–382. doi: 10.1038/nmeth.1315

Tang, Z., Fan, Y., Zhang, L., Zheng, C., Chen, A., Sun, Y., et al. (2021). Quantitative metabolome and transcriptome analysis reveals complex regulatory pathway underlying photoinduced fiber color formation in cotton. *Gene* 767, 145180. doi: 10.1016/j.gene.2020.145180

Tenorio Berrio, R., Verstaen, K., Vandamme, N., Pevernagie, J., Achon, I., Van Duyse, J., et al. (2022). Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. *Plant Physiol.* 188 (2), 898–918. doi: 10.1093/plphys/kiab489

Thibivilliers, S., Anderson, D., and Libault, M. (2020). Isolation of Plant Root Nuclei for Single Cell RNA Sequencing. *Curr. Protoc. Plant Biol.* 5 (4), e20120. doi: 10.1002/cppb.20120

Tirichine, L., Andrey, P., Biot, E., Maurin, Y., and Gaudin, V. (2009). 3D fluorescent *in situ* hybridization using Arabidopsis leaf cryosections and isolated nuclei. *Plant Methods* 5, 11. doi: 10.1186/1746-4811-5-11

Torii, K., Kubota, A., Araki, T., and Endo, M. (2020). Time-Series Single-Cell RNA-Seq Data Reveal Auxin Fluctuation during Endocycle. *Plant Cell Physiol.* 61 (2), 243–254. doi: 10.1093/pcp/pcz228

van den Brink, S. C., Sage, F., Vertesy, A., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., et al. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* 14 (10), 935–936. doi: 10.1038/nmeth.4437

Wang, J., Li, Y., Wu, T., Miao, C., Xie, M., Ding, B., et al. (2021a). Single-cell-type transcriptomic analysis reveals distinct gene expression profiles in wheat guard cells in response to abscisic acid. *Funct. Plant Biol.* 48 (11), 1087–1099. doi: 10.1071/FP20368

Wang, Q., Wu, Y., Peng, A., Cui, J., Zhao, M., Pan, Y., et al. (2022). Single-cell transcriptome atlas reveals developmental trajectories and a novel metabolic pathway of catechin esters in tea leaves. *Plant Biotechnol. J.* 20, 2089–2106. doi: 10.1111/pbi.13891

Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361 (6400), eaat5691. doi: 10.1126/science.aat5691

Wang, X., He, Y., Zhang, Q., Ren, X., and Zhang, Z. (2021b). Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. *Genomics Proteomics Bioinf.* 19 (2), 253–266. doi: 10.1016/j.gpb.2020.02.005

Wang, Y., Huan, Q., Li, K., and Qian, W. (2021c). Single-cell transcriptome atlas of the leaf and root of rice seedlings. *J. Genet. Genomics* 48 (10), 881–898. doi: 10.1016/j.jgg.2021.06.001

Wei, R., He, S., Bai, S., Sei, E., Hu, M., Thompson, A., et al. (2022). Spatial charting of single-cell transcriptomes in tissues. *Nat. Biotechnol.* 40 (8), 1190–1199. doi: 10.1038/s41587-022-01233-1

Wendrich, J. R., Yang, B., Vandamme, N., Verstaen, K., Smet, W., Van de Velde, C., et al. (2020). Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* 370 (6518), eaay4970. doi: 10.1126/science.aay4970

Wu, H., Kirita, Y., Donnelly, E. L., and Humphreys, B. D. (2019). Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J. Am. Soc. Nephrol.* 30 (1), 23–32. doi: 10.1681/ASN.2018090912

Xia, K., Sun, H. X., Li, J., Li, J., Zhao, Y., Chen, L., et al. (2022). The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in Arabidopsis leaves. *Dev. Cell* 57 (10), 1299–1310.e1294. doi: 10.1016/j.devcel.2022.04.011

Xie, Y., Jiang, S., Li, L., Yu, X., Wang, Y., Luo, C., et al. (2020). Single-Cell RNA Sequencing Efficiently Predicts Transcription Factor Targets in Plants. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.603302

Xu, X., Crow, M., Rice, B. R., Li, F., Harris, B., Liu, L., et al. (2021a). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev. Cell* 56 (4), 557–568.e556. doi: 10.1016/j.devcel.2020.12.015

Xu, X. F., Zhu, H. Y., Ren, Y. F., Feng, C., Ye, Z. H., Cai, H. M., et al. (2021b). Efficient isolation and purification of tissue-specific protoplasts from tea plants (Camellia sinensis (L.) O. Kuntze). *Plant Methods* 17 (1), 84. doi: 10.1186/s13007-021-00783-w

Yang, B., Minne, M., Brunoni, F., Plackova, L., Petrik, I., Sun, Y., et al. (2021). Non-cell autonomous and spatiotemporal signalling from a tissue organizer orchestrates root vascular development. *Nat. Plants* 7 (11), 1485–1494. doi: 10.1038/s41477-021-01017-6

Yoo, S. D., Cho, Y. H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2 (7), 1565–1572. doi: 10.1038/nprot.2007.199

Zhang, H., Yang, H., Hu, D., Li, B., Lin, Y., Yao, W., et al. (2021a). Single-cell RNA sequencing of meiocytes and microspores reveals the involvement of the Rf4 gene in redox homeostasis of CMS-C maize. *Crop J.* 9 (6), 1237–1247. doi: 10.1016/j.cj.2021.06.012

Zhang, M., Zou, Y., Xu, X., Zhang, X., Gao, M., Song, J., et al. (2020). Highly parallel and efficient single cell mRNA sequencing with paired picoliter chambers. *Nat. Commun.* 11 (1), 2118. doi: 10.1038/s41467-020-15765-0

Zhang, S., Xu, M., Qiu, Z., Wang, K., Du, Y., Gu, L., et al. (2016). Spatiotemporal transcriptome provides insights into early fruit development of tomato (Solanum lycopersicum). *Sci. Rep.* 6, 23173. doi: 10.1038/srep23173

Zhang, T. Q., Chen, Y., Liu, Y., Lin, W. H., and Wang, J. W. (2021b). Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat. Commun.* 12 (1), 2053. doi: 10.1038/s41467-021-22352-4

Zhang, T. Q., Chen, Y., and Wang, J. W. (2021c). A single-cell analysis of the Arabidopsis vegetative shoot apex. *Dev. Cell* 56 (7), 1056–1074.e1058. doi: 10.1016/j.devcel.2021.02.021

Zhang, T. Q., Xu, Z. G., Shang, G. D., and Wang, J. W. (2019). A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root. *Mol. Plant* 12 (5), 648–660. doi: 10.1016/j.molp.2019.04.004

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049. doi: 10.1038/ncomms14049

Zong, J., Wang, L., Zhu, L., Bian, L., Zhang, B., Chen, X., et al. (2022). A rice single cell transcriptomic atlas defines the developmental trajectories of rice floret and inflorescence meristems. *New Phytol.* 234 (2), 494–512. doi: 10.1111/nph.18008

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 August 2023
DOI 10.3389/fpls.2023.1233191



OPEN ACCESS

EDITED BY
Milen I. Georgiev,
Bulgarian Academy of Sciences, Bulgaria

REVIEWED BY
Mohsen Hesami,
University of Guelph, Canada
Miroslawa Chwil,
University of Life Sciences of Lublin, Poland

*CORRESPONDENCE
Hocelayne Paulino Fernandes
h.paulino.fernandes@biology.leidenuniv.nl
Francesco Tonolo
francesco.tonolo@protonmail.com

RECEIVED 02 June 2023
ACCEPTED 17 July 2023
PUBLISHED 11 August 2023

CITATION
Fernandes HP, Choi YH, Vrieling K, de Bresser M, Sewalt B and Tonolo F (2023) Cultivar-dependent phenotypic and chemotypic responses of drug-type *Cannabis sativa* L. to polyploidization. *Front. Plant Sci.* 14:1233191. doi: 10.3389/fpls.2023.1233191



# Cultivar-dependent phenotypic and chemotypic responses of drug-type *Cannabis sativa* L. to polyploidization

Hocelayne Paulino Fernandes[1,2]*, Young Hae Choi[2], Klaas Vrieling[1], Maikel de Bresser[3], Bobbie Sewalt[1] and Francesco Tonolo[1]*

[1]Aboveground-belowground Interaction Group, Plant Cluster, Institute of Biology, Leiden University, Leiden, Netherlands, [2]Natural Products Laboratory, Institute of Biology, Leiden University, Leiden, Netherlands, [3]F1 SeedTech, Eindhoven, Netherlands

*Cannabis sativa* L. is a plant with a wide range of potential medicinal applications. In recent years, polyploidy has gained attention as a potential strategy for rapidly improving *C. sativa*, which, unlike other modern crops, has not yet benefitted from this established biotechnological application. Currently, no reports on high THCA and CBDA drug-type polyploid cultivars have been published. Moreover, it still needs to be clarified if different cultivars react similarly to polyploidization. For these reasons, we set out to evaluate and compare the phenotype and chemotype of three high $\Delta^9$-tetrahydrocannabinolic acid (THCA) and one high cannabidiolic acid (CBDA) drug-type cultivars in their diploid, triploid and tetraploid state through agronomic and metabolomic approaches. Our observations on plant morphology revealed a significant increase in plant height and leaf size with increasing ploidy levels in a cultivar-dependent manner. In contrast, cannabinoids were negatively affected by polyploidization, with the concentration of total cannabinoids, THCA, CBDA and cannabigerolic acid (CBGA) decreasing significantly in higher ploidy levels across all four cultivars. Headspace analysis of volatiles revealed that total volatile content decreased in triploids. On the other hand, tetraploids reacted differently depending on the cultivars. Two THCA dominant cultivars showed an increase in concentrations, while in the other two cultivars, concentrations decreased. Additionally, several rare compounds not present in diploids appeared in higher ploidy levels. Moreover, in one high THCA cultivar, a couple of elite tetraploid genotypes for cannabinoid and volatile production were identified, highlighting the role of cultivar and genotypic variability as an important factor in *Cannabis sativa* L. polyploids. Overall, our observations on plant morphology align with the giga phenotype observed in polyploids of other plant species. The decrease in cannabinoids and volatiles production in triploids have relevant implications regarding their commercial use. On the other hand, this study found that tetraploidization is a suitable approach to improve *Cannabis sativa* L. medicinal potential, although the response is cultivar and genotype-dependent. This work lays

the ground for further improving, evaluating and harnessing *Cannabis sativa* L. chemical diversity by the breeding, biotechnological and pharmaceutical sectors.

KEYWORDS

***Cannabis sativa* L., polyploidy, cannabinoids, chemical diversity, terpenes, metabolomics, biotechnology, medicinal & aromatic plants**

## Introduction

Polyploidization is a widespread phenomenon in the plant kingdom and it plays a significant role in plant evolution and speciation. Indeed, genomic doubling provides additional genetic material upon which natural selection can act. At the same time, it establishes reproductive barriers, which can limit the capacity of polyploid organisms to mate with their diploid ancestors (Wood et al., 2009; Wendel et al., 2018). Polyploidization can lead to a multitude of effects on the plants' phenology, morphology, physiology and metabolic processes. One of the most common effects of polyploidization is the "giga" effect. Cell and nucleus volume increases compared to diploid ancestors, with cell areas being 1.5 times larger (Lavania, 2013). This can lead to an increase in organ size and, ultimately, biomass (Urwin, 2014). Other frequently observed effects in different species are changes in size and number of stomata, shoots, roots, leaves, flowers, seeds and plant height (Trojak-Goluch et al., 2021). Because of these effects, there are many examples of plant species that have improved agronomic qualities due to polyploidization. However, polyploidy does not necessarily work for the better, as documented by Kaensaksiri et al. (2011) and Xu et al. (2014), who found that for *Centella asiatica* L. and *Echinacea purpurea* L. the number of organs was reduced, and biomass did not increase. Nevertheless, polyploidization has been a powerful breeding tool for crop improvement (Sattler et al., 2016). Simmonds (1980) estimated that approximately 40% of the cultivated species are polyploids, and today, common crops such as *Vitis* spp. L. (Motosugi et al., 2002), *Actinidia chinensis*. (Hopping, 1994), *Solanum tuberosum* L. (Carputo et al., 2003) *Prunus domestica* L. (Bennett and Leitch, 2005), and *Fragaria* (Edger et al., 2020), among many others, are polyploids.

Polyploidization has also been shown to affect secondary metabolite production in several medicinal and aromatic plants (MAPs) (Iannicelli et al., 2020), including plant species of the Cannabaceae family such as *Humulus lupulus* L. (Trojak-Goluch and Skomra, 2013). Duplication of the genome can induce a series of changes, the most obvious being the increase in gene copy number. Genome duplications can also affect the genomic organization, gene regulation, transposon activation and can induce epigenetic reprogramming by modifying histones and chromatin structures (Song and Chen, 2015). The final outcome is a change in gene expression which can directly influence the metabolome, either through increased, decreased, or silenced transcription (Madani et al., 2021).

The general trend in polyploid MAPs is an increase in secondary metabolite production (Iannicelli et al., 2020), with some exceptions (Silvarolla et al., 1999; Lavania, 2005; Wohlmuth et al., 2005; Tan et al., 2017). Various studies have reported quantitative differences of diverse secondary metabolites in synthetic autotetraploids, which could be attributed to changes in gene expression. For example, genes involved in the artemisinin biosynthetic pathway were upregulated, resulting in an increased production of artemisinin in *Artemisia annua* plants (Lin et al., 2011), and a similar change was reported for morphine and related biosynthetic genes in *Papaver somniferum* L. (Mishra et al., 2010).

In light of these findings, *Cannabis sativa* L. (Cannabaceae) appears to be an interesting candidate for polyploidization as *C. sativa* has been widely cultivated due to its industrial (Karche and Singh, 2019), ornamental (Hesami et al., 2022), nutritional (Krüger et al., 2022), and broad medicinal potentials (Andre et al., 2016) and *C. sativa* is a prolific producer of secondary metabolites with at least 348 well-characterized compounds (Hanuš et al., 2016; Pollastro et al., 2018; Radwan et al., 2021) classified as cannabinoids (150), terpenoids (120), phenolics (42), flavonoids (34) and alkaloids (2).

For these reasons, lately, polyploidization has been at the center of attention in the *C. sativa* industry. In the past, due to legal restrictions, the generation and testing of *C. sativa* polyploids posed considerable challenges as permits for handling *C. sativa* material and hazardous chemicals as well as access to specialized equipment and technical knowledge were required. These hurdles delayed basic research needed to elucidate the changes brought about by polyploidization of *C. sativa* and already documented in other plant species.

To date, there are only a handful of studies assessing the polyploid state of *C. sativa*. Mansouri and Bagheri (2017) were the first to compare tetraploid to diploid *C. sativa*. They observed that the tetraploids had more total proteins, total flavonoids and starch while having reduced cellulose content compared to diploid plants. Unfortunately, samples were not decarboxylated before cannabinoid analysis and only $\Delta^9$-tetrahydrocannabinol (THC) and Cannabidiol (CBD) were measured therefore missing THCA and CBDA. Mansouri and Bagheri (2017) expanded their measurements in a second experiment. They found significant differences in leaf morphology, with tetraploids having shorter and broader leaves and a decrease in glandular trichome density, fiber content, height, and soluble sugars. At the same time, the amount of total proteins doubled compared to diploids. However, cannabinoids were measured as in their first publication, and no specifications were made regarding the starting plant material. Parson et al. (2019) made

observations on a single balanced CBDA/THCA drug-type tetraploid genotype and its diploid background. Measurements were conducted on ten replicates of a clone and found an 8% increase in CBDA but no increase in THCA, total cannabinoids and terpenes, while limonene decreased significantly compared to diploids. Like Mansouri and Bagheri (2017), they also observed that the central leaflet width was significantly bigger in tetraploids. Kurtz et al. (2020) published a safe and efficient method using pregerminated seeds and colchicine for developing tetraploids in *C. sativa*. They successfully generated tetraploids of 5 drug-type CBDA dominant cultivars and were able to generate triploids via embryo rescue. However, they only presented results on stomata for which they observed an increased size and reduced density with increasing ploidy levels. Lastly, Crawford et al. (2021) developed triploids and tetraploids of a CBGA dominant drug-type cultivar. Although not significant ($p<0.1$), they observed an increase in total cannabinoid concentrations, being 8.66%, 10.18%, and 12.38% of the dry flower weight, for diploid, triploid and tetraploid, respectively. Additionally, they also observed an increase in flower weight in triploids of 23% and 26% compared to diploids in outdoor and indoor trials, respectively.

*C. sativa* is a highly variable crop (Clarke and Merlin, 2016; Small, 2018) and Parson et al. (2019), as well as Kurtz et al. (2020), showed that the response of *C. sativa* to polyploidization induction is genotype and cultivar dependent. Additionally, observations of Parsons et al. (2019) and Crawford et al. (2021) on cannabinoid concentrations hinted that polyploidization might affect secondary metabolite quantity differently, likely dependent on the starting plant material. However, up to now, no tests on THCA and CBDA dominant drug-type varieties have been published, and it still needs to be clarified how different cultivars are affected by polyploidization. Therefore, we use a targeted metabolomics approach to investigate changes in cannabinoids and volatile content in triploids and tetraploids of 4 drug-type cultivars compared to their diploid ancestors. The cultivars used in the study, to fill the gaps in the field, were THCA and CBDA dominant and varied in secondary metabolite composition. Additionally, as leaf morphology has been proposed as proxy for polyploidy identification, we also measured leaf characteristics to check if previous observations apply to different cultivars. Here we set out to answer the following research questions: 1) Does ploidy level affect leaf morphology, cannabinoid and volatile quantity and diversity in *C. sativa*? 2) Is the response to polyploidization concerning leaf morphology, cannabinoid and volatile quantity and diversity cultivar-dependent in *C. sativa*?

# Material and methods

## Plant material

*C. sativa* seeds of four feminized commercial drug-type cultivars, in their respective diploid, triploid and tetraploid states, were obtained from F1 SeedTech. A pilot experiment was conducted, and among 29 cultivars tested, four cultivars were selected for their productivity, relatively low phenotypic variability, differential genetic background and chemical composition to evaluate their polyploid status. Cultivars used in the experiment were THCA dominant (cultivars A, B and C) and CBDA dominant (cultivar D).

## Plant growth

Plant growth was carried out following standard agricultural practices for the production of *C. sativa* inflorescences, as previously described by Magagnini et al. (2018). To minimize possible variation in the treatment of individual plants, the same nutrient solution concentration and regime of fertigation was used for all plants. Additionally, plants were placed in the growth chamber following a completely randomized design. Plants were grown in an indoor growth chamber fitted with 12, 400w HPS horticultural lights (Osram, Munich, Germany), providing a light intensity of 200 ± 50 µmol/m²/s PAR at canopy level. Environmental parameters were kept constant with a temperature of 25°C during the day and 21°C during the night, while RH was set to 70%. Diploid, triploid and tetraploid seeds were sterilized with 0.05% NaClO solution for 3 minutes and were subsequently rinsed three times with sterile deionized water and placed in Petri dishes lined with a moist paper filter. After 72h germinated seeds were then transferred to Rockwool plugs (Rockwool, Roermond, NL) moistened with hydroponic solution (Dutchpro Nutrients, Almere, NL). The plant nutrient solution was prepared by adding A and B of the vegetative hydroponic stock solution (Dutchpro) to deionized water to reach an electric conductivity (EC) of 0.4 mS/cm, and then pH was adjusted to 5.8 with NaOH. After a week from germination, plantlets were transplanted into 1l Rockwool cubes (Grodan) fitted with an automated irrigation system. During the vegetative phase, plants were fertigated three times a day with circa 150 ml vegetative hydroponic solution A and B (Dutchpro), supplemented with $MgSO_4$ and $CaCl_2$ to a final concentration of 0.127g/l and 0.05g/l, respectively. The EC was increased gradually as the plants grew to reach a value of 1.0 mS/cm while the surplus solution released by the Rockwool cubes was drained.

After three weeks from germination, flowering was induced by a change of photoperiod to short daylight conditions (12h/12h light/darkness). Plants then received the flowering solution A and B (Dutchpro), supplemented with $MgSO_4$ and $CaCl_2$, as stated above, until harvest. The EC value was increased to 1.6 mS/cm as plants entered the reproductive phase between week 7 to week 12 from germination. The nutrient solution concentration was then gradually lowered to EC 0.3 mS/cm during the last 3 weeks of flowering to promote maturation and senescence. Once all stigmas turned brown, flowers ceased to grow, and leaves progressed into senescence, plants were deemed mature. At week 15 (total of 12 weeks of flowering), flowers were harvested and dried in darkness in a climate chamber at 20°C and 30% RH for 7 days. For the number of *C. sativa* plants grown per cultivar and ploidy level, see Table 1. Samples for each plant were taken in equal proportion from the

TABLE 1 Number of *Cannabis sativa* L. plants grown for the experiment.

| Cultivar | Diploid | Triploid | Tetraploid |
|---|---|---|---|
| A | 10 | 4 | 5 |
| B | 10 | 7 | 4 |
| C | 9 | 5 | 12 |
| D | 9 | 5 | 9 |

apical, middle and lower flowers and pooled to account for possible variability in secondary metabolite accumulation throughout the plant height. As a result, a representative 1g of dried flower material per plant was obtained. Samples were stored at -20°C in sealed 10ml Falcon tubes, and before chemical analysis, samples were grounded with a mortar and pestle until a fine powder was obtained.

## Ploidy analysis

The method for ploidy analysis was carried out following a modified protocol by Zonneveld and Van Iren (2001). Nuclei isolation buffer containing 4.15g/l MOPS, 9.15 g/l $MgCl_2$, 8.8 g/l TriSodium citrate, 1.55 g/l DTT, 25g/l Polyvinylpyrrolidone and 1ml/l Triton-x was prepared before the analysis and stored at 4°C. Subsequently, 100-200 mg of fresh plant material was collected from newly emerged leaves and kept on ice until analysis. The samples were prepared by chopping the leaves with a razor blade in a Petri dish containing 1 ml of buffer. The buffer containing the free nuclei was then filtered through a 20 µm Minisart nylon filter (Sartorius, Goettingen, Germany) into a 1.5 ml Eppendorf and kept on ice while the other samples were processed. After all samples were prepared, Propidium iodide was added to a final concentration of 50 µg/ml. The samples were mixed and incubated in darkness on ice for 10 minutes prior to analysis. Ploidy analysis on a flow cytometer (Milliliter Guava Easycyte, Merck, Darmstadt, Germany) was performed following analytical parameters as stated by Parson et al. (2019) and Kurtz et al. (2020). The laser at 488nm was used for excitation, and fluorescence was recorded with Yellow H linear channel at 583/26 nm. Diploid *C. sativa* plants were used as standards. Samples were measured in triplicate and all plants part of the experiment were tested twice throughout their development (during the vegetative and flowering phase) to ensure ploidy stability, including control diploids.

## Morphological measurements

Leaf and plant height measurements were carried out two weeks before the final harvest. All plants were measured on the same day. For the leaflet measurements, the central leaflet of the leaf attached to the third node counting from the apical meristem of the plant was used. The length of the central leaflet and the width at the minor axis of the leaflet ellipse was measured. Plant height was measured from the base of the stalk to the apical meristem.

## Cannabinoid analysis

Cannabinoids were extracted with 1ml of methanol from 10 mg of homogenized flower material. Sonication was carried out for 20 minutes. Subsequently, the tube was centrifuged at 13000g to pellet plant material. The supernatant was filtered through a 20 µm RC Minisart filter (Sartorius). The extract was kept in a sealed dark glass vial and stored at -20°C until analysis. Cannabinoids were quantified following a modified protocol by Gul et al. (2015) with a reversed-phase HPLC (Agilent 1200 chromatographic system, Agilent, Folsom, CA, USA) equipped with a UV-photodiode array detector (UV-DAD) and an auto-sampler. The separation was achieved on a Luna-C18 column (Phenomenex, Utrecht, NL). The mobile phase consisted of solvent A 0.1% (v/v) formic acid in water and solvent B 0.1% (v/v) formic acid in acetonitrile. The flow rate was set at 1.2 ml/minute. Solvent B was initially set at 30%, and after 3 minutes, it was increased to 75%. At 15 minutes from the start of the program, solvent B was increased to 100% and held for 3 minutes, after which it was decreased to 30% in the final two minutes of the sample run. The auto-sampler automatically injected 5µl of the sample. Light absorption was detected by the DAD at 228, 230, 280, 320 and 360 nm. The wavelength chosen for quantification was 320 nm. Quantification was performed with standards, THCA, CBDA, CBGA, THC, CBD and Cannabinol (CBN) (Merck). A five-fold serial dilution of the standards was performed on the range of quantification, and linear regression in RStudio software (version 4.3.0) was performed. $R^2$ for all calibration regressions were above 0.99.

## Headspace analysis

Thirty milligrams of dried plant material were placed into a 20 mL headspace vial and sealed. The samples were incubated at 150°C for 30 minutes at 500 rpm before analyses. Before capping the vial, an aliquot of 15 µL of methyl palmitate (5mg/mL), used as internal standard, was added.

The GC-MS analysis was performed following a modified protocol (Shimadzu Scientific Instruments, 2016) using a 7890A gas chromatograph equipped with a 7693 automatic sampler and a 5975C single-quadrupole mass detector (Agilent). Compounds were separated on a DB-5 column (30 m x 0.25 mm, 0.25 µm film, J&W Science, Folsom, CS, USA). Helium (99.9% purity) was used as a carrier gas at a flow rate of 1.6 mL/min. The oven temperature was held at 80°C for 1 minute, then increased at 3°C/min to 150°C, and then increased to 250°C at 7°C/min and held for 3 min. The samples were injected with a 20:1 split and split flow of 8 mL/min. The injection volume was 500 µL. The ionization energy in EI mode was 70 eV, and a mass scan range of 50–550 amu. Data was

processed using MassHunter (B.07, Agilent). Peak identification was done by comparing the ion spectra obtained from samples to ion spectra in the NIST V.2008 library.

## Statistical analysis

Two-way ANOVAs were performed with cultivar and ploidy levels as fixed factors and plant height, leaflet length, leaflet width and cannabinoid levels as dependent variables. When the normality of residuals and homoscedasticity assumptions were not met, data were Box-Cox transformed. Differences between all possible pairs of means were obtained by performing a Tukey's Honest Significance Difference (HSD) test or a Dunnett *post hoc* test. Boxplots were made to visualize differences between treatments and/or cultivars.

The GC headspace chromatograms' total peak areas (TIC) of compounds relative to that of the internal standard (methyl palmitate) were used for multivariate analysis. SIMCA-P (16.0) software (Umetrics, Umeå, Sweden) was used to perform multivariate analysis, applying univariate scaling to the data. Loading plots, as well as VIP, were set up to visualize results and predict possible biomarkers. The unsupervised method PCA, not requiring any *a priori* knowledge, was applied first to explore the dataset. Subsequently, a supervised method OPLS-DA was applied to identify differential components among the samples. The VIP (Variable Influence on Projection) is a parameter used for calculating the cumulative measure of the influence of individual X-variables on the model. VIP values larger than 1, point to the most relevant variables, and generally, VIP values below 0.5 are considered irrelevant variables (Galindo-Prieto et al., 2013). Thus, in the present analysis, we considered only VIP values > 1. A two-way ANOVA was performed with cultivar and ploidy level as fixed factors and total peak area as dependent variable. For the number of peaks and peak area of 33 compounds with the highest VIP score, a non-parametric Kruskal-Wallis test with cultivar and ploidy as factors was performed. Following a Kruskal-Wallis test per each individual cultivar with ploidy as factor was performed. *Post hoc* analysis to detect differences between all possible pairs of means was performed with a non-parametric Dunn test. The software that was used for these analyses is R version 4.3.0. Boxplots were made to visualize differences between means using R version 4.3.0.

# Results

## Morphological measurements

### Plant height

Measurements of plant height, as expected, revealed significant differences between cultivars. Indeed, cultivar C did have a significantly larger height than cultivars A, B and D (Figure 1A). Also, ploidy level did affect plant height significantly, with tetraploid *C. sativa* plants being significantly taller (71.38 ± 15.42 cm) than diploids (61.54 ± 12.99 cm). At the same time, triploids were found in-between (65.3 ± 20.77 cm). Further analysis of separate cultivars revealed a differential response to polyploidization, with cultivar A showing a decreasing trend with increasing ploidy levels. On the other hand, triploids of cultivars C and D increased compared to diploids, with D being significantly taller than its diploid ancestors. In contrast, in cultivar B a slightly significant (0.070) difference was found between the triploids, which decreased and the tetraploids, which increased compared to diploids (Figure 1A).

### Leaf morphological traits

The middle leaflet length measurements also revealed significant differences between cultivars. Cultivars A and B had the shortest leaflet length, cultivar C had the longest leaflet length, while cultivar D was found in between. As expected, ploidy level also affected the leaflet length. Regardless of the cultivar, an increase was observed with increasing ploidy levels, with diploids (11.6 ± 2.0 cm) showing a significantly shorter leaflet length compared to triploids (13.04 ± 3.0 cm) and tetraploids (13.21 ± 2.6 cm) (Figure 1B). Both cultivar and ploidy had a significant effect on middle leaflet width. *C. sativa* cultivars B and C had significantly broader middle leaflets than cultivars A and D (Figure 1C). Regardless of the cultivars, the middle leaflet width increased with ploidy level, following a similar pattern observed with leaflet length. Diploids (3.17 ± 0.4 cm) showed a significantly shorter leaflet width compared to triploids (3.73 ± 0.47 cm) and tetraploids (4.09± 0.77 cm) (Figure 1C). Overall, leaf dimensions increased with ploidy level (Figure S1).

## Cannabinoids

As expected, for all measured cannabinoid concentrations, significant differences were found between cultivars except for CBN. In all cases, CBN levels were below the threshold of detection. These findings, in combination with the detection of very low levels of THC and CBD, indicate that degradation of the cannabinoid acids during the harvesting, sample preparation and storage was minimal. The total cannabinoid levels differed significantly between cultivars, with cultivar D, the only CBDA dominant cultivar, having a significantly lower total cannabinoid content (16.47 ± 3.98%) than cultivars A (21.43 ± 2.74%), B (21.72 ± 2.80%) and C (20.64 ± 3.62%) (Figure 2). Total cannabinoid content was influenced by ploidy level. A decreasing trend with increasing ploidy level was observed, with total cannabinoids of diploids (21.59 ± 3.25%) being significantly higher than triploids (19.19 ± 3.73%) and tetraploids (18.51 ± 4.26%) (Figure 2). The precursor cannabinoid CBGA displayed differences for both cultivars and ploidy (Figure 3). Cultivar A had the highest level (0.34 ± 0.12%), cultivar D had the lowest level (0.18 ± 0.11%), while cultivars B (0.28 ± 0.10%) and C (0.22± 0.10%) had intermediate levels. Significant differences were detected primarily between cultivar D and the three high THCA cultivars. Similar to total cannabinoids, CBGA content was significantly affected by ploidy level, with lower concentrations associated with increasing ploidy level. Diploids had significantly higher CBGA concentration (0.31 ± 0.12%) than triploids (0.24 ± 0.11%) and tetraploids (0.18 ± 0.10%) (Figure 3). Moreover, the interaction was also found significant. This significant difference was driven by the very low levels in triploids and tetraploids of cultivar D and by the differential response of

10.3389/fpls.2023.1233191

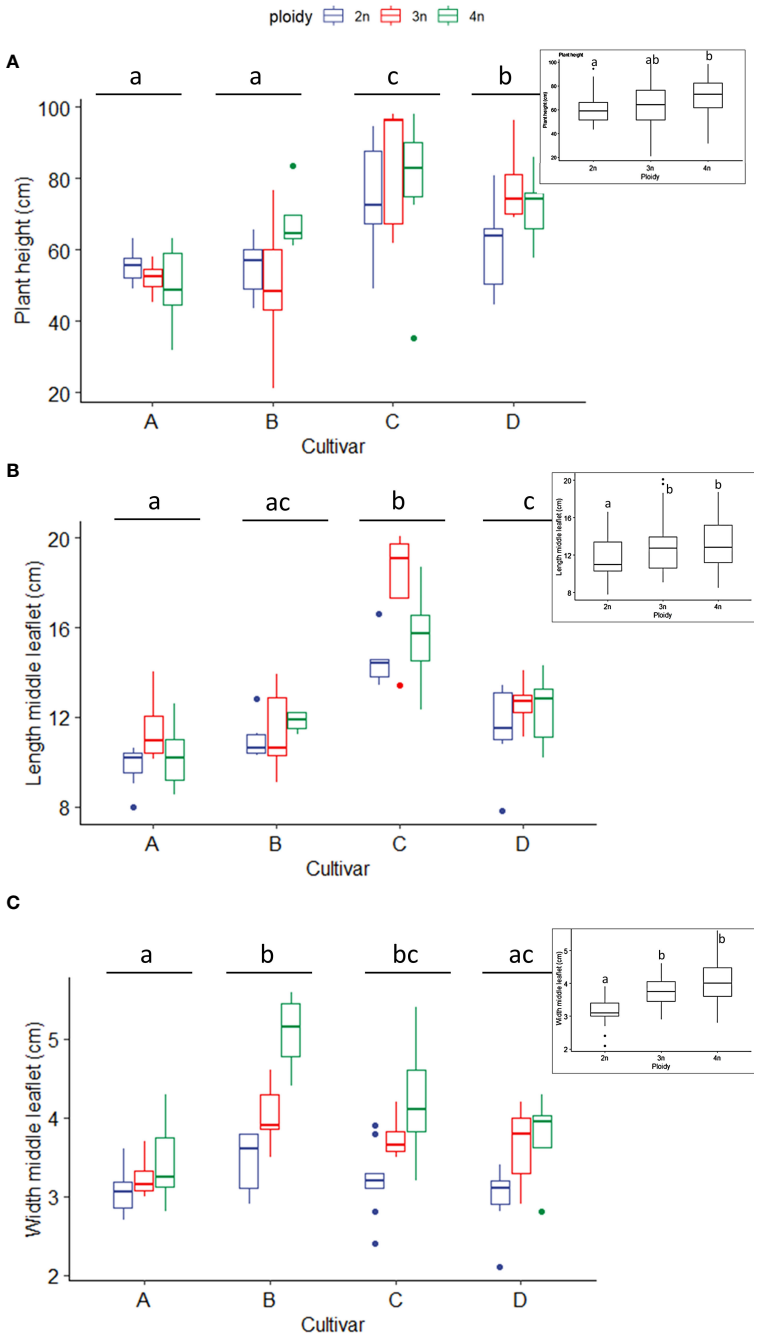



FIGURE 1
Boxplots of the **(A)** plant height (cm), **(B)** length of the middle leaflet (cm) and **(C)** width of the middle leaflet of a leaf (cm) of diploid (2n), triploid (3n) and tetraploid (4n) plants of four *Cannabis sativa* L. cultivars. **(A)**: Kruskal-Wallis rank sum test: Cultivar; $\chi 2$ = 38.269, p = 2.479e-08. Ploidy; $\chi 2$ = 7.7563, p = 0.02069. Different letters above bars indicate significant differences at p<0.05 with an all-pairs Dunn *post-hoc* analysis. The inserted graph displays the length of the middle leaflet length of the ploidy levels. **(B)** Logarithmic transformation was performed as indicated by a Box-Cox test. ANOVA: Cultivar; $F_{3,80}$ = 46.12, p < 2e-16. Ploidy; $F_{2,80}$ = 11.72, p = 3.42e-05. The inserted graph displays the width of the middle leaflet length of the ploidy levels. **(C)** Two-way ANOVA: Cultivar; $F_{3,77}$ = 8.629, p = 4.98e-05. Ploidy; $F_{2,77}$ = 27.244, p = 9.47e-10. The insert graphs show plant height between ploidy levels. Different letters above bars indicate significant differences at p<0.05 with a Tukey *post-hoc* analysis. For two samples no measurements were taken.

cultivar A, B and C. Indeed, for cultivar D, a significant -30.8% and a -71.9% decrease was detected in triploids and tetraploids, respectively. On the other hand, only a -10.5% decrease in triploids and a slight increase of +3.2% in tetraploids was detected for cultivar B. While for cultivars C and A, although a decrease was observed, it was not found to be significant, with a -37.0% and -32.2% for cultivar C, while -13.8% and -29.5% for cultivar A in the triploid and tetraploids, respectively.

As expected, the individual cannabinoids CBDA, CBD, THCA and THC significantly differed between cultivars since cultivar D is

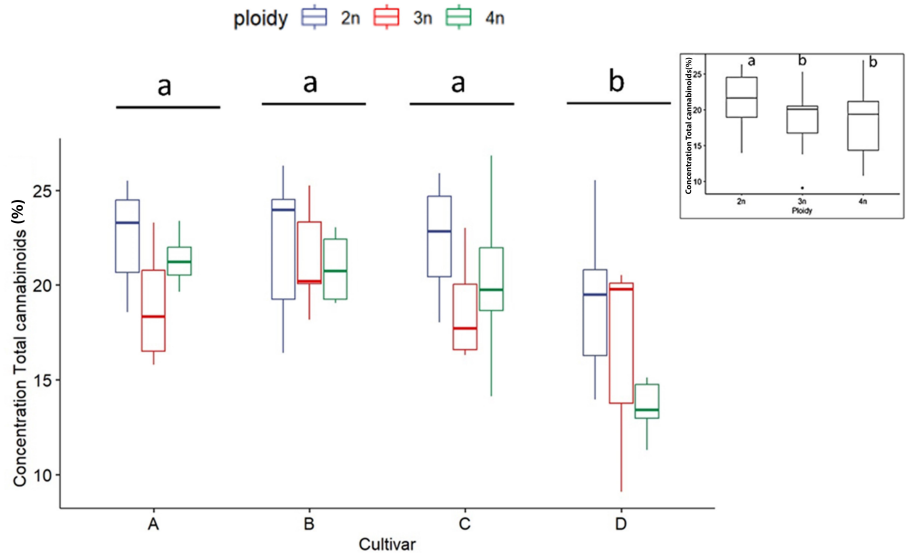



FIGURE 2
Boxplot of concentration total cannabinoids (expressed as % of dry flower weight) of a HPLC analysis of diploid (2n), triploid (3n) and tetraploid (4n) plants of four *Cannabis sativa* L. cultivars. Two-way ANOVA: Cultivar; $F_{3,82}$ = 12.248, p = 1.06e-06. Ploidy; $F_{2,82}$ = 8.672, p = 0.000383. The insert graph displays total cannabinoids of the ploidy levels: diploid, triploid and tetraploid. Different letters above bars indicate significant differences at p<0.05 with a Tukey *post-hoc* analysis.

high in CBDA and very low in THCA. For this reason, a statistical analysis of THCA, THC, CBDA and CBD was performed by separating the cultivars according to the chemotype.

In THCA dominant cultivars (A, B and C), THCA concentrations were influenced only by ploidy levels. For all three cultivars a significant decrease in the triploids (18.8 ± 2.8%) compared to diploids (21.0 ± 2.8%) was found. Tetraploids (19.5 ± 3.0%) were not significantly different from diploids and triploids (Figure 4A). In the three THCA dominant cultivars, the measured CBDA content was minimal, yet an effect of polyploidization could

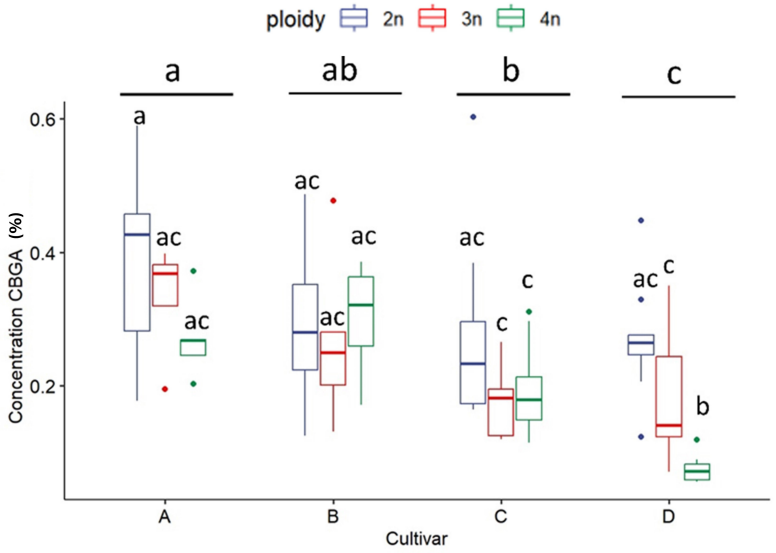



FIGURE 3
Boxplot of the concentration of CBGA (expressed as % of dry flower weight) of a HPLC analysis of flowers of diploid (2n), triploid (3n) and tetraploid (4n) plants of four *Cannabis sativa* L. cultivars. A logarithmic transformation was performed as indicated by a Box-Cox test. Two-way ANOVA: Cultivar; $F_{3,76}$ = 15.885, p = 4.12e-08. Ploidy; $F_{2,76}$ = 21.644, p = 3.63e-08. Cultivar x Ploidy; $F_{6,76}$ = 4.727, p = 0.000386. Different letters above bars indicate significant differences at p<0.05 with a Tukey *post-hoc* analysis.

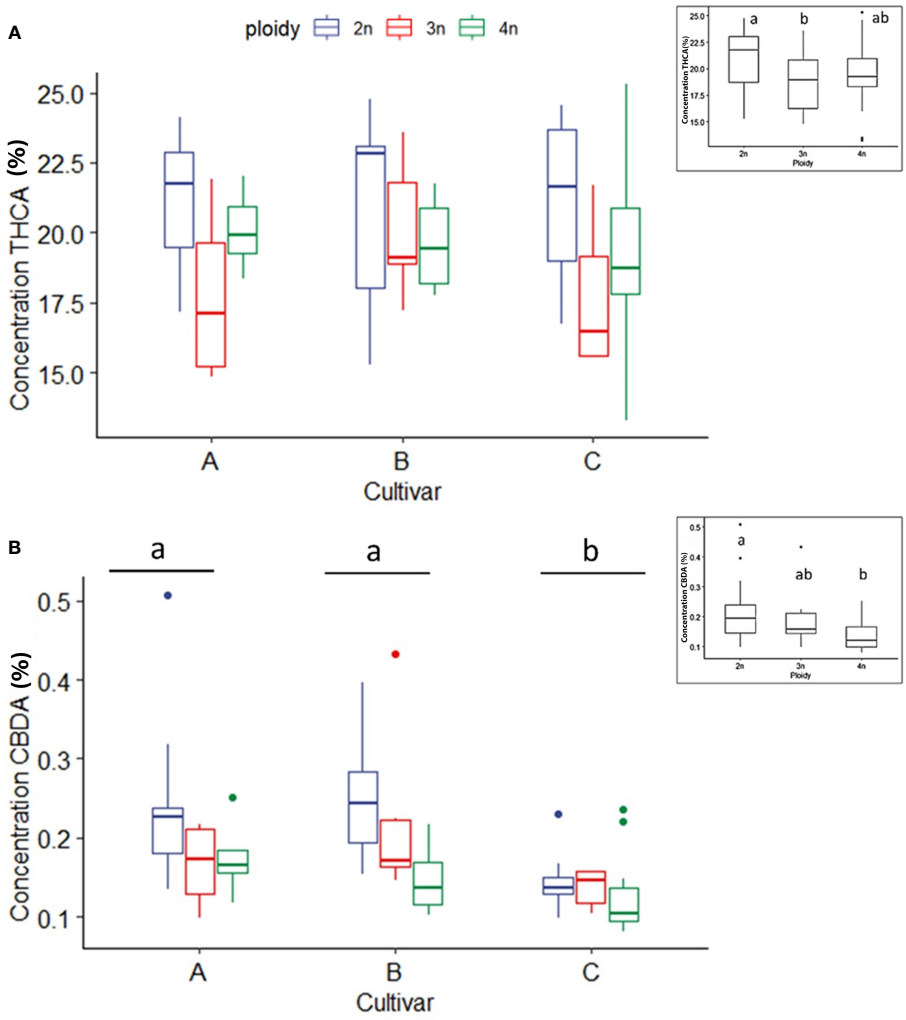



FIGURE 4
Boxplot of the concentration of THCA **(A)** and CBDA **(B)** (expressed as % of dry flower weight) of a HPLC analysis of flowers of diploid (2n), triploid (3n) and tetraploid (4n) plants of three high THCA Cannabis sativa L. cultivars. **(A)** One-way ANOVA: Ploidy; $F_{2,62} = 3.452$, $p = 0.0379$. **(B)** An inverted square root transformation was performed as indicated by a Box-Cox test. Two-way ANOVA: Cultivar; $F_{2,60} = 11.127$, $p = 0.000196$. Ploidy; $F_{2,60} = 9.879$, $p = 0.000196$. Different letters above bars indicate significant differences at $p<0.05$ with a Tukey *post-hoc* analysis. The insert graphs display CBDA and THCA concentration of the ploidy levels: diploid, triploid and tetraploid.

be detected. A significant decrease with increasing ploidy level was observed. Diploids displayed the highest levels of CBDA (0.21 ± 0.09%), followed by triploids (0.18 ± 0.08%) and tetraploids (0.14 ± 0.05%) (Figure 4B). Analysis of THC did not reveal any differences, while CBD could not be detected for the three high THCA cultivars (data not shown).

CBDA concentration in the CBDA dominant cultivar (D) was also affected significantly by ploidy level. In this case, the highest ploidy level presented the lowest CBDA concentration (12.14 ± 1.19%), while triploids had intermediate concentrations (15.02 ± 4.76%) in comparison to diploids (17.59 ± 3.20%) (Figure 5). For the high CBDA cultivar no differences between ploidy levels for CBD and THCA, could be found, while THC could not be detected in cultivar D.

## Headspace-GC-MS analysis

In total, 248 compounds were annotated (with an 85% NIST library score) from the flowers of the four *C. sativa* cultivars. As expected, most of the compounds annotated were monoterpenes (57) and sesquiterpenes (126). At the same time, also esters (2), alcohols (4), fatty acids (2), and cannabinoids such as $\Delta^9$-Tetrahydrocannabinol, $\Delta^1$-Tetrahydrocannabinol, Cannabichromene were also detected, while for the remaining 54 compounds the chemical family remained unassigned (Table S1). A comparison of the volatile compounds showed that cultivar C was the most diverse, containing the largest number of compounds detected (81 ± 14 peaks) compared to A (72 ± 10 peaks) and B (71 ± 14 peaks), while cultivar D was found in-between (77 ± 13 peaks). Interestingly no effect of ploidy or a ploidy x cultivar

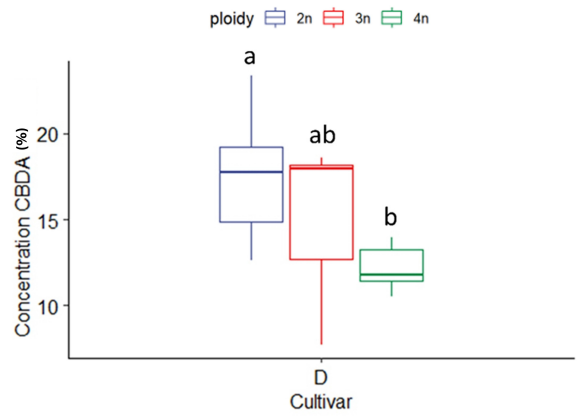



FIGURE 5
Boxplot of the concentration of CBDA (expressed as % of dry flower weight) of a HPLC analysis of flowers of diploid (2n), triploid (3n) and tetraploid (4n) plants of the high CBDA *Cannabis sativa* L. cultivar D. One-way ANOVA: Ploidy; $F_{2,20}$ = 7.281, p = 0.004. Different letters above bars indicate significant differences at p<0.05 with a Tukey *post-hoc* analysis.

interaction was detected for number of peaks (Figure 6A). Cultivar B showed the highest concentration of volatiles followed by C, A and D (Figure 6B). The interaction between ploidy and cultivar was found significant while the effect of ploidy was detected as almost significant (p = 0.071). A decrease of -21.89% was found in the triploids compared to diploids (Figure 6B). While analysis of the interaction revealed a differential response of the cultivars to polyploidization regarding volatile concentrations. The main differences were detected between tetraploids of cultivar B and C, presenting the highest overall concentrations compared to tetraploids of cultivar A and tetraploids and triploids of cultivar D, which displayed the lowest overall concentrations. Additionally, a significant decrease was also detected between the diploids and tetraploids of cultivar D (-62.7%), while although not significant in cultivar B, tetraploids displayed a +75.7% increase in volatile concentrations compared to B diploids (Figure 6B).

To investigate the changes in volatiles, a principal component analysis (PCA) was carried out. The multivariate analysis showed a trend in the sample separation driven by *C. sativa* cultivars (Figure S2A). The main separation along the PC1 was due mainly to cultivars C and D, while the PCA plot did not show separation among the three ploidy levels (Figure S2B). To better discriminate cultivars and ploidy level effects, a supervised discriminant analysis, Orthogonal partial least square analysis-discriminant analysis (OPLS-DA), was used. The OPLS-DA model showed separation between the cultivars C and D while A and B clustered together (Figure 7A). The model was highly validated ($R^2X$= 0.92, $Q^2$ = 0.75 and p-value in CV-ANOVA= 0). Using the same OPLS-DA model, but differentiating the samples based on ploidy levels, a gradient separation was achieved, though the model was not validated $R^2X$= 0.45, $Q^2$ = 0.07 and p-value in CV-ANOVA = 1 (Figure 7B).

Given the different volatile compositions of the cultivars, we therefore also analyzed the effect of ploidy level on each cultivar separately with OPLS-DA (Figure 8). Although this strategy increased the resolution of ploidy separation in all four cultivars, only for cultivar B the model was validated with a $Q^2$ value larger than 0.4 (Figure 8B).

To further investigate the changes in volatiles caused by the different ploidy levels on individual compounds, 25 compounds were extracted from the model separating the cultivars. In addition, 29 compounds with a VIP score larger than one were extracted from the model separating the ploidy levels, although this model was not validated. Checking for overlapping compounds between models revealed that of the 29 extracted from the ploidy model, 21 were already present in the cultivar model. This resulted in a pool of 33 individual compounds, which except for one case (Butanoic acid, hexyl ester), concerned monoterpenes and sesquiterpenes (Figures 9, 10). Further investigation revealed that changes in ploidy levels greatly influenced these individual compounds and often, cultivars were affected differently (Tables 2, 3).

Analysis of the 33 compounds in triploids revealed the same pattern observed in total volatile content but on a deeper level. In triploid plants, the majority of compounds decreased with cultivar A, showing a decrease in the concentration of 22 compounds while 8 increased compared to diploids. In cultivar B, 21 compounds decreased while 7 increased. In cultivar C, 18 compounds decreased while 11 increased and in cultivar D the concentration of 27 compounds decreased while only 5 increased compared to diploids (Table S2). Overall, this data indicates that although a small proportion of compounds increased in concentration, the majority decreased, with cultivar D being the most affected and cultivar C the least (Table S2).

An interesting trend was observed in tetraploids, with cultivars B and C showing an important difference compared to cultivars A and D. In cultivar A the concentrations of 24 compounds decreased while 6 increased compared to diploids. Cultivar D also showed a decrease in concentrations of 26 compounds and an increase in 5 compared to diploids. In contrast, in cultivar B, the concentration of only 8 compounds decreased while 18 increased. Similarly in cultivar C the concentration showed a reduction in 9 compounds while 20 increased compared to diploids (Table S2).

Additionally, several compounds present in diploids were no longer detected in triploids and tetraploids (Table S2). In cultivar A, β-bisabolene, Guaia-6,9-diene, α-Costol were not detected in both triploids and tetraploids. The compounds γ-Cadinene, Eudesm-7 (11)-en-4-ol and 1,4,7,-Cycloundecatriene, 1,5,9,9-tetramethyl-, Z, Z,Z- were not detected in triploids and Selina-5,11-diene was not detected in tetraploids A. In cultivar B α-Costol was not detected in both triploids and tetraploids and Eudesm-7(11)-en-4-ol and α-Amorphene were not detected in triploids while 3-Eudesmen-11-ol, Guaiol, *cis*-α-Bergamotene, Guai-1(10)-en-11-ol and γ-Muurolene were not detected in tetraploids. In cultivar C only α-Costol disappeared in triploids. On the other hand, several compounds were no longer detected in both triploids and tetraploids of cultivar D, Eudesm-7(11)-en-4-ol, Guaia-6,9-diene, α-Amorphene and *cis*-Caryophyllene while γ-Muurolene and Butanoic acid, hexyl ester were not present in triploids and Selina-5,11-diene was not present in tetraploids (Table S2).

Analysis across cultivars revealed that Selina-5,11-diene was not present in both tetraploids of cultivars A and D, Eudesm-7(11)-en-

10.3389/fpls.2023.1233191

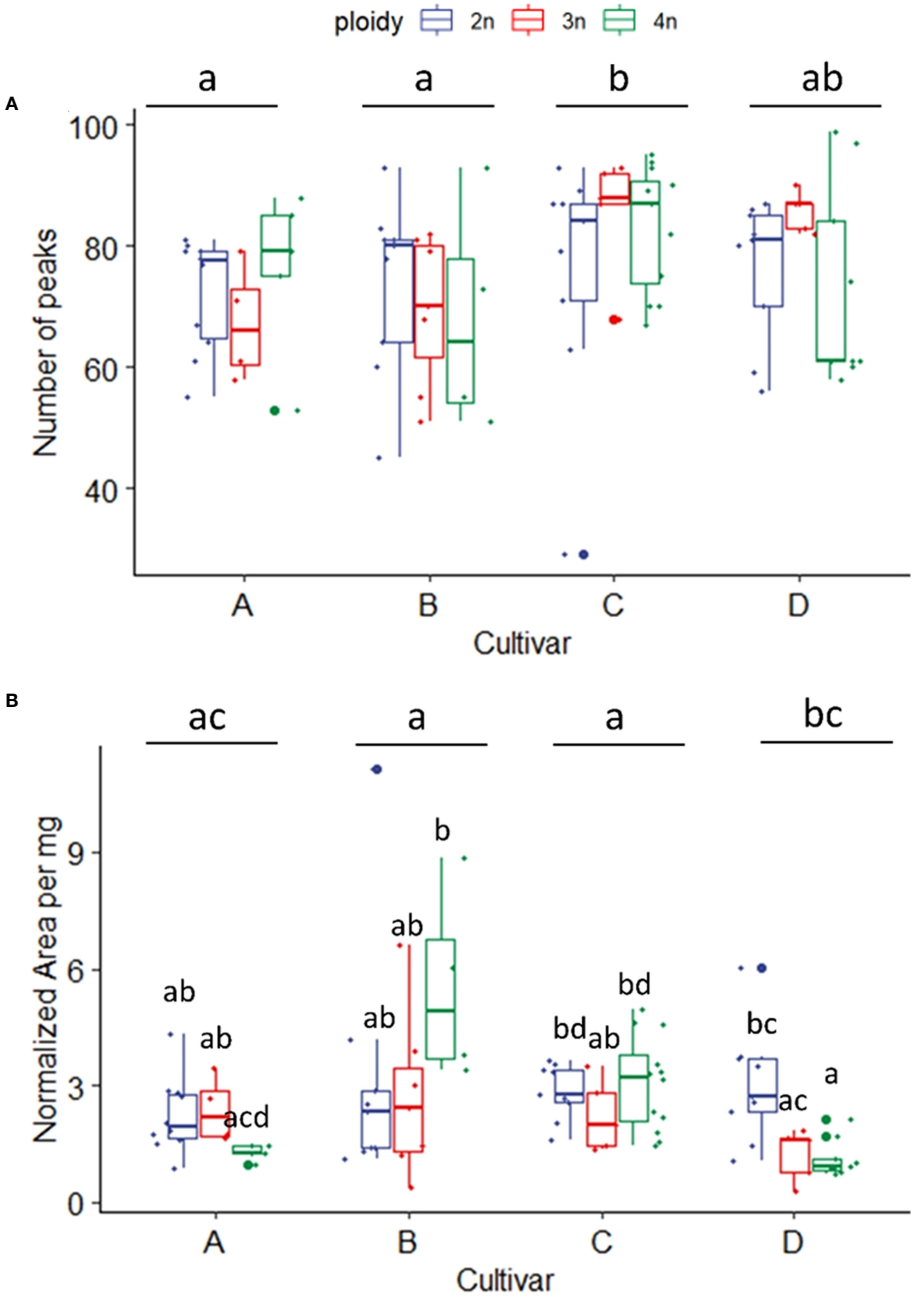



FIGURE 6
**(A)** Boxplot of total number of peaks of a headspace GC-MS analysis of diploid (2n), triploid (3n) and tetraploid (4n) plants of four *Cannabis sativa* L. cultivars. Kruskal-Wallis rank sum test: Cultivar; $\chi 2 = 12.624$, $p = 0.005$. Ploidy; $\chi 2 = 0.8091$, $p = 0.667$. Different letters above bars indicate significant differences at $p<0.05$ with a Kruskal-Wallis and a *post-hoc* all-pairs Dunn test. **(B)** Boxplot of total peak area per mg dry weight of a headspace GC-MS analysis of diploid, triploid and tetraploid plants of four *Cannabis sativa* L. cultivars. A logarithmic transformation was performed as indicated by a Box-Cox test. Two-way ANOVA: Cultivar; $F_{3,76} = 6.558$, $p = 0.00053$. Ploidy; $F_{2,76} = 2.741$, $p = 0.071$. Cultivar x Ploidy; $F_{6,76} = 4.082$, $p = 0.001$. Different letters above bars indicate significant differences at $p<0.05$ with a Tukey *post-hoc* analysis.

4-ol was not present in triploids of A, B and D, Guaia-6,9-diene was not present in both triploids and tetraploids of A and D. α-Costol was not present in triploids of A, B, C and tetraploids A and B while α-Amorphene was not present in triploids of A, B and D.

In cultivars C and D, some compounds that were not present in diploids were detected in the higher ploidy levels. In cultivar C, α-Amorphene was present in triploids and tetraploids, in concentrations that were more than doubled in tetraploids compared to triploids. Interestingly in cultivar D Citral and β-Cymene were present in both triploids and tetraploids while γ-Cadinene only was present in triploids.

Statistical analysis revealed that, for four compounds, significant differences in ploidy level were detected when pooling the four cultivars. Three compounds (*cis*-2-Norbornanol, Fenchol, and D-Limonene) showed the same pattern with strikingly similar numbers. Indeed, a significant decrease in triploids (*cis*-2-

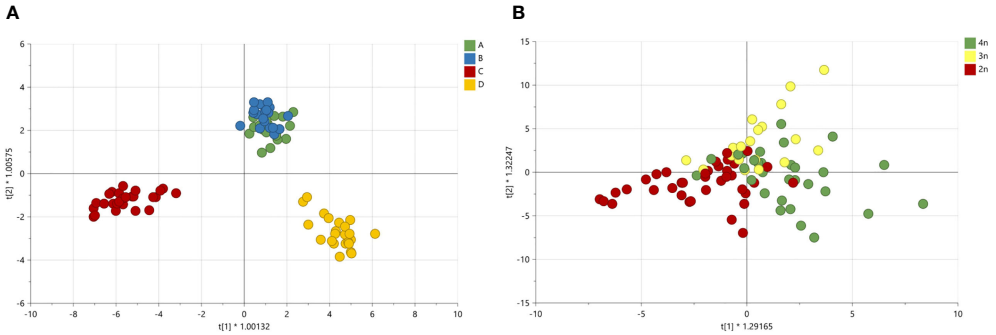



FIGURE 7
Orthogonal partial least square analysis-discriminant analysis of GC-MS analysis of 4 *Cannabis sativa* L. cultivars A, B, C and D **(A)**. The model was highly validated ($R^2X$= 0.92, $Q^2$ = 0.75 and p value in CV-ANOVA= 0). OPLS-DA on ploidy levels **(B)**. The model was not validated ($R^2X$= 0.45, $Q^2$ = 0.07 and p value in CV-ANOVA= 1).

Norbornanol -55.31%; Fenchol -43.86%; D-Limonene -47.22%) and a decrease in tetraploids (*cis*-2-Norbornanol -20.94%; Fenchol -18.91%; D-Limonene -27.54%) was observed (Figure 9). On the other hand, Eudesm-7(11)-en-4-ol showed an increase in tetraploids (+85.39%) while there was a decrease in triploids (-81.77%), leading to a significant difference between triploids and tetraploids plants. Further analysis on Eudesm-7(11)-en-4-ol revealed that in triploids of cultivar A, B and D, the compound was undetectable, while the increase in the tetraploids was driven by cultivar B (+111.23%) and C (+112.92%) (Figure 9).

Fenchone decreased in both ploidy levels across the four cultivars, apart from tetraploid C, where a slight increase of +5.45% was detected. Moreover, the decrease in cultivar D was found to be significant, with -81.52% in triploids and -91.73% in tetraploids. β-Bisabolene was undetectable in cultivars B and C, while in cultivar D, it significantly decreased in both triploids (-88.30%) and tetraploids (-94.43%) and decreased under detection level in both triploids and tetraploids of cultivar A. The remaining compounds all showed a differential response to polyploidization depending on the cultivars under analysis,

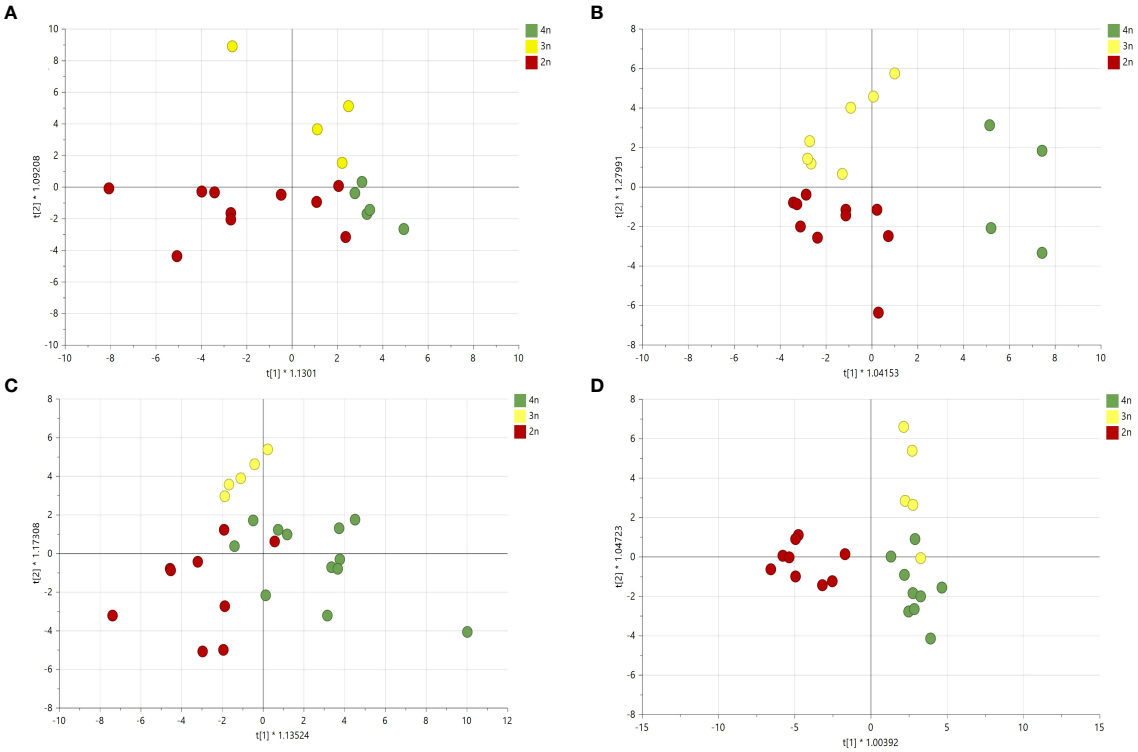



FIGURE 8
Orthogonal partial least square analysis-discriminant analysis of GC-MS analysis of four *Cannabis sativa* L. cultivars based on ploidy levels **(A)** cultivar A: $Q^2$ 0.24, **(B)** cultivar B: $Q^2$ 0.44, **(C)** cultivar C: $Q^2$ 0.04 and **(D)** cultivar D: $Q^2$ 0.27.

TABLE 2 Results of Kruskal-Wallis rank sum tests of the total peak area per mg flower dry weight of 25 compounds with the highest VIP score in an OPLS-DA analysis of a headspace GC-MS analysis of diploid, triploid and tetraploid plants of four different *Cannabis sativa* L. cultivars.

| Fig | Common name | Class | Cultivar | | Ploidy | | Ploidy levels per cultivar | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | A | | B | | C | | D | |
| | | | $\chi^2$ | *p* | $\chi^2$ | *p* | $\chi^2$ | *p* | $\chi^2$ | *p* | $\chi^2$ | *p* | $\chi^2$ | *p* |
| A | Fenchone | Monoterpene | 36.7 | 0.000 | 4.9 | 0.087 | 0.8 | 0.661 | 0.1 | 0.961 | 2.0 | 0.362 | 14.2 | 0.001 |
| B | Linalool | Monoterpene | 48.1 | 0.000 | 0.9 | 0.649 | 4.4 | 0.112 | 0.8 | 0.673 | 1.8 | 0.400 | 1.9 | 0.384 |
| C | Germacrene B | Sesquiterpene | 48.1 | 0.000 | 2.8 | 0.251 | 8.2 | 0.016 | 4.4 | 0.111 | 3.5 | 0.175 | 3.4 | 0.187 |
| D | L-α-Terpineol | Monoterpene | 47.2 | 0.000 | 1.6 | 0.457 | 2.5 | 0.283 | 9.3 | 0.009 | 5.4 | 0.067 | 4.1 | 0.128 |
| E | Fenchol | Monoterpene | 37.0 | 0.000 | 8.7 | 0.013 | 6.0 | 0.050 | 8.1 | 0.018 | 5.7 | 0.059 | 12.0 | 0.002 |
| F | γ-selinene | Sesquiterpene | 37.4 | 0.000 | 0.7 | 0.718 | 7.0 | 0.030 | 0.7 | 0.704 | 2.1 | 0.356 | 2.6 | 0.275 |
| G | Selina-3,7(11)-diene | Sesquiterpene | 34.7 | 0.000 | 2.6 | 0.268 | 7.8 | 0.020 | 4.8 | 0.093 | 2.7 | 0.264 | 5.3 | 0.070 |
| H | β-Maaliene | Sesquiterpene | 35.0 | 0.000 | 0.1 | 0.973 | 6.7 | 0.035 | 8.1 | 0.018 | 3.4 | 0.186 | 4.3 | 0.115 |
| I | *cis*-2-Norbornanol | Monoterpene | 30.6 | 0.000 | 6.4 | 0.039 | 0.9 | 0.633 | 3.7 | 0.155 | 4.8 | 0.089 | 3.5 | 0.171 |
| J | 3-Eudesmen-11-ol | Sesquiterpene | 26.2 | 0.000 | 3.3 | 0.195 | 1.2 | 0.543 | 5.5 | 0.063 | 1.3 | 0.530 | 16.0 | 0.000 |
| K | Butanoic acid, hexyl ester | Ester | 26.9 | 0.000 | 0.3 | 0.841 | 2.7 | 0.256 | 1.9 | 0.395 | NA | NA | 1.2 | 0.541 |
| L | Caryophyllene | Sesquiterpene | 43.8 | 0.000 | 2.0 | 0.368 | 3.7 | 0.150 | 7.5 | 0.023 | 9.8 | 0.007 | 16.9 | 0.000 |
| M | Bicyclogermacrene | Sesquiterpene | 39.8 | 0.000 | 2.1 | 0.348 | 0.3 | 0.830 | 4.2 | 0.120 | 2.2 | 0.338 | 1.5 | 0.463 |
| N | β-selinene | Sesquiterpene | 41.6 | 0.000 | 1.0 | 0.589 | 6.2 | 0.044 | 2.6 | 0.275 | 1.5 | 0.475 | 1.4 | 0.495 |
| O | Guaiol | Sesquiterpene | 20.4 | 0.000 | 2.7 | 0.255 | 2.9 | 0.237 | 3.5 | 0.171 | 1.3 | 0.534 | 14.9 | 0.001 |
| P | 1,4,7,-Cycloundecatriene, 1,5,9,9-tetramethyl-, Z,Z,Z- | Sesquiterpene | 15.9 | 0.001 | 1.3 | 0.509 | 3.2 | 0.204 | 1.7 | 0.437 | 1.8 | 0.399 | 3.1 | 0.217 |
| Q | Ylangene | Sesquiterpene | 39.9 | 0.000 | 1.7 | 0.419 | 6.3 | 0.043 | 0.2 | 0.884 | 8.3 | 0.016 | 1.0 | 0.613 |
| R | 3-Methylcamphenilol | Monoterpene | 25.5 | 0.000 | 4.2 | 0.121 | 1.4 | 0.491 | 2.3 | 0.313 | 2.4 | 0.302 | 3.7 | 0.158 |
| S | β-Bisabolene | Sesquiterpene | 24.0 | 0.000 | 4.2 | 0.121 | 0.9 | 0.638 | NA | NA | NA | NA | 7.2 | 0.027 |
| T | D-Limonene | Monoterpene | 26.2 | 0.000 | 7.4 | 0.024 | 5.0 | 0.082 | 3.1 | 0.218 | 4.0 | 0.135 | 12.2 | 0.002 |
| U | Selina-5,11-diene | Sesquiterpene | 18.9 | 0.000 | 4.1 | 0.127 | 4.2 | 0.120 | 0.7 | 0.714 | 2.3 | 0.317 | 5.4 | 0.066 |
| V | Eudesm-7(11)-en-4-ol | Sesquiterpene | 19.6 | 0.000 | 4.4 | 0.108 | 1.9 | 0.389 | 3.0 | 0.224 | 2.8 | 0.248 | 9.2 | 0.010 |
| W | *cis*-α-Bergamotene | Sesquiterpene | 19.5 | 0.000 | 0.6 | 0.705 | 3.8 | 0.150 | 2.4 | 0.299 | 1.0 | 0.618 | 1.7 | 0.417 |
| X | γ-Eudesmol | Sesquiterpene | 10.3 | 0.016 | 4.4 | 0.108 | 1.9 | 0.389 | 3.0 | 0.224 | 2.8 | 0.248 | 9.2 | 0.010 |
| Y | (-)-guaia-6,9-diene | Sesquiterpene | 31.9 | 0.000 | 4.2 | 0.117 | 1.9 | 0.387 | 1.9 | 0.387 | 3.4 | 0.179 | 1.6 | 0.459 |

Compounds are ranked with the first compound having the highest VIP score. Letters in the first row refer to corresponding boxplots in Figure 9. NA indicates that the compounds were not detected in the cultivar and therefore a statistical test could not be performed.

particularly comparing tetraploids of cultivar A and D versus B and C (Table S2).

Some of the most striking differences were detected in β-Maaliene, where a significant decrease was detected in tetraploids of cultivar A (-57.67%), while for cultivar B a significant increase was detected in tetraploids (+164.82%). Ylangene also showed a differential response of the cultivars in the tetraploid state, with cultivar A showing a significant decrease of -72.75%. In contrast, a significant increase of +182.13% for cultivar C was observed. Similarly, for β-Selinene, cultivar A showed a significant decrease in tetraploids (-56.66%) while cultivar B, although not significantly, increased by +113.96% in tetraploids. For γ-Eudesmol, cultivar D showed a significant -69.06% decrease in tetraploids, while cultivar A showed a +105.39% increase. Analysis of *cis*-α-Bergamotene revealed that concentrations in tetraploids of cultivar A increased by +78.04%, in cultivar C by +83.09%, while in tetraploids of cultivar B, the compound was not detected anymore. Interestingly, the rare compound Guaia-6,9-diene increased in tetraploids of cultivar C by +145.45% and +3378.74% in tetraploids of cultivar B. On the other hand, the compound could not be detected anymore in tetraploids of cultivars A and D.

Since cultivar B showed for several traits a positive reaction to polyploidization and the OPLS-DA model was validated, the analysis was also performed on the VIPs compounds of cultivar B

10.3389/fpls.2023.1233191

TABLE 3 Results of Kruskal-Wallis rank sum tests of the total peak area per mg flower dry weight of 8 compounds with the highest VIP score in an OPLS-DA analysis of a headspace GC-MS analysis of diploid, triploid and tetraploid plants.

| Fig | Common name | Class | Cultivar | | Ploidy | | Ploidy levels per cultivar | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | A | | B | | C | | D | |
| | | | $\chi^2$ | *p-value* | $\chi^2$ | *p-value* | $\chi^2$ | *p-value* | $\chi^2$ | *p-value* | $\chi^2$ | *p-value* | $\chi^2$ | *p-value* |
| A | α-Costol | Sesquiterpene | 10.5 | 0.015 | 3.7 | 0.154 | 1.900 | 0.387 | 1.222 | 0.543 | 3.192 | 0.203 | NA | NA |
| B | α-Amorphene | Sesquiterpene | 5.6 | 0.131 | 3.6 | 0.160 | 1.608 | 0.448 | 1.807 | 0.405 | 4.987 | 0.083 | 1.556 | 0.459 |
| C | γ-Cadinene | Sesquiterpene | 16.3 | 0.001 | 2.2 | 0.324 | 0.916 | 0.633 | 0.586 | 0.746 | 1.245 | 0.537 | 3.600 | 0.165 |
| D | *cis*-Caryophyllene | Sesquiterpene | 26.3 | 0.000 | 4.9 | 0.084 | NA | NA | NA | NA | 4.146 | 0.126 | 3.253 | 0.197 |
| E | Citral | Monoterpene | 11.7 | 0.008 | 6.2 | 0.045 | NA | NA | NA | NA | NA | NA | 7.196 | 0.027 |
| F | Guai-1(10)-en-11-ol | Sesquiterpene | 8.1 | 0.045 | 4.0 | 0.134 | 2.735 | 0.255 | 2.385 | 0.303 | 2.557 | 0.279 | 8.273 | 0.016 |
| G | β-Cymene | Monoterpene | 11.7 | 0.008 | 3.2 | 0.199 | NA | NA | NA | NA | NA | NA | 3.613 | 0.164 |
| H | γ-Muurolene | Sesquiterpene | 21.0 | 0.000 | 4.3 | 0.115 | 1.943 | 0.379 | 4.079 | 0.130 | 5.416 | 0.067 | 1.553 | 0.460 |

Compounds are ranked with the first compound having the highest VIP score. Letters in the first row refer to corresponding boxplots in Figure 10. NA indicates that the compounds were not detected in the cultivar and therefore a statistical test could not be performed.

separately (Table 4). The analysis was performed on 15 compounds, of which 8 showed a significant increase in cultivar B tetraploids compared to diploids (β-Maaliene +164.83%; Endo-Borneol +120.50%; Caryophyllene +118.4%; 3-Carene +1874.68%; L-α-Terpineol +150.87%; Fenchol +59.39%; β-Thujene +692.22%; *Trans*-2-Pinanol +309.3%) while two increased although the statistics could not detect a significant effect of ploidy (D-Guaiene +298.38%; Bicyclogermacrene +274.77%) (Figure 11).

Additionally, analysis of outliers detected an interesting effect of polyploidization in cultivar C. Several genotypes in this cultivar

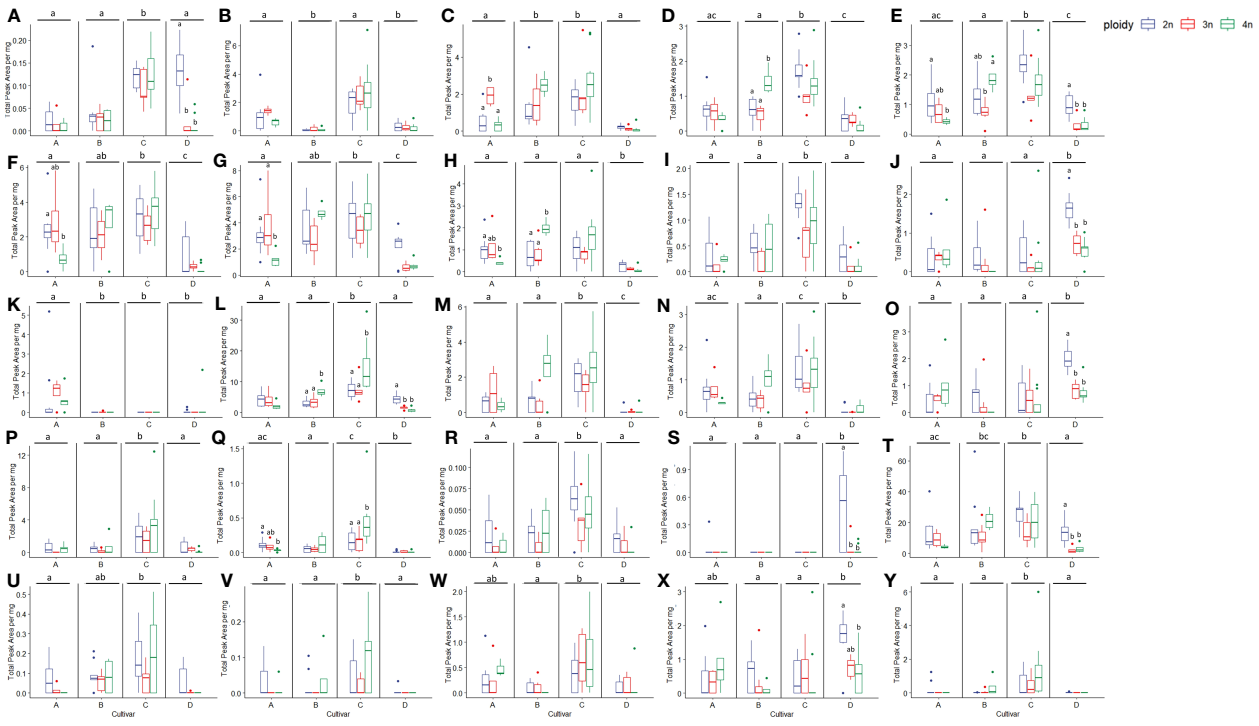



FIGURE 9
Boxplots of the total peak area per mg dry weight of 25 compounds with the highest VIP score in an OPLS-DA analysis separating cultivars of a headspace GC-MS analysis of diploid (2n), triploid (3n) and tetraploid (4n) plants of four *Cannabis sativa* L. cultivars. Kruskal-Wallis rank sum tests were performed for all compounds, for results of the tests see corresponding Table 2. Different letters above bars indicate significant differences at p<0.05 with a Kruskal-Wallis within the cultivar and between cultivars and *post hoc* all pairs Dunn test. **(A)** Fenchone, **(B)** Linalool, **(C)** Germacrene B, **(D)** L-α-Terpineol, **(E)** Fenchol, **(F)** γ-selinene, **(G)** Selina-3,7(11)-diene, **(H)** β-Maaliene, **(I)** *cis*-2-Norbornanol, **(J)** 3-Eudesmen-11-ol, **(K)** Butanoic acid, hexyl ester, **(L)** Caryophyllene, **(M)** Bicyclogermacrene, **(N)** β-Selinene, **(O)** Guaiol, **(P)** 1,4,7,-Cycloundecatriene,1,5,9,9-tetramethyl-,Z,Z,Z-, **(Q)** Ylangene, **(R)** 3-Methylcamphenilol, **(S)** β-Bisabolene, **(T)** D-Limonene, **(U)** Selina-5,11-diene, **(V)** Eudesm-7(11)-en-4-ol, **(W)** *cis*-α-Bergamotene, **(X)** γ-Eudesmol, **(Y)** (-)-Guaia-6,9-diene.

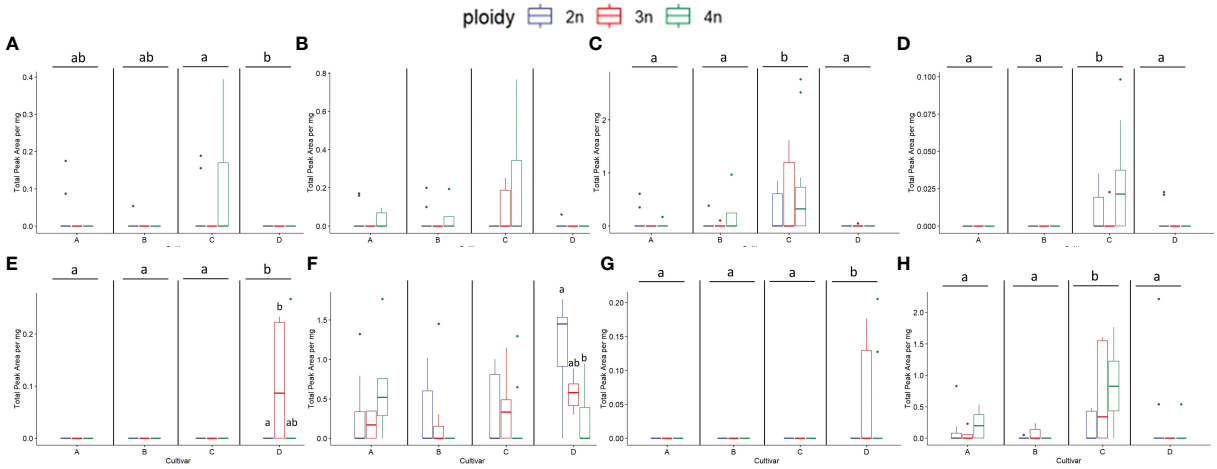



FIGURE 10
Boxplots of the total peak area per mg dry weight of 8 compounds with the highest VIP score in an OPLS-DA analysis separating ploidy levels of a headspace GC-MS analysis of diploid (2n), triploid (3n) and tetraploid (4n) plants of four *Cannabis sativa* L. cultivars. Kruskal-Wallis rank sum tests were performed for all metabolites, for their results see Table 3. **(A)** α-Costol, **(B)** α-Amorphene, **(C)** γ-Cardinene, **(D)** *cis*-Caryoplhyllene, **(E)** Citral, **(F)** Guai-1(10)-en-11-ol, **(G)** β-Cymene, **(H)** γ-Muurolene. Different letters above bars indicate significant differences at p<0.05 with a Kruskal-Wallis and an all-pairs Dunn test within the cultivar.

consistently presented very high levels of rare compounds compared to diploid ancestors, identifying them as elite tetraploid genotypes. Indeed, plant ID 82 presented a large increase in 8 compounds compared to diploid (β-Maaliene +336.89%; Caryophyllene +354.78%; 1,4,7,-Cycloundecatriene, 1,5,9,9-tetramethyl-, Z,Z,Z- +504.81%; Ylangene +830.39%; β-Selinene +145.86%; Guaia-6,9-diene +1024.29%; α-Amorphene +inf%; α-Costol +932.78%) while plant ID 91 showed a marked increase in 3 rare compounds (Guaiol + 569.55%; *cis*-Caryophyllene +913.95%; 3-Eudesmen-11-ol +527.32%). Moreover, further investigation revealed that plant ID 82 was also a top performer for total volatile content being sixth out of 89 plants. Given these results, cannabinoid data was also screened, and C tetraploid genotype ID 85 was found to be the best performer for total cannabinoids (26.84%) and linalool (+292% compared to average diploid).

## Discussion

In recent years, polyploidy has gained attention as a potential strategy for rapidly improving *C. sativa*. Compared to other modern crops, *C. sativa* has yet to benefit from this established biotechnological application. Indeed, the observed effects on plant morphology and

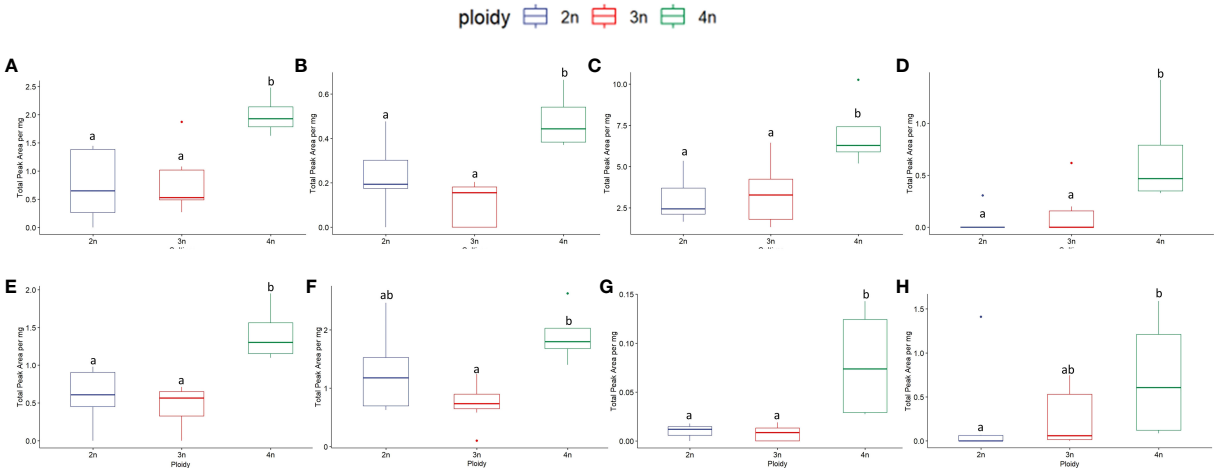



FIGURE 11
Boxplots of the total peak area per mg flower dry weight of 8 compounds with the highest VIP score in an OPLS-DA analysis separating ploidy levels of a headspace GC-MS analysis of diploid (2n), triploid (3n) and tetraploid (4n) plants of *Cannabis sativa* L. cultivar B. ANOVAs were performed to assess the differences between ploidy levels for this cultivar in the case of non-parametric data the non-parametric alternative Kruskal-Wallis rank sum tests was performed, for results see Table 4. Different letters above bars indicate significant differences at p<0.05 of an ANOVA and Tukey *post-hoc* analysis or a Kruskal-Wallis and an all-pairs Dunn test. **(A)** β-Maaliene, **(B)** *endo*-Borneol, **(C)** Caryophyllene, **(D)** 3-Carene, **(E)** L-α-Terpineol, **(F)** Fenchol, **(G)** β-Thujene, **(H)** *trans*-2-Pinanol.

TABLE 4 Results of ANOVAs or Kruskal-Wallis rank sum tests of the total peak area per mg dry flower weight of 8 compounds with the highest VIP score in an OPLS-DA analysis of a headspace GC-MS analysis of diploid, triploid and tetraploid plants of cultivar B. Letters in the first row refer to corresponding boxplots in Figure 11.

| Fig. | Common name | Class | Ploidy | | Ploidy | |
|---|---|---|---|---|---|---|
| | | | F-value | p-value | $\chi 2$ | p-value |
| A | β -Maaliene | Sesquiterpene | 8.565 | 0.003 | | |
| B | *endo*-Borneol | Monoterpene | 10.040 | 0.001 | | |
| C | Caryophyllene | Sesquiterpene | 8.460 | 0.003 | | |
| D | 3-Carene | Monoterpene | | | 10.723 | 0.005 |
| E | L- α-Terpineol | Monoterpene | 10.980 | 0.001 | | |
| F | Fenchol | Monoterpene | 6.562 | 0.008 | | |
| G | β -Thujene | Monoterpene | | | 9.505 | 0.009 |
| H | *trans*-2-Pinanol | Monoterpene | | | 5.776 | 0.056 |

secondary metabolite production in other medicinal and aromatic plants make it a very promising field of research.

Here we found that contrary to previous observations (Mansouri and Bagheri, 2017), plant height significantly increased with increasing ploidy levels. However, analysis of separate cultivars also highlighted a differential response, with cultivar D showing significantly taller triploids than diploids, while B triploids were shorter than their diploids counterpart. On the other hand, leaf morphology observations agreed with previous observations (Mansouri and Bagheri, 2017; Parsons et al, 2019). Leaf size increased significantly, with both leaflet length and width increasing with increasing ploidy level. Overall, our observations also align with the "giga" phenotype observed in other plant species after polyploidization (Sattler et al., 2016).

Investigation of secondary metabolites revealed several surprising changes. First, cannabinoids appear to be negatively affected by polyploidization, as the concentration of total cannabinoids decreased in higher ploidy levels. Moreover, major cannabinoids THCA, CBDA, as well as CBGA followed the same trend as observed for total cannabinoids. The significant decrease in concentration observed in total cannabinoids, -13.27% in triploids and - 28.89% in tetraploids of the CBDA dominant cultivar, and -11.76% in triploids and -6.89% in tetraploids of THCA dominant cultivars, has relevant implications regarding the uses of polyploids for producing major cannabinoids. These results are in contrast with the observation by Crawford et al. (2021), which observed an increase in total cannabinoids in polyploids of a drug-type CBGA dominant cultivar, highlighting once more the cultivar-dependent response to polyploidization. Regarding the concentration of CBGA, the results agree with the observation of Parson et al. (2019) who also detected a circa -30% decrease. Moreover, considering that the limitation on the availability of the precursor cannabinoid CBGA is a strong limiting factor for the accumulation of downstream cannabinoids (Laverty et al., 2019). These data point to the probable downregulation of the cannabinoid pathway in the higher ploidy levels. From these results, it appears that in modern drug-type *C. sativa* cultivars, where cannabinoid production has already been pushed to extremely high levels (>20% of dry flower weight) by artificial selection, in an event such as polyploidization, where plants undergo a highly stressful genomic reorganization, changes might favor other metabolic pathways as observed by Mansouri and Bagheri (2017).

A much more complex suite of changes was detected regarding the presence and concentration of volatiles. First of all, triploids showed a clear decrease in total volatile content. On the other hand, tetraploids depending on the cultivars, reacted differently, with cultivars B and C affected positively, while in cultivars A and D, concentrations decreased.

The VIPs extracted from the multivariate analysis yielded the 33 individual compounds having the most influence on the separation between cultivars and ploidy levels. Analysis of these compounds highlighted to a deeper level several interesting findings. Generally, triploids' volatiles were negatively affected across cultivars, apart from specific cases. Plants in the tetraploid state showed a differential response, with many compounds in C and B increasing while at the same time in A and D decreasing. Interestingly, several rare compounds appeared in higher ploidy levels. Cultivar C presented a couple of elite tetraploid genotypes, highlighting the role of within cultivar genotypic variability as an important factor in *C. sativa* polyploids. In cultivars A and D, the concentrations of most of the compounds decreased, and several were not detected in higher ploidy levels, pointing to the downregulation of the terpene pathway in these cultivars. On the other hand, in cultivar B, although an increase in concentration was detected for most of the compounds at higher ploidy levels, some also disappeared. These changes point to the probable gene-silencing effects caused by polyploidization in this specific cultivar (Eckardt, 2010).

This complex suite of changes have important implications for the medicinal value of the plants. Indeed, terpenes play an essential role in the bioactivity of *C. sativa* extracts and can contribute to pharmacological activity via the entourage effect (Russo, 2011). Some compounds that were found to increase significantly in cultivars B and C display important medicinal properties. "*Endo*-Borneol can increase

drug delivery across various physiological barriers (Kulkarni et al., 2021), 3-Carene has sleep-enhancing effects by targeting the GABAA-benzodiazepine receptors (Woo et al., 2019), L-α-Terpineol displays anticancer, anticonvulsant, antiulcer, antihypertensive, anti-nociceptive properties (Khaleel et al., 2018) and β-Caryophyllene might become effective for the treatment of diabetes and associated complications (Hashiesh et al., 2020). Moreover, Guaia-6,9-diene, a rare compound with promising anti-SARS-CoV-2 properties, also greatly increased (Amparo et al., 2021).

Likewise, a decrease in cultivars A and D of compounds with relevant medicinal properties would probably affect the cultivars' medicinal potential. The concentration of some compounds which decreased in higher ploidy levels show promising antitumor activity, such as Guaiol and β-bisabolene (Yeo et al., 2016; Yang et al., 2023) and the ability to counteract the accumulation of Amyloid-beta in Alzheimer's disease, Fenchol (Razazan et al., 2021)

To the best of our knowledge, this is the first report applying a metabolomics approach to investigate *C. sativa* polyploids in their diploid, triploid and tetraploid status across different cultivars. Overall, the response to polyploidization on morphological traits aligns with the observations on other plant species, showing larger organs with increasing ploidy levels. Moreover, analysis of secondary metabolites revealed for cannabinoids a marked decrease across cultivars in the higher ploidy levels. On the other hand, volatiles were negatively affected in triploids, while a differential response based on the cultivars' genetic background was observed in tetraploids.

Our observations revealed that the effect of gene doubling appears to be more evident on minor compounds, making polyploidy a very promising tool to increase their concentrations. *C. sativa* presents at least 150 minor cannabinoids apart from THCA and CBDA (Hanuš et al., 2016) and many other secondary metabolites (Radwan et al., 2021) with medicinal potential (Pollastro et al., 2018) that were not considered in this study and could be affected in the tetraploid state. Further research on *C. sativa* tetraploids applying metabolomic approaches could reveal changes also in minor cannabinoids and other metabolite classes with potential biological activity and medicinal properties. Research on preclinical and animal models pointed out that *C. sativa* botanical preparations appear more effective than pure THC (Blasco-Benito et al., 2018) or CBD (Gallily et al., 2015). Yet it is not clear what compounds or combination of compounds are responsible for such increased activity. Given the shift observed in secondary metabolite quantity and diversity, follow-up tests on tetraploid *C. sativa* botanical preparations could be a first step to reveal what kind of changes polyploidization bring about in the plant's pharmacological potential. Moreover, the only CBDA dominant cultivar of the study showed the most severe changes in cannabinoid and volatile reduction in the higher ploidy levels. Follow-up studies with more cultivars might elucidate what kind of relationship, if present, there might be between the effects of polyploidization and the genetic background of the plants.

The research on *C. sativa* polyploids is still in its infancy, and with this study, we have just scratched the surface. More research applying genomics, transcriptomics, metabolomics and phytocannabinomics (Cerrato et al., 2021) could reveal significant changes induced by genome doubling and elucidate mechanisms by which secondary metabolites are influenced in *C. sativa* polyploids.

Overall, this study found that polyploidization of *C. sativa* is a suitable approach to improve its medicinal potential, while the response is cultivar and genotype-dependent. The findings and methodology of this research lay the ground for further improving, evaluating and harnessing *C. sativa* chemical diversity by the breeding, biotechnological and pharmaceutical sectors.

## Data availability statement

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

## Author contributions

HPF: Headspace analysis, data acquisition, multivariate analysis, manuscript writing, review & editing. YHC: Supervising chemical analysis & manuscript review. KV: Conceptualization, project administration, formal analysis, supervision, manuscript writing, review & editing. MB: Conceptualization. BS: Plant growth, cannabinoid analysis, ploidy analysis, data acquisition, statistical analysis, formal analysis, manuscript writing, review & editing. FT: Conceptualization, supervision, methodology, plant growth, ploidy analysis, cannabinoid analysis, data acquisition, statistical analysis, formal analysis, investigation, manuscript writing, review & editing. All authors contributed to the article and approved the submitted version.

## Funding

The authors declare that this study received funding from F1 SeedTech. The funder provided the seeds of the cultivars studied. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

## Acknowledgments

We thank Lina Bayona Maldonado and Hye Kyong Kim for help with the method development of the cannabinoids analysis and F1 SeedTech for the financial support of the project.

## Conflict of interest

'Author MB is employed by the company F1 SeedTech.

The remaining authors declare that the research was conductedin the absence of any commercial or financial

relationships thatcould be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1233191/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The picture shows a side branch of a diploid (left side) and tetraploid (right side) genotypes of cultivar A.

SUPPLEMENTARY FIGURE 2
Principal component analysis PC1 against PC2, using UV scaling preprocessing data, of 4 *Cannabis sativa* L. cultivars A–D **(A)** and three ploidy levels 2n, 3n and 4n **(B)**. The PCA is based on 248 identified compounds. The explained variation is $R^2X$= 0.28.

## References


Amparo, T. R., Seibert, J. B., Silveira, B. M., Costa, F. S. F., Almeida, T. C., Braga, S. F. P., et al. (2021). Brazilian essential oils as source for the discovery of new anti-COVID-19 drug: a review guided by in silico study. *Phytochem. Rev.* 20, 1013–1032. doi: 10.1007/S11101-021-09754-4

Andre, C. M., Hausman, J. F., and Guerriero, G. (2016). Cannabis sativa: The plant of the thousand and one molecules. *Front Plant Sci.* 7, 19. doi: 10.3389/fpls.2016.00019

Bennett, M. D., and Leitch, I. J. (2005). Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Ann. Bot.* 95, 45–90. doi: 10.1093/AOB/MCI003

Blasco-Benito, S., Seijo-Vila, M., Caro-Villalobos, M., Tundidor, I., Andradas, C., García-Taboada, E., et al. (2018). Appraising the "entourage effect": Antitumor action of a pure cannabinoid versus a botanical drug preparation in preclinical models of breast cancer. *Biochem. Pharmacol.* 157, 285–293. doi: 10.1016/J.BCP.2018.06.025

Carputo, D., Frusciante, L., and Peloquin, S. J. (2003). The role of 2n gametes and endosperm balance number in the origin and evolution of polyploids in the tuber-bearing Solanums. *Genetics* 163, 287–294. doi: 10.1093/GENETICS/163.1.287

Cerrato, A., Citti, C., Cannazza, G., Capriotti, A. L., Cavaliere, C., Grassi, G., et al. (2021). Phytocannabinomics: Untargeted metabolomics as a tool for cannabis chemovar differentiation. *Talanta* 230, 122313. doi: 10.1016/J.TALANTA.2021.122313

Clarke, R. C., and Merlin, M. D. (2016). Cannabis domestication, breeding history, present-day genetic diversity, and future prospects. *CRC Crit. Rev. Plant Sci.* 35, 293–327. doi: 10.1080/07352689.2016.1267498

Crawford, S., Rojas, B. M., Crawford, E., Otten, M., Schoenenberger, T. A., Garfinkel, A. R., et al. (2021)Characteristics of the diploid, triploid, and tetraploid versions of a cannabigerol-dominant F1 hybrid industrial hemp cultivar, *Cannabis sativa* "Stem cell CBG." *Genes (Basel)* 12, 923 doi: 10.3390/GENES12060923

Eckardt, N. A. (2010). A double lock on polyploidy-associated epigenetic gene silencing. *Plant Cell* 22, 3. doi: 10.1105/TPC.110.220112

Edger, P. P., McKain, M. R., Yocca, A. E., Knapp, S. J., Qiao, Q., and Zhang, T. (2020). Reply to: Revisiting the origin of octoploid strawberry. *Nat. Genet.* 52, 1 52, 5–7. doi: 10.1038/s41588-019-0544-2

Gallily, R., Yekhtin, Z., and Hanuš, L. (2015). Overcoming the bell-shaped dose-response of cannabidiol by using *Cannabis* extract enriched in cannabidiol. *Pharmacol. Pharm.* 6, 75–85. doi: 10.4236/pp.2015.62010

Gul, W., Gul, S. W., Radwan, M. M., Wanas, A. S., Mehmedic, Z., Khan, I. I., et al. (2015). Determination of 11 cannabinoids in biomass and extracts of different varieties of cannabis using high-performance liquid chromatography. *J. AOAC Int.* 98 (6), 1523–1528. doi: 10.5740/jaoacint.15-095

Hanuš, L. O., Meyer, S. M., Muñoz, E., Taglialatela-Scafati, O., and Appendino, G. (2016). Phytocannabinoids: a unified critical inventory. *Nat. Prod Rep.* 33 (12), 1357–1392. doi: 10.1039/c6np00074f

Hashiesh, H. M., Nagoor Meeran, M. F., Sharma, C., Sadek, B., Kaabi, J., and Al, Ojha, S. K. (2020). Therapeutic potential of β-caryophyllene: A dietary cannabinoid in diabetes and associated complications. *Nutrients* 12, 1–30. doi: 10.3390/NU12102963

Hesami, M., Pepe, M., Baiton, A., Salami, S. A., and Jones, A. M. P. (2022). New insight into ornamental applications of cannabis: perspectives and challenges. *Plants (Basel).* 11 (18), 2383. doi: 10.3390/plants11182383

Hopping, M. E. (1994). Flow cytometric analysis of *Actinidia* species. *New Z. J. Bot.* 32 (1), 85–93. doi: 10.1080/0028825X.1994.10410410

Iannicelli, J., Guariniello, J., Tossi, V. E., Regalado, J. J., Di Ciaccio, L., van Baren, C. M., et al. (2020). The "polyploid effect" in the breeding of aromatic and medicinal species. *Scientia Hortic.* 260, 108854. doi: 10.1016/j.scienta.2019.108854

Kaensaksiri, T., Soontornchainaksaeng, P., Soonthornchareonnon, N., and Prathanturarug, S. (2011). In vitro induction of polyploidy in *Centella asiatica* (L.) Urban. *Plant Cell Tissue Organ Cult.* 107, 187–194. doi: 10.1007/s11240-011-9969-8

Karche, T., and Singh, M. R. (2019). The application of hemp (*Cannabis sativa* L.) for a green economy: a review. *Turkish J. Bot.* 43, 6. doi: 10.3906/bot-1907-15

Khaleel, C., Tabanca, N., and Buchbauer, G. (2018). α-Terpineol, a natural monoterpene: A review of its biological properties. *Open Chem.* 16, 349–361. doi: 10.1515/CHEM-2018-0040/ASSET/GRAPHIC/J_CHEM-2018-0040_FIG_002.JPG

Krüger, M., van Eeden, T., and Beswa, D. (2022). *Cannabis sativa* cannabinoids as functional ingredients in snack foods-historical and developmental aspects. *Plants (Basel).* 11 (23), 3330. doi: 10.3390/plants11233330

Kulkarni, M., Sawant, N., Kolapkar, A., Huprikar, A., and Desai, N. (2021). Borneol: a promising monoterpenoid in enhancing drug delivery across various physiological barriers. *AAPS PharmSciTech* 22, 145. doi: 10.1208/S12249-021-01999-8

Kurtz, L. E., Brand, M. H., and Lubell-Brand, J. D. (2020). Production of tetraploid and triploid hemp. *HortScience* 55, 1703–1707. doi: 10.21273/HORTSCI15303-20

Lavania, U. C. (2005). Genomic and ploidy manipulation for enhanced production of phyto-pharmaceuticals. *Plant Genet. Resour.* 3, 170–177. doi: 10.1079/PGR200576

Lavania, U. C. (2013). Polyploidy, body size, and opportunities for genetic enhancement and fixation of heterozygosity in plants. *Nucleus (India)* 56, 1–6. doi: 10.1007/S13237-013-0075-7

Laverty, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., et al. (2019). A physical and genetic map of *Cannabis sativa* identifies extensive rearrangements at the THC/CBD *acid synthase* loci. *Genome Res.* 29, 146–156. doi: 10.1101/GR.242594.118

Lin, X., Zhou, Y., Zhang, J., Lu, X., Zhang, F., Shen, Q., et al. (2011). Enhancement of artemisinin content in tetraploid *Artemisia annua* plants by modulating the expression of genes in artemisinin biosynthetic pathway. *Biotechnol. Appl. Biochem.* 58, 50–57. doi: 10.1002/BAB.13

Madani, H., Escrich, A., Hosseini, B., Sanchez-Muñoz, R., Khojasteh, A., and Palazon, J. (2021). Effect of polyploidy induction on natural metabolite production in medicinal plants. *Biomolecules* 11, 899. doi: 10.3390/BIOM11060899

Magagnini, G., Grassi, G., and Kotiranta, S. (2018). The effect of light spectrum on the morphology and cannabinoid content of Cannabis sativa L. *Med. Cannabis Cannabinoids.* 1 (1), 19–27. doi: 10.1159/000489030

Mansouri, H., and Bagheri, M. (2017). "Induction of polyploidy and its effect on *Cannabis sativa* L." in *Cannabis sativa L. - Botany and biotechnology*. Eds. (Cham: Springer), 365–383. doi: 10.1007/978-3-319-54564-6_17

Mishra, B. K., Pathak, S., Sharma, A., Trivedi, P. K., and Shukla, S. (2010). Modulated gene expression in newly synthesized auto-tetraploid of *Papaver somniferum* L. *South Afr. J. Bot.* 76, 447–452. doi: 10.1016/J.SAJB.2010.02.090

Motosugi, H., Okudo, K., Kataoka, D., and Naruo, T. (2002). Comparison of growth characteristics between diploid and colchicine-induced tetraploid grape rootstocks. *J. Japanese Soc. Hortic. Sci.* 71, 335–341. doi: 10.2503/JJSHS.71.335

Parsons, J. L., Martin, S. L., James, T., Golenia, G., Boudko, E. A., and Hepworth, S. R. (2019). Polyploidization for the genetic improvement of *Cannabis sativa*. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00476

Pollastro, F., Minassi, A., and Fresu, L. G. (2018). Cannabis phenolics and their bioactivities. *Curr. Med. Chem.* 25 (10), 1160–1185. doi: 10.2174/0929867324666170810164636

Radwan, M. M., Chandra, S., Gul, S., and Elsohly, M. A. (2021). Cannabinoids, phenolics, terpenes and alkaloids of *Cannabis*. *Molecules* 26, 2774. doi: 10.3390/MOLECULES26092774

Razazan, A., Karunakar, P., Mishra, S. P., Sharma, S., Miller, B., Jain, S., et al. (2021). Activation of microbiota sensing – free fatty acid receptor 2 signaling ameliorates amyloid-β Induced neurotoxicity by modulating proteolysis-senescence axis. *Front. Aging Neurosci.* 13. doi: 10.3389/FNAGI.2021.735933/BIBTEX

Russo, E. B. (2011). Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *Br. J. Pharmacol.* 163, 1344. doi: 10.1111/J.1476-5381.2011.01238.X

Sattler, M. C., Carvalho, C. R., and Clarindo, W. R. (2016). The polyploidy and its key role in plant breeding. *Planta* 243, 281–296. doi: 10.1007/S00425-015-2450-X

Shimadzu Scientific Instruments (2016) *Simplified cannabis terpene profiling by GCMS*. Available at: https://www.ssi.shimadzu.com/sites/ssi.shimadzu.com/files/pim/pim_document_file/ssi/others/14371/GCMS-1604-TerpeneProfilingCannabis.pdf.

Silvarolla, M. B., Mazzafera, P., Lima, M. M. A., de, Medina Filho, H. P., and Fazuoli, L. C. (1999). Ploidy level and caffeine content in leaves of Coffea. *Sci. Agric.* 56, 661–663. doi: 10.1590/S0103-90161999000300021

Simmonds, N. (1980). Polyploidy in plant breeding. *Span* 23 (2), 73–75.

Small, E. (2018). Dwarf germplasm: the key to giant *Cannabis* hempseed and cannabinoid crops. *Genet. Resour Crop Evol.* 65, 1071–1107. doi: 10.1007/s10722-017-0597-y

Song, Q., and Chen, J. Z. (2015). Epigenetic and developmental regulation in plant polyploids. *Curr. Opin. Plant Biol.* 24, 101–109. doi: 10.1016/J.PBI.2015.02.007

Tan, F. Q., Tu, H., Wang, R., Wu, X. M., Xie, K. D., Chen, J. J., et al. (2017). Metabolic adaptation following genome doubling in citrus doubled diploids revealed by non-targeted metabolomics. *Metabolomics* 13, 143. doi: 10.1007/S11306-017-1276-X

Trojak-Goluch, A., Kawka-Lipińska, M., Wielgusz, K., and Praczyk, M. (2021). Polyploidy in industrial crops: applications and perspectives in plant breeding. *Agron. 2021* 11, 2574. doi: 10.3390/AGRONOMY11122574

Trojak-Goluch, A., and Skomra, U. (2013). Artificially induced polyploidization in *Humulus lupulus* L. and its effect on morphological and chemical traits. *Breed Sci.* 63 (4), 393–399. doi: 10.1270/jsbbs.63.393

Urwin, N. A. R. (2014). Generation and characterisation of colchicine-induced polyploid *Lavandula x intermedia*. *Euphytica* 197, 331–339. doi: 10.1007/S10681-014-1069-5

Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi: 10.1016/J.GDE.2018.01.004

Wohlmuth, H., Leach, D. N., Smith, M. K., and Myers, S. P. (2005). Gingerol content of diploid and tetraploid clones of ginger (*Zingiber officinale* Roscoe). *J. Agric. Food Chem.* 53, 5772–5778. doi: 10.1021/JF050435B

Woo, J., Yang, H., Yoon, M., Gadhe, C. G., Pae, A. N., Cho, S., et al. (2019). 3-carene, a phytoncide from pine tree has a sleep-enhancing effect by targeting the GABAA-benzodiazepine receptors. *Exp. Neurobiol.* 28, 593. doi: 10.5607/EN.2019.28.5.593

Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., and Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13875–13879. doi: 10.1073/PNAS.0811575106

Xu, C., Tang, Y., Chen, R., Liang, C., Liu, X., Wu, C., et al. (2014). A comparative study of bioactive secondary metabolite production in diploid and tetraploid *Echinacea purpurea* (L.) Moench. *Plant Cell Tissue Organ Cult.* 116, 323–332. doi: 10.1007/s11240-013-0406-z

Yang, X., Yang, J., Gu, X., Tao, Y., Ji, H., Miao, X., et al. (2023). (-)-Guaiol triggers immunogenic cell death and inhibits tumor growth in non-small cell lung cancer. *Mol. Cell Biochem.* 478, 1611–1620. doi: 10.1007/S11010-022-04613-Y

Yeo, S. K., Ali, A. Y., Hayward, O. A., Turnham, D., Jackson, T., Bowen, I. D., et al. (2016). β-bisabolene, a sesquiterpene from the essential oil extract of opoponax (*Commiphora guidottii*), exhibits cytotoxicity in breast cancer cell lines. *Phytother. Res.* 30, 418–425. doi: 10.1002/PTR.5543

Zonneveld, B. J. M., and Van Iren, F. (2001). Genome size and pollen viability as taxonomic criteria: Application to the genus *Hosta*. *Plant Biol.* 3, 176–185. doi: 10.1055/S-2001-12900

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 September 2023
DOI 10.3389/fpls.2023.1228961



OPEN ACCESS

EDITED BY
Kashif Ali,
Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Pakistan

REVIEWED BY
William Bryan Terzaghi,
Wilkes University, United States
Pengbo Xu,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Christian Schulze Gronover
christian.schulze.gronover@ime.fraunhofer.de

†These authors have contributed equally to this work

RECEIVED 30 May 2023
ACCEPTED 21 August 2023
PUBLISHED 28 September 2023

CITATION
Wolters SM, Benninghaus VA, Roelfs K-U, van Deenen N, Twyman RM, Prüfer D and Schulze Gronover C (2023) Overexpression of a pseudo-etiolated-in-light-like protein in *Taraxacum koksaghyz* leads to a pale green phenotype and enables transcriptome-based network analysis of photomorphogenesis and isoprenoid biosynthesis.
*Front. Plant Sci.* 14:1228961.
doi: 10.3389/fpls.2023.1228961



# Overexpression of a pseudo-etiolated-in-light-like protein in *Taraxacum koksaghyz* leads to a pale green phenotype and enables transcriptome-based network analysis of photomorphogenesis and isoprenoid biosynthesis

Silva Melissa Wolters[1†], Vincent Alexander Benninghaus[1†], Kai-Uwe Roelfs[1], Nicole van Deenen[2], Richard M. Twyman[3], Dirk Prüfer[1,2] and Christian Schulze Gronover[1*]

[1]Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Münster, Germany, [2]Institute for Biology and Biotechnology of Plants, University of Münster, Münster, Germany, [3]TRM Ltd, Scarborough, United Kingdom

**Introduction:** Plant growth and greening in response to light require the synthesis of photosynthetic pigments such as chlorophylls and carotenoids, which are derived from isoprenoid precursors. In *Arabidopsis*, the pseudo-etiolated-in-light phenotype is caused by the overexpression of *repressor of photosynthetic genes 2* (*RPGE2*), which regulates chlorophyll synthesis and photosynthetic genes.

**Methods:** We investigated a homologous protein in the Russian dandelion (*Taraxacum koksaghyz*) to determine its influence on the rich isoprenoid network in this species, using a combination of *in silico* analysis, gene overexpression, transcriptomics and metabolic profiling.

**Results:** Homology-based screening revealed a gene designated *pseudo-etiolated-in-light-like* (*TkPEL-like*), and *in silico* analysis identified a light-responsive G-box element in its promoter. *TkPEL-like* overexpression in dandelion plants and other systems reduced the levels of chlorophylls and carotenoids, but this was ameliorated by the mutation of one or both conserved cysteine residues. Comparative transcriptomics in dandelions overexpressing *TkPEL-like* showed that genes responsible for the synthesis of isoprenoid precursors and chlorophyll were downregulated, probably explaining the observed pale green leaf phenotype. In contrast, genes responsible for

carotenoid synthesis were upregulated, possibly in response to feedback signaling. The evaluation of additional differentially expressed genes revealed interactions between pathways.

**Discussion:** We propose that TkPEL-like negatively regulates chlorophyll- and photosynthesis-related genes in a light-dependent manner, which appears to be conserved across species. Our data will inform future studies addressing the regulation of leaf isoprenoid biosynthesis and photomorphogenesis and could be used in future breeding strategies to optimize selected plant isoprenoid profiles and generate suitable plant-based production platforms.

KEYWORDS

*Taraxacum*, chlorophyll biosynthesis, isoprenoids, light-dependent regulation, photomorphogenesis, pseudo-etiolation-in-light, RPGE

## 1 Introduction

Plants are photoautotrophic organisms that mostly depend on light for their energy supply. They also need to balance light exposure or mitigate excess irradiation to restrict photo-oxidative damage. Accordingly, they possess a complex network of light perception, signal transduction and effector proteins for the coordination of essential developmental processes such as photomorphogenesis and flowering (reviewed by Kami et al., 2010; Jing and Lin, 2020; Li et al., 2022). This starts with the perception of light signals by multiple families of photoreceptors that are translocated from the cytosol to the nucleus following activation (Genoud et al., 2008; Galvão and Fankhauser, 2015). In the nucleus, photoreceptors interact with E3 ubiquitin ligase complexes and transcription factors, which tune the activity of light-sensitive genes and thereby activate and coordinate downstream transcription cascades (Xu et al., 2014; Xu, 2019).

During photomorphogenesis, plants start to green and expand their leaves, requiring the synthesis of chlorophylls and other photosynthetic pigments. Chlorophyll and carotenoid biosynthesis requires isoprenoid precursors because the porphyrin ring of chlorophyll carries a phytol side chain and carotenoids are $C_{40}$ isoprenoids (Cunningham and Gantt, 1998; Beale, 1999). Isoprenoids, also known as terpenoids, are one of the largest and most structurally diverse classes of natural products, consisting of many primary and secondary metabolites with various roles in basic cellular processes (Croteau et al., 2000; Tholl, 2015; Yazaki et al., 2017). The central metabolic precursors for the biosynthesis of all isoprenoids are the $C_5$ isomers isopentenyl pyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP), both of which can be synthesized via two metabolic routes: the cytosolic mevalonate (MVA) pathway and the plastidial methylerythritol phosphate (MEP) pathway (Vranová et al., 2012; Liao et al., 2016; Frank and Groll, 2017). The MVA pathway mainly supplies precursors for isoprenoids produced in the cytosol and mitochondria, such as sterols and brassinosteroids, whereas precursors produced by the MEP pathway form plastidial isoprenoids such as chlorophylls, carotenoids and plastoquinone (Lichtenthaler, 1999; Vranová et al., 2012). Both pathways are embedded in a tightly regulated, light-dependent network that is distributed over several cellular compartments. The pathways are regulated at transcriptional, post-transcriptional, translational and post-translational levels, and via feedback as well as feedforward signaling. This allows the metabolic flux to be allocated between different downstream pathways under normal and challenging conditions (reviewed in detail by Hemmerlin et al., 2012; Vranová et al., 2012; Tholl, 2015). However, the extent to which light perception/signaling and isoprenoid biosynthesis overlap is unclear and more work is required to understand how this complex network is regulated.

A systematic gain-of-function mutant screen in *Arabidopsis thaliana* revealed a line with pale green leaves and rapid development, a phenotype designated PEL meaning 'pseudo-etiolation in light' (Ichikawa et al., 2006). The gene overexpressed in this line encoded "plant protein 1589 of unknown function" (At3g55240). Given that all enzymes in the MEP and chlorophyll biosynthesis pathways are already known, the PEL gene was predicted to be a negative regulator (Beale, 1999; Vranová et al., 2012). The overexpression phenotype was confirmed in a later study and the gene was named *repressor of photosynthetic genes 2* (*RPGE2*) based on the identification of its paralog *RPGE1* as an early red-light-responsive gene regulated by phytochrome interacting factors (PIFs), and the repression of photosynthetic genes in *RPGE1* and *RPGE2* overexpression lines (Leivar et al., 2009; Zhang et al., 2013; Kim K. et al., 2016). PIFs have been described as light-dependent negative transcriptional regulators of photomorphogenesis and the MEP pathway (Chenge-Espinosa et al., 2018), as well as chlorophyll and carotenoid biosynthesis (Huq et al., 2004; Moon et al., 2008; Toledo-Ortiz et al., 2010; Tang et al., 2012; Job and Datta, 2021). They are potential cofactors of COP1 (constitutive photomorphogenic 1), an E3 ubiquitin ligase that represses photomorphogenesis in the dark by accumulating in the nucleus and mediating the degradation of light-dependent positive regulators such as the bZIP transcription factor Long hypocotyl 5 (HY5) (Deng et al., 1991; Osterlund et al., 1999; Saijo et al., 2003; Jang et al., 2005; Xu et al., 2014). Very recent findings in *Arabidopsis* have shown that AtRPGE1 and AtRPGE2

suppress chlorophyll biosynthesis by inhibiting the transcription factor Golden2-like 1 (GLK1) (Kim et al., 2023), which activates genes involved in chlorophyll biosynthesis, light harvesting and electron transport (Waters et al., 2009).

Comparative transcriptomics in carrots (*Daucus carota*) with high and low pigment levels revealed an association between carotenoid accumulation in roots with one or two recessive mutations in the gene *DCAR_032551*, which is homologous to *AtRPGE2*, expanding the function of this protein to root tissues and carotenoid synthesis (Iorizzo et al., 2016). Furthermore, multiple light-induced genes and genes involved in the MEP pathway and carotenoid biosynthesis are upregulated in strongly pigmented carrots (Iorizzo et al., 2016). The function of the DCAR_032551 protein is unclear, but it was proposed to regulate carotenoid accumulation resulting from de-etiolation in roots that have lost the ability to repress the transcriptomic cascade connected to photomorphogenesis (Iorizzo et al., 2016).

The *AtRPGE2* and *DCAR_032551* genes encode interesting regulatory proteins in the light signaling network affecting isoprenoid metabolism, and the analysis of orthologs in other species could provide more insight into their mode of action and evolutionary origin. A plant ideally suited to analyze photomorphogenic isoprenoid metabolism is the Russian dandelion (*Taraxacum koksaghyz*), because it produces high levels of various isoprenoid compounds in different tissues. It is also a valuable source of economically important isoprenoids (van Beilen and Poirier, 2007), in particular due to the large amounts of high-quality natural rubber formed in the latex of its roots (Ulmann, 1951; van Beilen and Poirier, 2007; Ramirez-Cadavid et al., 2017). Research has focused on domesticating this wild plant to make a profitable and sustainable rubber crop, especially by investigating the transcriptomic, proteomic and metabolomic properties of its roots (Niephaus et al., 2019; Pütter et al., 2019; Benninghaus et al., 2020). Only recently has the scope of this work expanded to include the transcriptomic and metabolic profiles of the leaves (Cheng et al., 2022; Panara et al., 2022; Zhang et al., 2022).

To broaden our knowledge of the rich isoprenoid metabolic network in the Russian dandelion, we screened for homologs of AtRPGE2/DCAR_032551 to investigate their regulatory functions. We identified the *T. koksaghyz* gene *pseudo-etiolated-in-light-like* (*TkPEL-like*), which like its orthologs in *Arabidopsis* and carrot reduced leaf chlorophyll and carotenoid levels when overexpressed. The function of TkPEL-like depends on two conserved cysteine residues. Comparative transcriptomics in dandelions overexpressing *TkPEL-like* vs controls provided information about affected pathways and their interactions, and highlighted the suggested conservation of these properties across species.

# 2 Materials and methods

## 2.1 Cloning of *TkPEL-like* and vector construction

The TkPEL-like coding sequence was amplified by PCR from *T. koksaghyz* cDNA using primers TkPEL-like_fw and TkPEL-like_rev (Supplementary Table S1). It was purified using the PCR clean-up gel extraction kit (Macherey-Nagel, Germany) and inserted into the cloning vector pJET1.2/blunt using the CloneJET PCR Cloning Kit (Thermo Fisher Scientific, USA). For the *TkPEL-like* overexpression construct, the strong quadruple cauliflower mosaic virus promoter (pQ35S) was amplified from pFGC5941-GW (GenBank: DQ231581.1) using primers Q35S-P_XmaI_fw and Q35S-P_SOE_rev_XhoI_mut, and the TMV Ω 5′-leader sequence was amplified from the same plasmid using primers TMV_Omega_TL_SOE_fw_XhoI_mut and TMV_Omega_TL_XhoI_rev. The PCR products were diluted 1:100 and used as templates for overlap extension PCR with primers Q35S-P_XmaI_fw and TMV_Omega_TL_XhoI_rev. This disrupted the XhoI site between pQ35S and the TMV Ω 5′-leader, enabling the subsequent cloning step. The PCR product was purified and digested (XmaI + XhoI) and inserted into pLab12.1 (Post et al., 2012), which had been digested with the same enzymes. The *TkPEL-like* sequence was amplified from pJET-*TkPEL-like* using primers TkPEL-like_NcoI_fw and TkPEL-like_XhoI_rev. After purification and digestion (XhoI + XbaI), the *TkPEL-like* sequence was inserted into the XhoI/XbaI-digested custom vector for constitutive expression. For transient expression in *Nicotiana benthamiana*, the *TkPEL-like* sequence was amplified from pJET-*TkPEL-like* using forward primer TkPEL-like_NcoI_fw and either TkPEL-like_XhoI_rev or TkPEL-like_XhoI_rev_nsc, yielding fragments with and without a stop codon. These fragments were purified, digested (NcoI + XhoI) and inserted into the NcoI/XhoI-digested Gateway vector pENTR4 (Invitrogen, USA), resulting in the cloning vectors pENTR4-*TkPEL-like* and pENTR4-*TkPEL-like*_NSC (NSC = no stop codon). Site-directed mutagenesis was carried out with primer pairs TkPEL-like_muta.t55a and TkPEL-like_muta.t55a_antis or TkPEL-like_muta.t85a and TkPEL-like_muta.t85a_antis using the QuikChange Lightning Site-Directed Mutagenesis Kit and the publicly available QuikChange Primer Design Program (Agilent Technologies, USA), leading to the nucleotide substitutions *c.*55T>A and/or *c.*85T>A. This yielded the following cloning vectors: pENTR4-*TkPEL-like*_t55a, pENTR4-*TkPEL-like*_t85a, pENTR4-*TkPEL-like*_t55a/t85a, pENTR4-TkPEL-like_NSC_t55a, pENTR4-*TkPEL-like*_NSC_t85a and pENTR4-*TkPEL-like*_NSC_t55a/t85a. Finally, the mutated and non-mutated sequences were transferred to the Gateway-compatible vectors pBatTL-*ccdB*, pBatTL-*Cerulean-ccdB* and pBatTL-*ccdB-Cerulean* (Epping et al., 2015) by LR recombination using LR Clonase (Thermo Fisher Scientific), resulting in the final plant transformation vectors pBatTL-*TkPEL-like_t55a*, pBatTL-*TkPEL-like_t85a*, pBatTL-*TkPEL-like_t55a/t85a*, pBatTL-*Cerulean-TkPEL-like_t55a*, pBatTL-*Cerulean-TkPEL-like_t85a*, pBatTL-*Cerulean-TkPEL-like_t55a/t85a*, pBatTL-*TkPEL-like_t55a-Cerulean*, pBatTL-*TkPEL-like_t85a-Cerulean* and pBatTL-*TkPEL-like_t55a/t85a-Cerulean*. All constructs were sequenced to confirm their integrity.

## 2.2 Plant cultivation and tissue processing

All *T. koksaghyz* plants were cultivated under controlled conditions in an indoor greenhouse (18°C, 16-h photoperiod, 260 PPFD high-pressure sodium lamp with enhanced yellow and red spectrum) as previously described (Unland et al., 2018). Plant

tissues were separated during harvesting, then immediately flash-frozen in liquid nitrogen and lyophilized. The root tissue was pulverized using a ZM 200 Ultra Centrifugal Mill (Retsch, Germany), whereas petiole, leaf and flower tissues were pulverized under liquid nitrogen with mortar and pestle.

## 2.3 Generation of pQ35S::*TkPEL-like T. koksaghyz* plants and transgene verification

The *TkPEL-like* overexpression construct was introduced into *T. koksaghyz* plants (accession number 203; kindly provided by the Botanical Garden of the University of Münster, Germany) as previously described (Stolze et al., 2017). The resulting transgenic plants were selected by cultivation on phosphinothricin-containing medium. The presence of the transgenes was verified in crude leaf extracts by PCR with gene-specific primers (Supplementary Table S1) using the KAPA3G Plant PCR Kit (Kapa Biosystems, USA).

## 2.4 RNA extraction and cDNA synthesis

Total RNA was extracted from *T. koksaghyz* root, latex, leaf, petiole and flower tissues using the innuPREP RNA Mini Kit (Analytik Jena, Germany) according to the manufacturer's instructions. To synthesize full-length cDNA, 500 ng of total RNA was reverse transcribed using PrimeScript RT Master Mix (TAKARA Clontech, USA) according to the manufacturer's instructions.

## 2.5 Gene expression analysis by qRT-PCR

Endogenous spatiotemporal expression patterns were determined by quantitative real-time PCR (qRT-PCR) as previously described (Laibach et al., 2015) using RNA from the root, latex, petiole, leaf without midrib and flower tissues of 12-week-old *T. koksaghyz* wild-type plants grown outdoors from May to August 2017 (University of Münster, Germany) or leaf RNA derived from wild-type plants aged 3, 4, 5, 6, 8, 10 or 12 weeks grown under controlled greenhouse conditions as above. For the spatial expression pattern, three pools consisting of four individual plants as biological replicates were tested for each tissue. For the temporal expression pattern, three independent wild-type plants as biological replicates were tested for each time point. *TkPEL-like* overexpression in transgenic plants was assessed for eight individual plants from each of the lines #3.4, #4.1 and #5.7 in comparison to nine individual plants of near-isogenic lines (NILs). For the validation of the transcriptomic data, leaf RNA from three pools each consisting of four plants was analyzed for the transgenic lines #3.4, #4.1 and #5.7. In comparison, three pools consisting of leaf RNA from two to three plants each was analyzed for the corresponding NILs (NIL#3.4, NIL #4.1, NIL #5.7). All samples were additionally analyzed in three technical replicates. We used *elongation factor 1a* (*TkEf1a*) and *ribosomal protein L27* (*TkRP*) to calculate the relative normalized expression levels (Pütter et al., 2017) using Bio-Rad CFX Manager v3.1 (Bio-Rad Laboratories, USA). All oligonucleotides are listed in Supplementary Table S1.

## 2.6 Infiltration of *N. benthamiana* leaves and microscopy

Transient expression in 4-week-old *N. benthamiana* leaves was achieved by agroinfiltration (Müller et al., 2010). *N. benthamiana* seeds (kindly provided by the Sainsbury Laboratory, John Innes Centre, Norwich, UK) were germinated and the plants were cultivated under controlled greenhouse conditions (Unland et al., 2018). After infiltration, the plants were grown for 4–5 days under steady light in a growth chamber (CLF Plant Climatics, Germany) at 19°C. The subcellular localization of recombinant proteins in epidermal cells was determined by screening tissue explants from infiltrated leaves for Cerulean fluorescence using the Leica TCS SP5 X confocal system (Leica Microsystems, Germany) with excitation at 405/458 nm and emission at 469–488 nm.

## 2.7 Quantification of carotenoids and chlorophylls in leaves

Leaf material was harvested, immediately flash-frozen in liquid nitrogen, lyophilized in a freeze dryer Alpha 1-4 LSCplus (Christ Gefriertrocknungsanlagen, Germany) and cryogenically ground in a Retsch mixer mill MM 400. We then extracted 10 mg of the pulverized leaf powder twice with 500 µl acetone (30 min, room temperature, on a shaking platform in the dark) and removed the cell debris by centrifugation (1000 g, 5 min, room temperature). The final 1 ml acetone leaf extract was diluted 1:10 with acetone and analyzed by absorption spectrophotometry at 470, 645, 662 and 750 nm using a quartz cuvette and a BioSpectrometer (Eppendorf, Germany). The carotenoid and chlorophyll contents were calculated as previously described (Lichtenthaler, 1987).

## 2.8 Quantification of root metabolites

Poly(*cis*-1,4-isoprene), squalene/2,3-oxidosqualene and pentacyclic triterpenes/triterpenoids were measured in pulverized root extracts by $^{1}$H-NMR and GC-MS (Stolze et al., 2017).

## 2.9 Chlorophyll fluorescence measurements

The maximum potential quantum efficiency of photosystem II (PSII) and non-photochemical quenching (NPQ) were measured in a single leaf from 12-week-old transgenic *T. koksaghyz* plants and controls using a Maxi-Imaging Pulse-Amplitude-Modulation (PAM) chlorophyll fluorimeter (Walz, Germany). Before measurement, the leaves were dark adapted for 20 min, exposed to photosynthetically active radiation with an intensity of 1076 µmol $m^{-2}$ $s^{-1}$ for 5 min, followed by relaxation in the dark for ~15

min. Fluorescence measurements were captured from 14 independent areas per leaf. The photosynthetic parameters, described in detail by Baker (2008), were calculated using ImagingWin v2.41a (Walz). NPQ was calculated using Equation 1, the maximum quantum yield of PSII after dark adaption ($F_v/F_m$) using Equation 2, and the operating efficiency of PSII under illumination ($F_q'/F_m'$ or $\Phi_{PSII}$) using Equation 3.

Equation 1: $NPQ = \frac{Fm-Fm'}{Fm'}$

Equation 2: $\frac{F_v}{F_m} = \frac{F_m-F0}{F_m}$

Equation 3: $\frac{F_q'}{F_m'} = \frac{F_m'-F'}{F_m'}$,

## 2.10 *In silico* analysis

The detection of distant *TkPEL-like* homologs was achieved by screening the non-redundant NCBI peptide database (NR, http://www.ncbi.nlm.nih.gov). Multiple sequence alignment was then carried out using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and MegAlign Pro 17 (DNASTAR, USA). MOTIF (https://www.genome.jp/tools/motif/) was used to screen the NCBI-CDD library (Marchler-Bauer et al., 2013), as well as the Pfam (Bateman et al., 2002) and PROSITE (Falquet et al., 2002) databases for matching protein domains. Evolutionarily conserved amino acids were visualized using the ConSurf server (http://consurf.tau.ac.il; Ashkenazy et al., 2016), and secondary structures were predicted using the JPred 4 server (Drozdetskiy et al., 2015). The subcellular localization of proteins was predicted using DeepLoc (Armenteros et al., 2017; Armenteros et al., 2019). The *TkPEL-like* promoter was analyzed using NSITE-PL (Shahmuradov and Solovyev, 2015).

## 2.11 Transcriptomic comparison

RNA was extracted from the leaf material of four plants from each pQ35S::*TkPEL-like* line (#3.4, #4.1 and #5.7) and four plants of the corresponding NILs. The RNA obtained from four individual plants belonging to one line was pooled in equal amounts. The three RNA pools per genotype were then pooled again in equal amounts resulting in two RNA pools consisting of 12 plants each that were used for Illumina 150-bp paired-end sequencing (Novogene, USA). Reads were mapped to the *T. koksaghyz* reference genome (Lin et al., 2022) using Subread v2.0.3 and raw counts were determined using Feature Counts. Transcripts were annotated by conducting a BLASTX search against the NCBI NR and UniProt databases with a minimum e-value of $1 \times 10^{-3}$. Differential gene expression was analyzed using DEGseq v1.50.0. Datasets were filtered for minimum counts per million (CPM) of 0.2 and differential expression was assessed by comparing CPMs in NILs and pQ35S::*TkPEL-like* plants. Differentially expressed genes (DEGs) were accepted if they met the threshold for $\log_2$ fold changes – $\log_2$(FC) – in normalized expression ≥ 1 or ≤ –1 and a q-value (Storey) < 0.05 (Supplementary Data S1). Gene Ontology (GO) term enrichment among modulated genes ($-2.32 \leq \log_2(FC) \geq 2.32$) was calculated using WEGO (https://wego.genomics.cn/) with all transcripts detected in NILs as the background. Genes were also annotated against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using eggNOG-mapper v2 (http://eggnog-mapper.embl.de/). KEGG pathway enrichment analysis was carried out using the enricher tool of the clusterProfiler package in R Studio (Posit PBC, USA).

# 3 Results

## 3.1 Identification and *in silico* characterization of *TkPEL-like*

A homology search against *T. koksaghyz* sequence databases identified Contig19977 (length 751 bp, e-value $2 \times 10^{-29}$) with a translated sequence similar to DCAR_032551. Amplification with flanking primers yielded a shorter, 327-bp coding sequence, which we named *TkPEL-like*. A subsequent, less stringent BLASTP search with the obtained 108 aa sequence identified homologous protein sequences to be most prominent in clade Streptophyta (Supplementary Figure S1). They included only a few annotated proteins: a DNA replication complex GINS protein, a putative angiotensin-converting enzyme 2 and a polyribonucleotide nucleotidyltransferase. Nucleotide-based homology searching revealed an additional match to an argininosuccinate lyase (Supplementary Table S2). Multiple sequence alignment of the TkPEL-like protein, DCAR_032551 and AtRPGE2 showed that the 57 N-terminal amino acids are strongly conserved (Figure 1A), with three overlapping domains: 'A_thal_3526' (CDD: 401595) and 'TIGR01589: A_thal_3526' (CDD: 130650) are only found in plant proteins but their functions are unknown, whereas the PIN_Fcf1-like domain (CDD: 350212) is present in a large nuclease superfamily involved in RNA processing (Figure 1D). The N-terminal sequence also contains two highly conserved cysteines that are predicted to act as structural residues (arrows in Figure 1B). Secondary structure prediction revealed five potential α-helices (H, marked in green) distributed throughout the peptide sequence (Figure 1D). Additionally, the protein contains two putative *N*-glycan accepter sites (AA90-93, 105-108) and two phosphorylation sites (AA25-28,75-78). Like its carrot and *Arabidopsis* counterparts, TkPEL-like was predicted to be a cytosolic protein (Supplementary Figure S1). Given that the *AtRPGE2* promoter contains sequences recognized by PIFs (Leivar et al., 2009; Zhang et al., 2013; Kim K. et al., 2016) and the bZIP transcription factor HY5, a key regulator of light-mediated transcriptional programming (Lee et al., 2007), we also analyzed the *TkPEL-like* promoter for the corresponding binding sites. First, we searched the *T. koksaghyz* genome assembly (Lin et al., 2022) for homology to the 327-bp amplified gene sequence to identify the promoter region, revealing two matching loci with 95–99% identity. We then screened 1 kb upstream of each coding sequence to identify *cis*-acting regulatory motifs. We found 20 motifs for GWHGBCHF020456 on pseudochromosome four, and 21 for GWHGBCHF024124 on pseudochromosome five, including a G-box motif with the core sequence CACGTG. This is known to

10.3389/fpls.2023.1228961

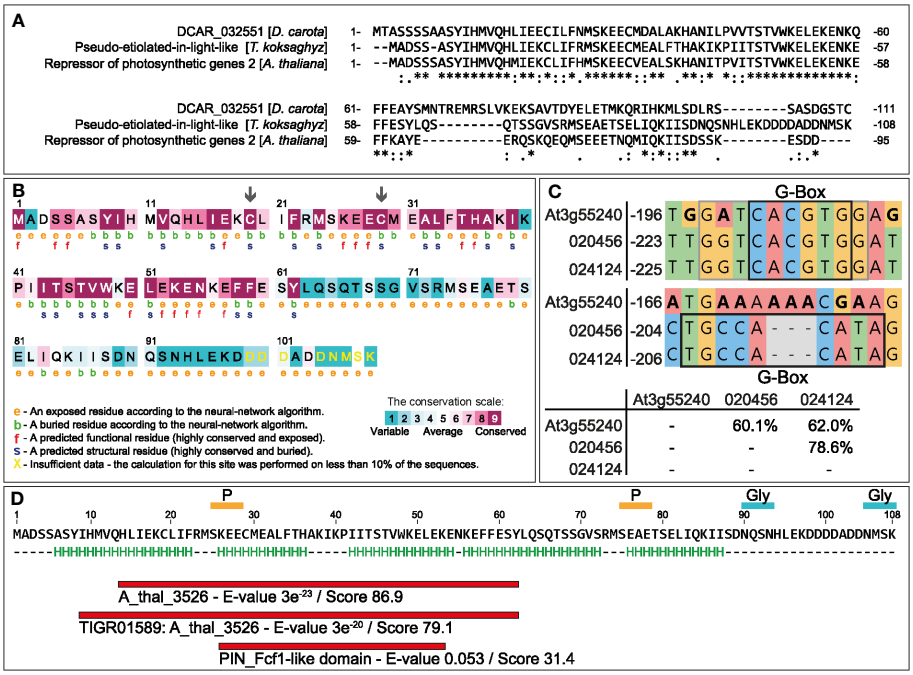



FIGURE 1
*In silico* analysis of the TkPEL-like protein and the corresponding promoter. **(A)** Multiple sequence alignment of TkPEL-like, DCAR_032551 from *D. carota* (GenBank: KZM94650.1) and AtRPGE2 (At3g55240). * (asterisk) indicates positions which have a single, fully conserved residue; : (colon) indicates conservation between groups of strongly similar properties - roughly equivalent to scoring > 0.5 in the Gonnet PAM 250 matrix; . (period) indicates conservation between groups of weakly similar properties - roughly equivalent to scoring ≤ 0.5 and > 0 in the Gonnet PAM 250 matrix. **(B)** Identification of functional and conserved regions within the TkPEL-like protein using ConSurf based on multiple sequence alignments. **(C)** Promoter sequence comparisons between two *TkPEL-like* loci in *T. koksaghyz* and *AtRPGE2* based on multiple sequence alignments. IDs refer to gene IDs of the *T. koksaghyz* genome (Lin et al., 2022). Black boxes indicate G-box core sequences. Gray box indicates variable G-box extensions. **(D)** TkPEL-like amino acid sequence with proposed helix structures (H) and predicted conserved protein domains identified by Motiffinder.

mediate light-responsive transcriptional regulation by binding to HY5 and PIFs (Toledo-Ortiz et al., 2003; Lee et al., 2007; Kim J. et al., 2016) (Figure 1C; Supplementary Table S3). This motif also shows similarity to the light-dependent regulatory element ACE$^{CHS}$, which is present in the *chalcone synthase* (*CHS*) promoter and bound by the bZIP transcription factors CPRF1 and CPRF4 to regulate flavonoid biosynthesis (Kircher et al., 1998; Sprenger-Haussels and Weisshaar, 2000).

## 3.2 *TkPEL-like* is predominantly expressed in the leaves of adult *T. koksaghyz* plants

To examine the spatial expression profile of *TkPEL-like*, we used cDNA obtained from latex, root, leaf, petiole and flower tissues of 12-week-old *T. koksaghyz* plants for qRT-PCR analysis. This revealed predominant expression in leaves (Figure 2A). We then evaluated the temporal expression profile of *TkPEL-like* in leaves,

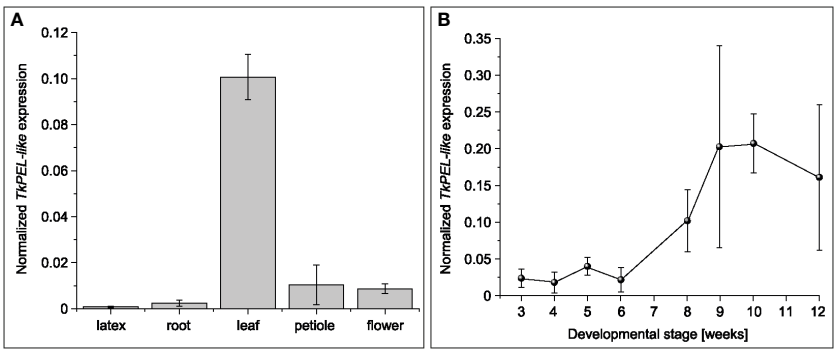



FIGURE 2
*TkPEL-like* is predominantly expressed in the mature leaves of wild-type *T. koksaghyz* plants. **(A)** Normalized transcript levels of *TkPEL*-like in different tissues of wild-type *T. koksaghyz*. Data are means (± SD) of three pools of four individual plants. **(B)** Time course of *TkPEL-like* expression in wild-type *T. koksaghyz* leaves. Plants were grown under controlled greenhouse conditions. Expression levels were normalized against *elongation factor 1a* (*TkEf1a*) and *ribosomal protein L27* (*TkRP*). Data are means (± SD) of three independent plants.

which indicated low levels of gene expression during the growth stages (weeks 3–6) followed by a ~5-fold increase in weeks 8–12 (Figure 2B).

## 3.3 Heterologous expression of *TkPEL-like* in *N. benthamiana* reveals cytosolic and nuclear protein localization and reduced levels of chlorophylls and carotenoids

We generated expression constructs in which the fluorophore Cerulean was fused in-frame with the TkPEL-like coding sequence at the N-terminus or C-terminus in order to investigate its intracellular localization *in vivo*. Transient expression in *N. benthamiana* leaf epidermal cells was monitored by confocal laser scanning microscopy, showing Cerulean fluorescence signals in the cytosol and nucleus for TkPEL-like fusions as well as Cerulean itself (Figure 3B). This suggests TkPEL-like is localized in these compartments, although we cannot exclude the possibility that the spatial distribution was due to the small size of the fusion protein. Leaves transiently expressing native *TkPEL-like* or the *Cerulean* fusions showed visible signs of chlorosis after 4 days (Figure 3A). We therefore measured the pigment content of the leaves and compared them with non-infiltrated leaves and those infiltrated with a *Cerulean* construct alone. We found that chlorophyll and carotenoid levels were significantly lower in leaves expressing *TkPEL-like* constructs compared to controls (Figures 3C, D; Supplementary Table S4).

To assess the role of the two conserved cysteine residues in the highly conserved N-terminus, we introduced either single mutations (C19S or C29S) or a double mutation (C19S, C29S) in the fusion constructs. Microscopy revealed comparable cytosolic and nuclear fluorescence for the mutants and the wild-type variant (Supplementary Figure S2). This was supported by *in silico* secondary structure predictions, which showed that the five α-helices were preserved in the mutants, confirming no significant impact on the integrity of the protein's structure (Supplementary Figure S3). Pigment analysis of leaves expressing one of the three mutant *TkPEL-like* sequences alone or as *Cerulean* fusions again showed a reduction in chlorophyll and carotenoid levels (Figures 3C, D; Supplementary Table S4). However, the amounts were comparable to those detected in leaves expressing *Cerulean* alone. Accordingly, leaves infiltrated with the *TkPEL-like* wild-type constructs had significantly lower levels of photosynthetic pigments than leaves expressing the mutant variants or *Cerulean*, suggesting the conserved cysteine residues are important for the function of TkPEL-like.

## 3.4 Overexpression of *TkPEL-like* in *T. koksaghyz* causes a pale green phenotype

The ubiquitous overexpression of *TkPEL-like* in *T. koksaghyz* was achieved by placing the gene under the control of the pQ35S promoter (Figure 4A). Several transgenic lines were generated and crossed with wild-type *T. koksaghyz* to obtain the $T_1$ generation.

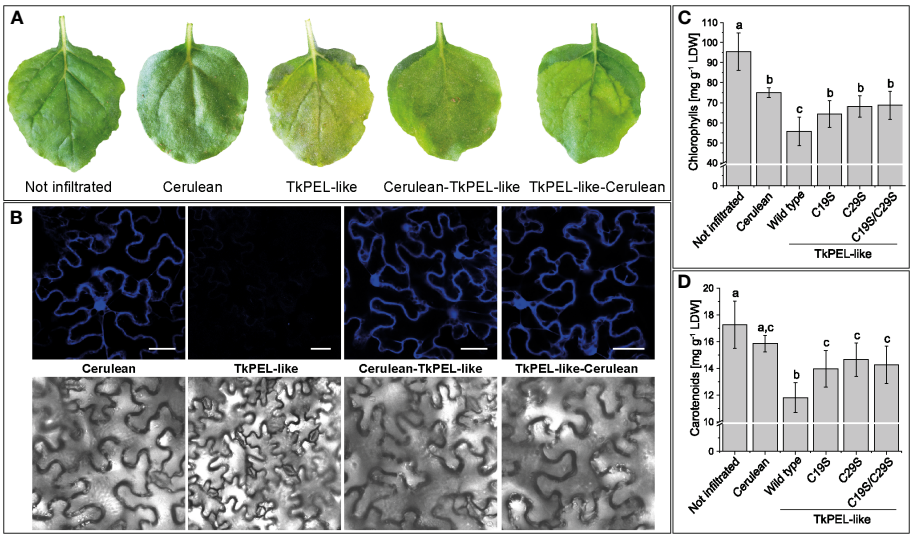



FIGURE 3
Heterologous expression of *TkPEL-like* in *N. benthamiana* leaf epidermal cells results in cytosolic and nuclear protein localization and reduced carotenoid and chlorophyll levels. **(A)** Phenotype of leaves expressing N-terminal and C-terminal TkPEL-like Cerulean fusions and corresponding controls. Leaves expressing *TkPEL-like* appear light green. **(B)** Confocal laser scanning microscopy images of leaves expressing TkPEL-like Cerulean fusions and corresponding controls. Scale bar = 40 µm. **(C, D)** Chlorophyll and carotenoid levels in leaves expressing wild-type *TkPEL-like* and three different mutant versions lacking conserved cysteine residues at position 19 and/or 29 compared to *Cerulean*-expressing and non-infiltrated controls. The chlorophyll content was calculated based on chlorophyll *a* and *b* (Supplementary Table S3). Data are means (± SD) of four independently infiltrated leaves per construct. Significant differences were assessed by ANOVA with Tukey's honest significant difference test ($p < 0.05$). The lower case letters represent the statistically different groups.

10.3389/fpls.2023.1228961

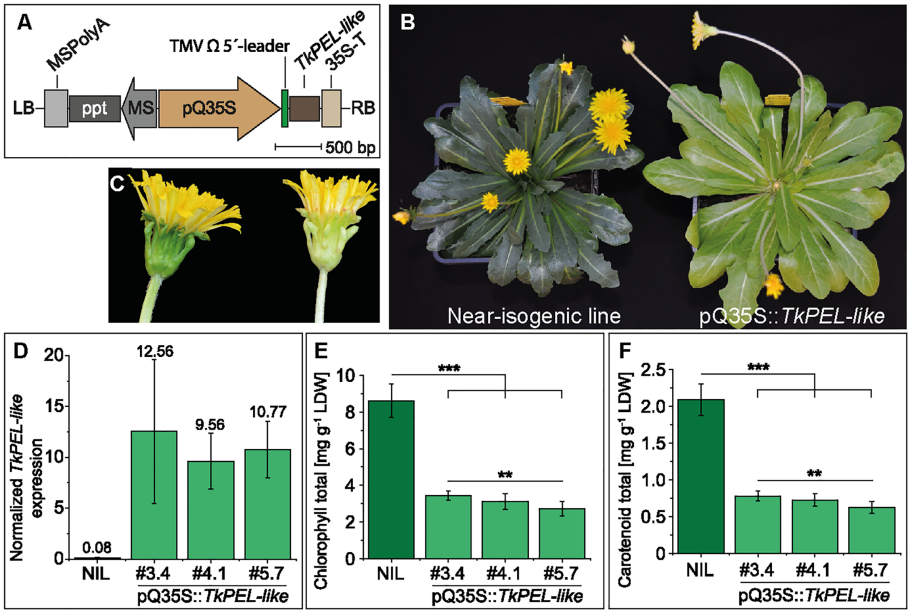



FIGURE 4
Generation and characterization of pale green *T. koksaghyz* plants overexpressing *TkPEL-like*. **(A)** Schematic representation of the T-DNA carrying the pQ35S promoter and TMV Ω 5′-leader sequence to achieve strong, constitutive expression of the *TkPEL-like* coding sequence in plants. Resistance gene cassettes: MS, *mannopine synthase* promoter; ppt, phosphinothricin resistance gene; MSPolyA, poly(A) sequences from the mannopine synthase gene. **(B)** 12-week-old $T_1$ generation plants of a near isogenic line (NIL) and a *TkPEL-like* overexpression line pQ35S::*TkPEL-like*. **(C)** Close-up of flowers from the NIL (left) and pQ35S::*TkPEL-like* (right) representing the $T_1$ generation. **(D–F)** Analysis of nine individual NIL plants and eight individual plants as biological replicates from three independent pQ35S::*TkPEL-like* lines (#3.4, #4.1 and #5.7) of the $T_2$ generation grown under controlled greenhouse conditions for 12 weeks. **(D)** Normalized expression levels of *TkPEL-like* confirm overexpression in leaves of pQ35S::*TkPEL-like* lines. Expression levels were normalized against *elongation factor 1a* (*TkEf1a*) and *ribosomal protein L27* (*TkRP*). Numbers above SD error bars are mean values. **(E)** Leaf chlorophyll content based on chlorophyll *a* and *b* (Supplemental Table S4). **(F)** Leaf carotenoid levels. Statistical differences in chlorophyll and carotenoid contents were assessed using non-parametric Mann-Whitney U-tests (**p < 0.01, ***p < 0.001).

The resulting progeny showed two alternative genotypes: transgenic plants that carried the *TkPEL-like* overexpression construct, which were named pQ35S::*TkPEL-like*, and NILs lacking the transgene. In common with the $T_0$ generation, the $T_1$ generation of pQ35S::*TkPEL-like* plants showed a pale green leaf phenotype and a white midrib (Figure 4B). The normally green parts of the petioles and flowers also appeared whitish (Figures 4B, C).

For in-depth analysis, we used three independent transgenic lines (#3.4, #4.1 and #5.7) grown from seeds harvested from $T_1$ plants crossed with wild-type plants, thus representing the $T_2$ generation. The resulting pQ35S::*TkPEL-like* plants showed no abnormal morphology (Supplementary Figure S4) or differences in growth rate, size of the aboveground tissues or flowering (Supplementary Figure S5) compared to NIL controls, but still had a pale green/whitish appearance, which confirmed the stable inheritance of this phenotype. Next, we used leaf cDNA for qRT-PCR analysis, which confirmed a massive increase in *TkPEL-like* expression (137-fold, on average) in the pQ35S::*TkPEL-like* plants compared to NIL controls (Figure 4D). We also observed a significant reduction in the amount of chlorophylls and carotenoids in the transgenic leaves compared to NIL leaves (Figures 4E, F and Supplementary Table S5), explaining the observed phenotype. After observing these effects on leaf isoprenoids, we tested the roots of the pQ35S::*TkPEL-like* plants because *T. koksaghyz* roots normally accumulate large quantities of isoprenoid end-products such as pentacyclic triterpenes and poly (*cis*-1,4-isoprene), the main component of natural rubber (Niephaus et al., 2019; Pütter et al., 2019). *TkPEL-like* overexpression was confirmed in the roots of all three transgenic lines, but the quantity of pentacyclic triterpenes and poly(*cis*-1,4-isoprene) remained similar to normal (Supplementary Figure S6; Supplementary Table S6). However, the dry root weight of lines #3.4 and #5.7 was significantly lower than that of NIL controls (Supplementary Figure S6).

## 3.5 The maximum quantum yield of PSII increases in pQ35S::*TkPEL-like* plants

To compare the photosynthetic performance of wild-type and pale green leaves, we measured the leaf fluorescence of 12-week-old NIL and pQ35S::*TkPEL-like* plants. We measured the maximum quantum yield ($F_v/F_m$) and the effective quantum yield ($F_q'/F_m'$) of PSII in three consecutive periods: dark-adaption, exposure to intense light, and during recovery in darkness (Figure 5A). During short-term illumination, both genotypes showed stable and comparably low $F_q'/F_m'$ values. In contrast, the $F_v/F_m$ values of both genotypes in the absence of light appeared to match those of leaves that were typically non-stressed (Björkman and Demmig, 1987), but this value was consistently and significantly higher for

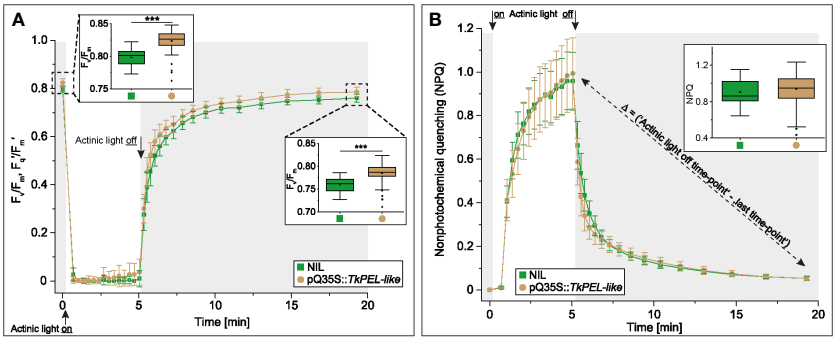



FIGURE 5
Quantum efficiency of photosystem II increases in 12-week-old pQ35S::*TkPEL-like* plants but non-photochemical quenching is not affected.
**(A)** Measurement of photosystem II (PSII) maximum ($F_v/F_m$) and effective ($F_q'/F_m'$) quantum yield and **(B)** Measurement of non-photochemical quenching (NPQ) following transition from dark to intense light. Data are means (± SD) of 13 pQ35S::*TkPEL-like* plants and 13 NILs. One leaf from each plant was harvested for measurement. Leaves were dark adapted for 20 min before applying photosynthetically active radiation (actinic light) with an intensity of 1076 µmol quanta $m^{-2}$ $s^{-1}$ for 5 min. This was followed by relaxation in the dark for ~15 min. The fluorescence levels of 14 independent areas per leaf were measured. The framed bar charts in panel A depict the first (left) and last (right) data points of the measurement. The framed bar chart in panel B depicts the calculated difference in NPQ of the data point 'actinic light off' and the last data point. Statistical differences were assessed using non-parametric Mann-Whitney U-tests (***p < 0.001).

the pale green leaves of pQ35S::*TkPEL-like* plants. We also calculated the NPQ, which was the same for both genotypes (Figure 5B). NPQ appeared quickly under short-term exposure to intense light and then relaxed in the same way in both genotypes.

## 3.6 Transcriptomic comparison of pQ35S::*TkPEL-like* and NIL plants

Given the clear negative impact of *TkPEL-like* overexpression on photosynthetic isoprenoid compounds in the leaf, but not on root isoprenoids, we investigated the underlying mechanism by comparing the transcriptomes of pQ35S::*TkPEL-like* and NIL leaves. We extracted RNA from the leaves of four individual 12-week-old plants from pQ35S::*TkPEL-like* lines #3.4, #4.1 and #5.7 and NILs as a control. The samples of each genotype were pooled so that two samples each representing 12 different plants were compared. The sequencing reads were mapped against the *T. koksaghyz* reference genome (Lin et al., 2022) with 93% efficiency, and were filtered for a minimum CPM of 0.2. The number of reads mapped to a gene was similar for both datasets (Supplementary Figure S7) and 30,093 transcripts could be detected for NIL plants and 29,830 for the pQ35S::*TkPEL-like* lines.

Differentially expressed genes (DEGs) were defined as those meeting the threshold $\log_2$FC (normalized expression) ≥ 1 or ≤ –1 with a q-value (Storey) of < 0.05. We identified 2646 (upregulated) and 3580 (downregulated) genes that were modulated at least two-fold in the pQ35S::*TkPEL-like* plants vs NIL, of which 1215 (upregulated) and 837 (downregulated) genes were modulated at least five-fold (Figures 6A, B). For independent validation of the transcriptomic data, expression levels of single genes were analyzed in the three separate lines representing each genotype using qRT-PCR (Supplementary Figure S8). This confirmed the overall direction of transcriptional changes in pQ35S::*TkPEL-like* lines. Here it became obvious that, for specific genes, the transcriptional change seemed to be dependent on the level of *TkPEL-like* (over)expression.

Gene Ontology (GO) terms were compared with the background of all genes expressed in NIL leaves, showing that several terms were significantly (p < 0.05) enriched among the DEGs (Figure 6C). We focused on DEGs meeting the threshold $-2.32 \leq \log_2(FC) \geq 2.32$. Among the genes downregulated at least five-fold, we observed enriched GO terms in the superordinate category 'Biological Process' related to 'signal transduction', 'signaling' and 'secondary metabolic process' and responses to different stimuli such as 'response to stress', 'response to abiotic/biotic stimuli' and 'response to endogenous/external stimuli'. Furthermore, the terms 'response to other organism', 'multi-organism process', 'secondary metabolic process', and 'response to biotic stimulus' were enriched for both the upregulated and downregulated genes, and the terms 'cell growth', 'cell wall organization or biogenesis' and terms related to 'cell or biological adhesion' were overrepresented only among the upregulated genes. Also among the downregulated genes, we observed enriched GO terms in the superordinate category 'Molecular Function' relating to 'DNA-binding transcription factor activity', 'transcription regulator activity' and 'signaling receptor activity', whereas the term 'oxidoreductase activity' was enriched for both the upregulated and downregulated genes. In the superordinate category 'Cellular Component', the upregulated genes were enriched for the terms 'apoplast', 'cell junction', 'external encapsulating structure' and 'cell periphery'.

We also screened for KEGG pathways significantly enriched (p < 0.05) among the DEGs ($-1 \leq \log_2(FC) \geq 1$) compared to all genes expressed in the NIL plants (Figure 6D). We identified 42 pathways that were overrepresented among the DEGs, many of which related to processes and diseases only found in animals. This probably reflects the conservation of fundamental enzymes among eukaryotes, and such pathways were ignored for subsequent analysis. Enriched pathways present in plants included plant

10.3389/fpls.2023.1228961

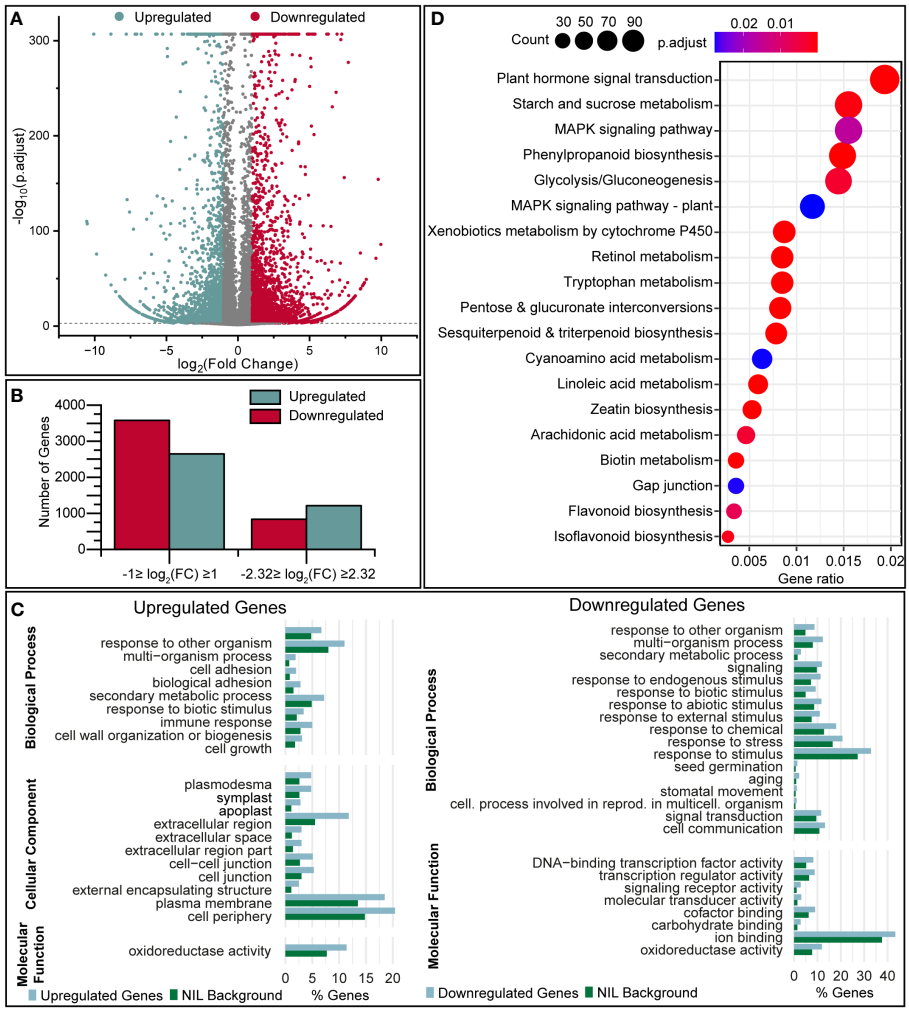



FIGURE 6
Transcriptomic analysis of pQ35S::*TkPEL-like* plants reveals a large number of differentially expressed genes. **(A)** Volcano plot showing the transcriptomic comparison of pQ35S::*TkPEL-like* and NIL plants. $Log_2$(FC) values are plotted against $-log_{10}$(p.adjust). Negative $log_2$(FC) represents transcriptional upregulation in pQ35S::*TkPEL-like* plants compared to NIL. **(B)** Numbers of DEGs detected using different filtering criteria. Upregulation and downregulation refers to transcript levels in pQ35S::*TkPEL-like* plants. **(C)** GO terms enriched among DEGs ($-5 \geq FC \geq 5$) in pQ35S::*TkPEL-like* plants, showing terms of levels 1–3 significantly enriched ($p < 0.05$) among the DEGs compared to the NIL background. **(D)** KEGG pathways enriched among DEGs ($-1 \geq FC \geq 1$) in pQ35S::*TkPEL-like* plants. Gene ratios describe the number of genes assigned to a specific pathway compared to all genes that were assigned a pathway. Bubble size represents the absolute number of genes associated with the pathway and colors represent the adjusted p-value for enrichment compared to all transcripts identified in the NIL background. Pathways referring to processes in animals have been omitted.

hormone signal transduction and the biosynthesis of phenylpropanoids, sesquiterpenoids, triterpenoids, flavonoids and isoflavonoids.

### 3.6.1 Genes representing chlorophyll and carotenoid biosynthesis and related pathways are modulated in pQ35S::*TkPEL-like* leaves

The lower abundance of chlorophylls and carotenoids in pQ35S::*TkPEL-like* leaves suggests transcriptional changes in the MVA and MEP pathways as well as chlorophyll and carotenoid biosynthesis. In the MEP pathway, the genes representing the first step (*DXS*, encoding 1-deoxy-D-xylulose-5-phosphate synthase) and final step (*IspH*, encoding 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase) were downregulated in the leaves of pQ35S::*TkPEL-like* plants, with *DXS* expression levels reduced by ~50% compared to the NIL control (Figure 7). In contrast, the gene responsible for the second step (*DXR*, encoding 1-deoxy-D-xylulose 5-phosphate reductoisomerase) was upregulated in these plants. On the other hand, *HMGR*, encoding hydroxymethylglutaryl-CoA reductase, the rate-limiting enzyme of the MVA pathway (Gondet et al., 1992; Suzuki et al., 2004; Hemmerlin et al., 2012), was also strongly repressed. Our data suggest that lower amounts of $C_5$ isoprene diphosphate precursors are produced by the MEP and MVA pathways in leaves overexpressing *TkPEL-like*.

We detected four DEGs representing the downstream metabolic pathway leading to carotenoids. Interestingly, genes representing two different isoforms of geranylgeranyl diphosphate synthase (GGPS), which condenses four $C_5$ isoprenoid precursors to

10.3389/fpls.2023.1228961

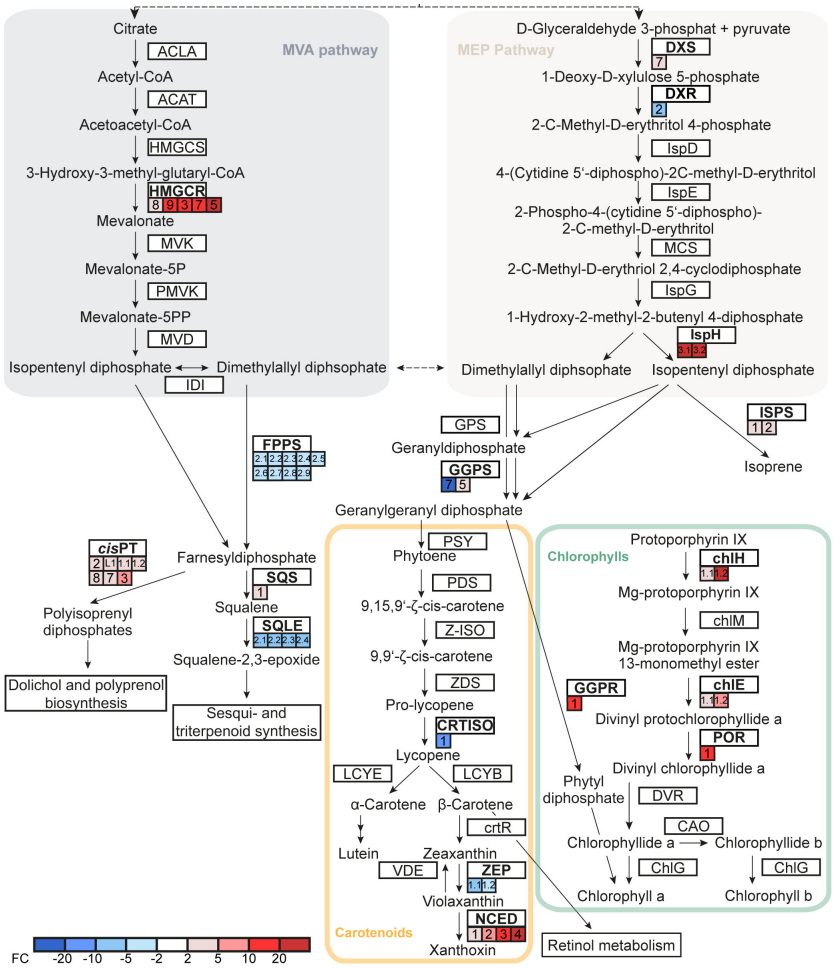



FIGURE 7
*TkPEL-like* overexpression affects genes involved in chlorophyll and carotenoid biosynthesis as well as precursors and connected pathways. Expression fold-changes of NIL compared to pQ35S::*TkPEL-like* plants are represented by colored boxes. Multiple boxes for one gene represent independent RNA IDs. Numbers in boxes assign transcripts to IDs based on the published genome (Lin et al., 2018) and Supplemental Table S9. ACLA, ATP citrate (pro-S)-lyase; ACAT, acetyl-CoA C-acetyltransferase; HMGCS, hydroxymethylglutaryl-CoA synthase; HMGCR, hydroxymethylglutaryl-CoA reductase; MVK, mevalonate kinase; PMVK, phosphomevalonate kinase; MVD, diphosphomevalonate decarboxylase; IDI, isopentenyldiphosphate δ-isomerase; DXS, 1-deoxy-D-xylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; IspD, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; IspE, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; IspF, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; IspG, (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; IspH, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase; ISPS, isoprene synthase; GPS, geranyldiphosphate synthase; FPPS, farnesyldiphosphate synthase; GGPS, geranylgeranyldiphosphate synthase; PSY, phytoene synthase; PDS, 15-*cis*-phytoene desaturase; Z-ISO, ζ-carotene isomerase; ZDS, ζ-carotene desaturase; CRTISO, prolycopene isomerase; LCYB, lycopene β-cyclase; LCYE, lycopene ϵ-cyclase; crtR, β-carotene hydroxylase; ZEP, zeaxanthin epoxidase; NCED, 9-*cis*-epoxycarotenoid dioxygenase; VDE, violaxanthin de-epoxidase; chlH, Mg chelatase subunit H; chlM, Mg-protoporphyrin O-methyltransferase; chlE, Mg-protoporphyrin IX monomethyl ester cyclase; POR, protochlorophyllide reductase; GGPR, geranylgeranyldiphosphate reductase; DVR, divinyl chlorophyllide a 8-vinyl-reductase; CAO, chlorophyllide a oxygenase; ChlG, chlorophyll synthase; SQS, squalene synthase; SQLE, squalene monooxygenase; *cis*PT, *cis*-prenyltransferase.

geranylgeranyl diphosphate (GGPP) as a precursor for carotenoid biosynthesis, were antagonistically modulated in the pQ35S::*TkPEL-like* lines. Furthermore, the genes encoding prolycopene isomerase (*CRTISO*) and zeaxanthin epoxidase (*ZEP*) in the carotenoid pathway were strongly upregulated (Figure 7), whereas the *NCED* gene encoding 9-*cis*-epoxycarotenoid dioxygenase, which directs metabolic flux from the carotenoid pathway toward abscisic acid biosynthesis, was strongly repressed.

GGPP is not only a precursor for carotenoid biosynthesis but also for the production of intact chlorophyll *a* and *b*. It is reduced to phytyl diphosphate by geranylgeranyl diphosphate reductase (GGPR) and attached to chlorophyllide derived from protoporphyrin by chlorophyll synthase (Rüdiger, 1993; Keller et al., 1998). We found that *GGPR* and three genes involved in porphyrin ring synthesis were downregulated in leaves of pQ35S::*TkPEL-like* plants (Figure 7), providing a possible explanation for the lower chlorophyll levels (Figure 4). Genes representing competing metabolic pathways that build on farnesyl diphosphate (FPP, $C_{15}$) as a substrate were also affected. For example, significant transcriptional changes were observed for the genes encoding squalene synthase and squalene monooxygenase, which direct metabolic flux towards triterpenoid and phytosterol synthesis,

and *cis*-prenyltransferases, which form polyisoprenyl diphosphates essential for *N*-glycan synthesis and photosynthetic performance by modulating thylakoid membrane dynamics (Surmacz and Swiezewska, 2011; Akhtar et al., 2017; van Gelder et al., 2018).

### 3.6.2 Several genes related to photosynthetic complexes, circadian rhythm and light regulation are downregulated in pQ35S::*TkPEL-like* leaves

Given the downregulation of photosynthesis-related genes in *Arabidopsis AtRPGE2* overexpression lines (Kim et al., 2023), the presence of a light-sensitive transcription factor binding site in the *TkPEL-like* promoter and the identification of several light-responsive genes among the DEGs involved in carotenoid and chlorophyll synthesis, we next focused on the analysis of genes involved in photosynthesis (other than chlorophyll biosynthesis), circadian rhythm and light-dependent signaling (Supplemental Table S6). Notably, we observed the downregulation of the genes *COP1* and *PIF1*, which as discussed above encode negative regulators of photomorphogenesis. Regulators of the circadian clock, which control flowering time, were also strongly downregulated in the *TkPEL-like* overexpression lines, including *ELF3*, *CO* and *FT*. Several FAR1-related sequence (*FRS*) homologs, encoding proteins that interact with AtRPGE2 (Dreze et al., 2011), were also downregulated. Interestingly, although GLK1 and GLK2 interact with AtRPGE2 and are transcriptionally correlated with *AtRPGE2* overexpression (Kim et al., 2023), we did not detect this sequence among the DEGs. However, we found that a number of photosynthesis-related genes, including those encoding subunits of PS I and II, light-harvesting complex II chlorophyll *a*/*b* binding protein 1, and ATPase subunits, were repressed. Similar profiles were reported for the *AtRPGE2* overexpression lines (Kim et al., 2023). Only genes encoding a cytochrome b6-f complex iron-sulfur subunit and a photosystem I subunit III were induced by *TkPEL-like* overexpression (Supplemental Table S7). Other DEGs encoded homologs of transcription factor MYB75/PAP1 (which may be transcriptionally regulated by HY5). MYB75/PAP1 is a master regulator of anthocyanin biosynthesis and targets *CHS* among other genes (Shin et al., 2013).

### 3.6.3 RNA degradation genes are differentially expressed in pQ35S::*TkPEL-like* leaves

Initial homology searches revealed similarities between TkPEL-like and a polyribonucleotide nucleotidyltransferase, also known as polyribonucleotide phosphorylase (PNPase), which is involved in RNA maturation and degradation (Bollenbach et al., 2004; Germain et al., 2011). Therefore, we screened genes representing the RNA degradation and surveillance pathways to see if they were also modulated by *TkPEL-like* overexpression (Supplemental Table S8). Enolase and 6-phosphofructokinase are thought to form an RNA degradosome complex with PNPase and other proteins in *Bacillus subtilis*, and *T. koksaghyz* homologs were found to be differentially expressed in pQ35S::*TkPEL-like* plants. In addition, the DEGs included several cofactors of RNA exosomes and subunits of the Ccr4-NOT complex, an important modulator of gene expression at the mRNA level (Collart, 2016). One particular downregulated gene encoded a putative ATP-dependent RNA helicase DOB1/MTR4 homolog. In yeast, the helicase activity of this protein and its ability to act as a cofactor in poly(A)polymerase complexes involved in RNA degradation are thought to be important for the function of exosomes (Vaňáčová et al., 2005; Falk et al., 2017; Schmid and Jensen, 2019). In addition, a poly(A)-binding protein with multiple functions in gene regulation (e.g., by mediating poly(A) tail synthesis and nuclear mRNA export) was upregulated in the pQ35S::*TkPEL-like* plants (Mangus et al., 2003).

## 4 Discussion

### 4.1 The *TkPEL-like* overexpression phenotype resembles the effects of *AtRPGE2* and *DCAR_032551* in other plants

Recent progress has provided insight into the complexity of light-dependent regulation and isoprenoid biosynthesis in plants (reviewed by Vranová et al., 2012; Tholl, 2015; Li et al., 2022; Wang et al., 2022). Previous studies have indicated that the carrot protein DCAR_032551 (Iorizzo et al., 2016) and the *Arabidopsis* protein RPGE2 (Ichikawa et al., 2006; Kim K. et al., 2016; Kim et al., 2023) are involved in light signal transduction and the regulation of isoprenoids and isoprenoid-containing compounds such as carotenoids and chlorophylls. We have now identified TkPEL-like, a putative homolog of DCAR_032551/AtRPGE2 in *T. koksaghyz*. The overexpression of *TkPEL-like* in *N. benthamiana* and *T. koksaghyz* led to a pale green leaf phenotype and a reduction in chlorophyll and carotenoid levels (Figure 4). This indicates the cross-species functional conservation of this protein in photosynthetic organisms, which is also supported by the expression of rice (*Oryza sativa*) *RPGE* in *Arabidopsis* and the interaction between RPGE and GLK homologs of different species (Kim et al., 2023). However, the faster stem elongation and early flowering observed in *Arabidopsis* plants overexpressing *AtRPGE2* (Ichikawa et al., 2006) was not observed in our pQ35S::*TkPEL-like T. koksaghyz* plants. Contrary to the normal flowering time, different genes involved in the circadian rhythm, which regulate flowering in other plants, were among the DEGs in pQ35S::*TkPEL-like* vs. NILs (Supplementary Table S7). To our knowledge, the control of flowering time in *T. koksaghyz* is not yet understood and we do not know the exact functions of the corresponding genes because different functions have been described for different isoforms (Beinecke et al., 2018; Lin et al., 2019; Schmidt et al., 2020). The vernalization-dependent flowering of *T. koksaghyz* appears to play an important role, but the various protein isoforms that regulate flower development in *T. koksaghyz* require further detailed analysis.

*DCAR_032551* is more strongly expressed in highly pigmented carrot roots compared to pale roots, but the gene carries frameshift mutations that probably cause functional disruption (Iorizzo et al., 2016). The absence of a functional *DCAR_032551* gene product therefore appears to be associated with high carotenoid levels whereas a functional DCAR_032551 protein has the opposite

effect. This matches the phenotype of *TkPEL-like* overexpression in *T. koksaghyz*. Furthermore, neither the carrot nor dandelion plants discussed above showed any developmental abnormalities, suggesting that the synthesis of photosynthetic isoprenoids is affected rather than photomorphogenesis as a whole. However, in contrast to carrots, the isoprenoid content of *T. koksaghyz* roots was unaffected by *TkPEL-like* overexpression (Supplementary Figure S6). The transcriptional downregulation of the MVA pathway (as detected in the leaves) may therefore be restricted to green tissues and root isoprenoid synthesis might be regulated differently. MEP pathway downstream products such as carotenoids have not been identified as typical latex/root metabolites, although some MEP pathway proteins have been detected in *T. koksaghyz* latex (Niephaus et al., 2019; Xie et al., 2019).

## 4.2 *TkPEL-like* overexpression may primarily repress chlorophyll and isoprenoid precursor biosynthesis resulting in the secondary induction of carotenoid biosynthesis genes

Transcriptomic analysis of pQ35S::*TkPEL-like* lines revealed the downregulation of genes involved in chlorophyll biosynthesis and the MEP and MVA pathways in leaves, providing an explanation for the observed phenotype. In highly pigmented carrots, the *DXS* gene (representing the first step in the MEP pathway) was induced, whereas in the pale green pQ35S::*TkPEL-like* dandelion plants it was strongly suppressed (Iorizzo et al., 2016). This suggests TkPEL-like is a negative regulator of *DXS*, and that other transcriptional changes may be secondary effects. For example, inhibiting the MEP pathway in *Arabidopsis* revealed that the MEP and tetrapyrrole pathways are co-regulated to maintain a metabolic balance and to prevent photo-oxidative damage by intermediates (Kim et al., 2013).

*IspH* expression was also strongly reduced in the pQ35S::*TkPEL-like* plants, in contrast to the outcome in carrots. In *Arabidopsis* and tomato (*Lycopersicon esculentum*), IspH is a key enzyme in MEP-derived plastidial isoprenoid biosynthesis, and the upregulation of *IspH* increased the incorporation of isoprenoid precursors into a downstream recombinant pathway by 100% compared to the upregulation of *DXS* (Estévez et al., 2001; Botella-Pavía et al., 2004; Kishimoto and Ohmiya, 2006; Banerjee and Sharkey, 2014). The simultaneous downregulation of *IspH* and *DXS* may have reduced the allocation of IPP and DMAPP from the MEP pathway, which could not be compensated by the upregulation of *DXR* due to the limited amount of substrate.

The downregulation of chlorophyll biosynthesis genes and other photosynthesis-related genes was similar to the transcriptional changes observed following *AtRPGE2* overexpression in *Arabidopsis*, although not all the genes downregulated in *Arabidopsis* were among the DEGs in *T. koksaghyz* (Kim K. et al., 2016; Kim et al., 2023) (Supplementary Table S7). In *Arabidopsis*, RPGE2 prevents GLK target gene activation by the formation of heterodimers, which probably contributed to the transcriptional changes observed following *AtRPGE2* overexpression (Kim et al., 2023). The upregulation of *GLK* as described in Arabidopsis could not be detected in our RNA-seq data. To test for a conserved molecular mode of action for TkPEL-like, as suggested in rice (Zhang et al., 2021), GLK homologs in *T. koksaghyz* and their potential interaction with TkPEL-like should be investigated in the future.

The higher maximum quantum yield in pQ35S::*TkPEL-like* plants compared to NIL controls (Figure 5) was likewise reported for plants overexpressing AtRPGE2 (Kim et al., 2023). However, the associated lower total seed yields reported for *Arabidopsis* were not evaluated in our experiments. The increase in $F_v/F_m$ indicates an increase of the proportion of the absorbed light energy that is used for photochemical reactions, which could be a consequence of the lower chlorophyll content in order to maintain sufficient photosynthesis. In chlorophyll-deficient rice mutants, an increased photosynthetic rate per chlorophyll molecule compensated for the negative impact of chlorophyll reduction (Li et al., 2013). It is possible that such an increase also contributed to sufficient photosynthesis and normal biomass accumulation in pQ35S::*TkPEL-like* plants (Supplementary Figure S5). The lower chlorophyll content could have also reduced the photochemical damage and heat stress in leaves absorbing more light than required for maximum photosynthesis, and energy otherwise used for repair could therefore be used elsewhere, contributing to the comparable biomass accumulation in pQ35S::*TkPEL-like* and NILs (Hamblin et al., 2014). The comparable NPQ in the pQ35S::*TkPEL-like* plants and NIL controls supports the hypothesis that *TkPEL-like* overexpression does not trigger stress. This is particularly notable given that many essential photosynthetic components, such as plastoquinone (an intramembrane electron acceptor downstream of PSII) are also derived from the MEP pathway that was transcriptionally downregulated in pQ35S::*TkPEL-like* plants (Nowicka and Kruk, 2010).

Despite the lower carotenoid levels in pQ35S::*TkPEL-like* plants, carotenoid pathway genes tended to be upregulated. This may also be a primary effect, but we hypothesized it might have been a secondary feedback mechanism in response to low carotenoid levels (Avendaño-Vázquez et al., 2014). However, the limited pool of isoprenoid precursors may not have allowed for compensation via transcriptional upregulation. The possible upregulation of *CCD*, encoding a carotenoid cleavage dioxygenase, suggests an exacerbation of low carotenoid synthesis, given that this enzyme is responsible for the degradation of carotenoids in chrysanthemum, resulting in white petals (Ohmiya et al., 2006).

Three DEGs were found in the pathway leading from IPP and DMAPP to the formation of the first carotenoids. IPP and DMAPP are condensed by prenyltransferases to form pools of $C_{10}$ (geranyl diphosphate), $C_{15}$ (FPP) and $C_{20}$ (GGPP) precursors (Hemmerlin et al., 2012). For carotenoid biosynthesis, two GGPP molecules are condensed by phytoene synthase (PSY) to form phytoene, which is desaturated and isomerized in several steps (Shumskaya and Wurtzel, 2013). Interestingly, the genes encoding FFP synthase (*FDPS*) and GGPP synthase (*GGPS*) were upregulated in leaves of pQ35S::*TkPEL-like* plants, possibly to increase the flux from IPP/DMAPP in this direction. Another *GGPS* isoform was

downregulated in the transgenic plants. Given the widespread occurrence of *GGPS* gene families together with the specialization of the individual enzymes (Beck et al., 2013; Ruiz-Sola et al., 2016), the opposing transcriptomic responses of the dandelion *GGPS* genes may reflect their spatial expression profiles or protein localization. In tobacco (*Nicotiana tabacum*), NtGGPPS1 and a light-regulated small subunit of GGPS were shown to interact with PSY, allowing the channeling of GGPP between successive pathway enzymes. The differential expression of *GGPPS* may therefore affect the efficiency of PSY in pQ35S::*TkPEL-like* plants, although the *PSY* gene itself was not differentially expressed. The specific functions of the differentially regulated isoforms should be examined in detail to validate the hypotheses derived from our expression data.

The induced genes in the downstream parts of the carotenoid pathway included *ZEP*, encoding zeaxanthin epoxidase (responsible for violaxanthin biosynthesis) whereas the gene encoding 9-cis-epoxycarotenoid dioxygenase (which converts violaxanthin to xanthoxin) was suppressed, indicating that violaxanthin levels were specifically increased in the pQ35S::*TkPEL-like* lines. In the xanthophyll cycle, violaxanthin is reversibly de-epoxidized to zeaxanthin by violaxanthin de-epoxidase under intense light to efficiently reduce photo-oxidative stress and lipid peroxidation (Havaux et al., 2007). The induction of *ZEP* may therefore prepare the cell to deal with reactive oxygen species by providing enough substrate for the detoxifying reaction that converts violaxanthin to zeaxanthin, and the limited quantity of carotenoids available was optimally deployed.

Finally, genes involved in flavonoid/isoflavonoid metabolism were also significantly upregulated in pQ35S::*TkPEL-like* plants compared to NIL controls (Figure 6B). One example is the key enzyme CHS, which also plays a role in cell wall organization and biosynthesis (Zuk et al., 2016), a process significantly enriched among upregulated genes (Figure 6C). *CHS* is also regulated by light, via HY5 and COP1 (Ang et al., 1998; Thain et al., 2002). These data suggest that flavonoid pigments, such as anthocyanins, were enriched to compensate for the low carotenoid levels in the plastids thus ensuring photoprotection, given there was no difference in fitness between the pQ35S::*TkPEL-like* plants and NILs. Such compensatory functions have been proposed in response to UV radiation (Guidi et al., 2016).

## 4.3 TkPEL-like may act downstream of HY5 to enable light-dependent regulation

*TkPEL-like* overexpression was negatively correlated with the expression of various modulators of the light response, in contrast to the highly pigmented carrots in which there was a positive correlation (Iorizzo et al., 2016). Transcription factors (HY5 and PIFs) and the E3 ubiquitin ligase COP1 are key regulators of the light response. PIFs negatively regulate photomorphogenesis and repress the MEP pathway (Chenge-Espinosa et al., 2018), as well as chlorophyll and carotenoid biosynthesis (Huq et al., 2004; Moon et al., 2008; Toledo-Ortiz et al., 2010; Tang et al., 2012; Job and Datta, 2021). PIF1 and PIF3 repress transcription, and their degradation in the presence of light thus enables the expression of MEP pathway genes such as *IspH* (Chenge-Espinosa et al., 2018). In *pif1* and *pif3* mutant seedlings, protochlorophyllide accumulates in the dark, leading to bleaching under illumination, in parallel to higher levels of light-induced chlorophyll and increased carotenoid accumulation in the dark (Huq et al., 2004; Toledo-Ortiz et al., 2010; Job and Datta, 2021).

PIFs are also potential COP1 cofactors that represses photomorphogenesis in the dark by accumulating in the nucleus, facilitating the degradation of positive regulators of light signaling such as HY5 (Deng et al., 1991; Osterlund et al., 1999; Saijo et al., 2003; Jang et al., 2005; Xu et al., 2014). Accordingly, *cop1* mutants show photomorphogenesis in the dark (Deng et al., 1991). These reports contrast with the indicated downregulation of *PIF* and *COP1* in pale green pQ35S::*TkPEL-like* plants. The identification of a G-box element in the *TkPEL-like* promoter hints that *TkPEL-like* may be regulated by HY5, PIFs, or both. *HY5* expression was unaffected in pQ35S::*TkPEL-like* plants, suggesting that HY5 may act upstream of TkPEL-like in the regulatory hierarchy. In contrast, *PIFs* were differentially expressed in the pQ35S::*TkPEL-like* plants, but *AtRPGE2* was identified as a direct target of PIFs (Kim K. et al., 2016). The exact relationship between PIFs and *TkPEL-like* should be analyzed in future studies to gain a more detailed understanding of this regulatory network in *T. koksaghyz*.

One explanation for these contradictory results is that *COP1* and *PIF1* were downregulated as a feedback response to low carotenoid and chlorophyll levels in order to de-repress photomorphogenesis and associated pigment synthesis, but the limited pool of isoprenoid precursors prevented the normalization of carotenoid and chlorophyll levels. Furthermore, although three transcripts annotated as *PIF1* or *PIF3* with sequence identities > 50% were indeed downregulated in pQ35S::*TkPEL-like* plants, one further transcript with high similarity to the ATP-dependent DNA helicase PIF1-like from *Cynara cardunculus* var. *scolymus* was strongly upregulated. The overexpression of a single *PIF1* isoform together with *TkPEL-like* may have been sufficient to effectively repress photomorphogenesis, including the downregulation of MEP/MVA and chlorophyll biosynthesis pathway gene expression. However, we found no evidence for the transcriptional modulation of PIF target genes such as *PSY*, as described in other plants (Toledo-Ortiz et al., 2010). *PSY*, among other photomorphogenic targets, was shown to be antagonistically regulated by PIFs and HY5 (Toledo-Ortiz et al., 2014). Therefore, unaffected *HY5* expression may have been sufficient to regulate HY5 targets in a normal manner, despite differential *PIF* expression. This would further support the idea that TkPEL-like acts downstream of HY5 in the signaling hierarchy.

We observed the downregulation of *COP1*, but this gene was co-expressed with nonfunctional *DCAR_032551* in highly pigmented carrots (Iorizzo et al., 2016). COP1 is excluded from the nucleus under prolonged illumination (Osterlund and Deng, 1998; Subramanian et al., 2004), so the transcriptional modulation of *COP1* in response to light may not be critical for photomorphogenesis as long as nuclear exclusion is not impaired. Notably, translational and post-translational regulation can also influence metabolic pathways in a way that is not fully evident from our data (Hemmerlin et al., 2012; Shumskaya and Wurtzel, 2013; Banerjee and Sharkey, 2014).

## 4.4 TkPEL-like may play a role in RNA degradation

TkPEL-like is related to a PNPase and the presence of a PIN_Fcf1-like domain prompted us to analyze genes involved in mRNA processing. This indeed revealed several DEGs that are homologs of genes involved in RNA degradation and surveillance. PNPases play various roles in RNA metabolism. In *Arabidopsis*, PNPase was shown to facilitate plastid mRNA, tRNA and 23S rRNA metabolism, including poly(A) tail formation and degradation, as well as RNA degradation in general (Walter et al., 2002; Holec et al., 2006; Germain et al., 2011). Knockout mutants revealed a phenotype similar to pQ35S::*TkPEL-like* plants, with pale young leaves but near normal mature tissues (Sauret-Güeto et al., 2006). The PIN_Fcf1-like domain (~120 residues) may facilitate pre-rRNA cleavage, nonsense mediated mRNA decay and RNA interference (Clissold and Ponting, 2000; Fatica et al., 2004). However, the level of similarity between TkPEL-like, and PNPase and the PIN_Fcf1-like domain was only low, and PNPases are much larger (e.g., 922 amino acids for AtPNPase) than the 108-amino-acid TkPEL-like protein (Walter et al., 2002). Furthermore, plant PNPases have only been found in chloroplasts and mitochondria (Li et al., 1998; Yehudai-Resheff et al., 2001; Holec et al., 2006; Germain et al., 2012), whereas our data suggest that TkPEL-like is localized in the nucleus and cytosol. The secondary structure of the PIN_Fcf1-like domain also differs from that predicted for TkPEL-like (Figure 1) (Senissar et al., 2017). However, a number of conserved acidic residues are found in both sequences. These data suggests that the function of TkPEL-like differs from that described in other species.

## 4.5 TkPEL-like probably suppresses genes involved in chlorophyll and isoprenoid precursor biosynthesis through its conserved N-terminal domain in a light-dependent manner

Our results support the hypothesis that the protein family containing AtRPGE2 and TkPEL-like acts as a negative regulator of photomorphogenesis (specifically chlorophyll and carotenoid accumulation) and represents an early step in the light-dependent signal transduction system (Iorizzo et al., 2016). The involvement of TkPEL-like proteins in light signal transduction and photomorphogenesis would also explain why this protein family is almost exclusive to photosynthetic organisms thus far, with highly conserved functions (Kim et al., 2023). However, RNAi-induced knockdown in *Arabidopsis* (Ichikawa et al., 2006) and our attempt to generate *T. koksaghyz* knockdown lines consistently resulted in a lethal phenotype, whereas an *Atrpge1/2/3* triple mutant was viable (Kim et al., 2023). The phenotype of the triple mutant contrasted with the results of overexpressing *AtRPGE1*, *AtRPGE2* or *TkPEL-like*, which suggested partial redundancy among the different isoforms in *Arabidopsis* (Kim et al., 2023). Additionally, highly pigmented carrots expressing a presumably nonfunctional version of *DCAR_032551* were also viable (Iorizzo et al., 2016). This might reflect a functional specialization of DCAR_032551 so that only root isoprenoids derived from the MEP pathway are affected. The authors did not report low chlorophyll levels in the aboveground organs, and *TkPEL-like* overexpression did not influence MVA pathway isoprenoids in *T. koksaghyz* roots.

Predominant expression of *TkPEL-like* in leaves from the ninth week onwards (Figure 5B) coincided with the development of a visible leaf rosette under our greenhouse conditions (Supplementary Figure S5). This expression profile correlated with that of *AtRPGE2* (Klepikova et al., 2016). Given the hypothesis that the corresponding protein is a repressor of photosynthetic gene expression, de-etiolation and photomorphogenesis, low expression levels in young plants may promote leaf development and coloration to cope with the initial exposure to light, and higher expression levels in mature leaves are necessary to restrict the aforementioned processes to an appropriate level.

Our data also support recent findings on the molecular function of AtRPGE2 (Kim et al., 2023). The interaction between AtRPGE2 and GLK in the cytosol and nucleus of *N. benthamiana* cells agrees with our results showing the localization of Cerulean–TkPEL-like fusion proteins in the cytosol and the nucleus in the same species. However, TkPEL-like lacks a nuclear localization signal, and may be transported into the nucleus passively due to its small size or actively transported in response to light, as reported for photoreceptors (Genoud et al., 2008; Galvão and Fankhauser, 2015).

TkPEL-like mutants with either one or two conserved cysteines replaced did not suppress chlorophyll or carotenoid biosynthesis to the same degree as the wild-type protein, suggesting the cysteine residues are functionally important (Figure 3). The cysteine residues found in the N-terminal segment of TkPEL-like are highly conserved between *Arabidopsis*, carrot and *T. koksaghyz* (Figure 1A, B), and may therefore be required for their molecular function. The conserved cysteine residues could also serve as points of attack for redox signaling (Couturier et al., 2013), which is known to regulate transcription, posttranslational modifications and retrograde signaling (Surpin et al, 2002; Balmer et al., 2003; Lemaire et al., 2004).

## 4.6 Conclusions

TkPEL-like is a promising candidate for the regulation of isoprenoid biosynthesis in leaves because it affects leaf isoprenoid but not root isoprenoid levels when constitutively overexpressed. It may act within the light-dependent signal transduction pathway and react to the redox status of the cell, thereby enabling responses to environmental cues. Our transcriptomic data provide a broad overview of the pathways affected by *TkPEL-like* overexpression and can be used as a basis for future functional studies to validate our hypotheses and to fully understand the complex regulatory network controlling isoprenoid biosynthesis and photomorphogenesis. A better understanding of the metabolic network will facilitate future breeding approaches aiming to modulate plant isoprenoid levels such as chlorophylls and carotenoids, or possibly other metabolites that we have not studied yet. It could also lead to the development of a suitable plant-based production platform for valuable metabolites.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/sra/PRJNA985648.

## Author contributions

SMW, VAB, K-UR and CSG conceived and designed the experiments. SMW, VAB and NvD conducted the experiments. SMW, VAB and K-UR analyzed the data. NvD, DP and CSG contributed the reagents, materials, and analytical tools. SMW and RMT wrote the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Fraunhofer Internal Programs and the Federal Ministry of Food and Agriculture Grant No. 2219NR415.

## Acknowledgments

We thank Denise Weinberg (Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Münster), Daniela Ahlert and Sascha Ahrens (both University of Münster) for their technical assistance and Prof. Dr. Michael Hippler and Dr. Philipp Gäbelein (both University of Münster) for their support with the chlorophyll fluorescence analysis.

## Conflict of interest

RMT was employed by TRM Ltd during the study.

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1228961/full#supplementary-material

## References

Akhtar, T. A., Surowiecki, P., Siekierska, H., Kania, M., van Gelder, K., Rea, K. A., et al. (2017). Polyprenols are synthesized by a plastidial cis-prenyltransferase and influence photosynthetic performance. *Plant Cell. Oxford Univ. Press* 29 (7), 1709. doi: 10.1105/TPC.16.00796

Ang, L. H., Chattopadhyay, S., Wei, N., Oyama, T., Okada, K., Batschauer, A., et al. (1998). Molecular interaction between COP1 and HY5 defines a regulatory switch for light control of Arabidopsis development. *Mol. Cell. Cell Press* 1 (2), 213–222. doi: 10.1016/S1097-2765(00)80022-2

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., et al. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinf. (Oxford England) Bioinf.* 33 (21), 3387–3395. doi: 10.1093/BIOINFORMATICS/BTX431

Armenteros, J. J. A., Sønderby, C. K., Sønderby, S. K., Nielsen, H., and Winther, O. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance* 2 (5), 1–14. doi: 10.26508/lsa.201900429

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., et al. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules', *Nucleic acids research*. *Nucleic Acids Res.* 44 (W1), W344–W350. doi: 10.1093/NAR/GKW408

Avendaño-Vázquez, A. O., Cordoba, E., Llamas, E., San Román, C., Nisar, N., De la Torre, S., et al. (2014). An uncharacterized apocarotenoid-derived signal generated in ζ-carotene desaturase mutants regulates leaf development and the expression of chloroplast and nuclear genes in Arabidopsis. *Plant Cell. Oxford Univ. Press* 26 (6), 2524. doi: 10.1105/TPC.114.123349

Baker, N. R. (2008). Chlorophyll fluorescence: a probe of photosynthesis in *vivo*. *Annu. Rev. Plant Biol. Annu. Rev. Plant Biol.* 59, 89–113. doi: 10.1146/ANNUREV.ARPLANT.59.032607.092759

Balmer, Y., Koller, A., Del Val, G., Manieri, W., Schürmann, P., Buchanan, B. B., et al. (2003). Proteomics gives insight into the regulatory function of chloroplast thioredoxins. *Proc. Natl. Acad. Sci. U. States A. Natl. Acad. Sci.* 100 (1), 370–375. doi: 10.1073/PNAS.232703799/SUPPL_FILE/7037TABLE4.HTML

Banerjee, A., and Sharkey, T. D. (2014). Methyl erythritol 4-phosphate (MEP) pathway metabolic regulation. *This J. is © R. Soc. Chem.* 00, 1–3. doi: 10.1039/x0xx00000x

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., et al. (2002). The Pfam protein families database. *Nucleic Acids Res. Nucleic Acids Res.* 30 (1), 276–280. doi: 10.1093/NAR/30.1.276

Beale, S. I. (1999). Enzymes of chlorophyll biosynthesis. *Photosynthesis Res. Kluwer Acad. Publishers* 60 (1), 43–73. doi: 10.1023/A:1006297731456

Beck, G., Coman, D., Herren, E., Ruiz-Sola, M. Á., Rodríguez-Concepción, M., Gruissem, W., et al. (2013). Characterization of the GGPP synthase gene family in Arabidopsis thaliana. *Plant Mol. Biol. Springer Netherlands* 82 (4–5), 393–416. doi: 10.1007/s11103-013-0070-z

Beinecke, F. A., Grundmann, L., Wiedmann, D. R., Schmidt, F. J., Caesar, A. S., Zimmermann, M., et al. (2018). The FT/FD-dependent initiation of flowering under long-day conditions in the day-neutral species Nicotiana tabacum originates from the facultative short-day ancestor Nicotiana tomentosiformis. *Plant J. John Wiley Sons Ltd.* 96 (2), 329–342. doi: 10.1111/TPJ.14033

Benninghaus, V. A., Van Deenen, N., Müller, B., Roelfs, K. U., Lassowskat, I., Finkemeier, I., et al. (2020). Comparative proteome and metabolome analyses of latex-exuding and non-exuding Taraxacum koksaghyz roots provide insights into laticifer biology. *J. Exp. Botany Oxford Univ. Press* 71 (4), 1278. doi: 10.1093/JXB/ERZ512

Björkman, O., and Demmig, B. (1987). Photon yield of O2 evolution and chlorophyll fluorescence characteristics at 77 K among vascular plants of diverse origins. *Planta. Springer-Verlag* 170 (4), 489–504. doi: 10.1007/BF00402983

Bollenbach, T. J., Schuster, G., and Stern, D. B. (2004). Cooperation of endo- and exoribonucleases in chloroplast mRNA turnover. *Prog. Nucleic Acid Res. Mol. Biol. Acad. Press Inc.* 78, 305–337. doi: 10.1016/S0079-6603(04)78008-3

Botella-Pavía, P., Besumbes, Ó., Phillips, M. A., Carretero-Paulet, L., Boronat, A., and Rodríguez-Concepción, M. (2004). Regulation of carotenoid biosynthesis in plants: evidence for a key role of hydroxymethylbutenyl diphosphate reductase in controlling the supply of plastidial isoprenoid precursors. *Plant J. John Wiley Sons Ltd.* 40 (2), 188–199. doi: 10.1111/J.1365-313X.2004.02198.X

Cheng, Y., Luo, J., Li, H., Wei, F., Zhang, Y., Jiang, H., et al. (2022). Identification of the WRKY gene family and characterization of stress-responsive genes in Taraxacum kok-saghyz Rodin. *Int. J. Mol. Sci.* 23 (18), 10270. doi: 10.3390/IJMS231810270

Chenge-Espinosa, M., Cordoba, E., Romero-Guido, C., Toledo-Ortiz, G., and León, P.. (2018). Shedding light on the methylerythritol phosphate (MEP)-pathway: long hypocotyl 5 (HY5)/phytochrome-interacting factors (PIFs) transcription factors modulating key limiting steps. *Plant J. John Wiley Sons Ltd.* 96 (4), 828–841. doi: 10.1111/TPJ.14071

Clissold, P. M., and Ponting, C. P. (2000). PIN domains in nonsense-mediated mRNA decay and RNAi. *Curr. Biol.* 10 (24), R888–R890. doi: 10.1016/S0960-9822(00)00858-7

Collart, M. A. (2016). 'The Ccr4-Not complex is a key regulator of eukaryotic gene expression', Wiley Interdisciplinary Reviews. *RNA. Wiley-Blackwell* 7 (4), 438. doi: 10.1002/WRNA.1332

Couturier, J., Chibani, K., Jacquot, J. P., and Rouhier, N. (2013). Cysteine-based redox regulation and signaling in plants. *Front. Plant Sci. Front. Res. Foundation* 4 (APR). doi: 10.3389/FPLS.2013.00105/BIBTEX

Croteau, R., Kutchan, T. M., and Lewis, N. G. (2000). "Natural products (Secondary metabolites)," in *Biochemistry & Molecular Biology of Plants*. Eds. B. Buchanan, W. Gruissem and R. Jones (Rockville, Md: American Society of Plant Physiologists), 1250–1318.

Cunningham, F. X., and Gantt, E. (1998). Genes and enzymes of carotenoid biosynthesis in plants. *Annu. Rev.* 49 (1), 557–583. doi: 10.1146/annurev.arplant.49.1.557

Deng, X.-W., Caspar, T., and Quail, P. H. (1991). cop1: a regulatory locus involved in light-controlled development and gene expression in Arabidopsis. *Genes & Development* 5 (7), 1172–1182.

Dreze, M., Carvunis, A.-R., Charloteaux, B., Galli, M., Pevzner, S. J., Tasan, M., et al. (2011). Evidence for network evolution in an Arabidopsis interactome map. *Science* 333 (6042), 601. doi: 10.1126/SCIENCE.1203877

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* 43 (W1), W389–W394. doi: 10.1093/NAR/GKV332

Epping, J., van Deenen, N., Niephaus, E., Stolze, A., Fricke, J., Huber, C., et al. (2015). A rubber transferase activator is necessary for natural rubber biosynthesis in dandelion. *Nat. Plants* 1 (5). doi: 10.1038/nplants.2015.48

Estévez, J. M., Cantero, A., Reindl, A., Reichler, S., and León, P. (2001). 1-deoxy-d-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem. Elsevier* 276 (25), 22901–22909. doi: 10.1074/JBC.M100854200

Falk, S., Bonneau, F., Ebert, J., Kögel, A., and Conti, E. (2017). Mpp6 incorporation in the nuclear exosome contributes to RNA channeling through the Mtr4 helicase. *Cell Rep. Cell Press* 20 (10), 2279–2286. doi: 10.1016/J.CELREP.2017.08.033

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J.A., Hofmann, K., et al. (2002). The PROSITE database, its status in 2002. *Nucleic Acids Res. Nucleic Acids Res.* 30 (1), 235–238. doi: 10.1093/NAR/30.1.235

Fatica, A., Tollervey, D., and Dlakić, M. (2004). PIN domain of Nob1p is required for D-site cleavage in 20S pre-rRNA. *RNA. Cold Spring Harbor Lab. Press* 10 (11), 1698–1701. doi: 10.1261/RNA.7123504

Frank, A., and Groll, M. (2017). The methylerythritol phosphate pathway to isoprenoids. *Chem. Rev. Am. Chem. Soc.* 117 (8), 5675–5703. doi: 10.1021/ACS.CHEMREV.6B00537/ASSET/IMAGES/MEDIUM/CR-2016-005373_0021.GIF

Galvão, V. C., and Fankhauser, C. (2015). Sensing the light environment in plants: photoreceptors and early signaling steps. *Curr. Opin. Neurobiol. Elsevier Curr. Trends* 34, 46–53. doi: 10.1016/J.CONB.2015.01.013

Genoud, T., Schweizer, F., Tscheuschler, A., Debrieux, D., Casal, J. J., Schäfer, E., et al. (2008). FHY1 mediates nuclear import of the light-activated phytochrome A photoreceptor. *PloS Genet. Public Library Sci.* 4 (8), e1000143. doi: 10.1371/JOURNAL.PGEN.1000143

Germain, A., Herlich, S., Larom, S., Kim, S. H., Schuster, G., Stern, D. B., et al. (2011). Mutational analysis of Arabidopsis chloroplast polynucleotide phosphorylase reveals roles for both RNase PH core domains in polyadenylation, RNA 3'-end maturation and intron degradation. *Plant J.: Cell Mol. Biol.* 67 (3), 381–394. doi: 10.1111/J.1365-313X.2011.04601.X

Germain, A., Kim, S. H., Gutierrez, R., and Stern, D. B.. (2012). Ribonuclease II preserves chloroplast RNA homeostasis by increasing mRNA decay rates, and cooperates with polynucleotide phosphorylase in 3′ end maturation. *Plant J. John Wiley Sons Ltd.* 72 (6), 960–971. doi: 10.1111/TPJ.12006

Gondet, L., Weber, T., Maillot-Vernier, P., Benveniste, P., and Bach, T. J. (1992). Regulatory role of microsomal 3-hydroxy-3-methylglutaryl-coenzyme a reductase in a tobacco mutant that overproduces sterols. *Biochem. Biophys. Res. Commun. Acad. Press* 186 (2), 888–893. doi: 10.1016/0006-291X(92)90829-A

Guidi, L., Brunetti, C., Fini, A., Agati, G., Ferrini, F., Gori, A., et al. (2016). UV radiation promotes flavonoid biosynthesis, while negatively affecting the biosynthesis and the de-epoxidation of xanthophylls: Consequence for photoprotection? *Environ. Exp. Botany Elsevier* 127, 14–25. doi: 10.1016/J.ENVEXPBOT.2016.03.002

Hamblin, J., Stefanova, K., and Angessa, T. T. (2014). Variation in chlorophyll content per unit leaf area in spring wheat and implications for selection in segregating material. *PloS One* 9 (3), 92529. doi: 10.1371/JOURNAL.PONE.0092529

Havaux, M., Dall'Osto, L., and Bassi, R. (2007). Zeaxanthin has enhanced antioxidant capacity with respect to all other xanthophylls in Arabidopsis leaves and functions independent of binding to PSII antennae. *Plant Physiol. Oxford Acad.* 145 (4), 1506–1520. doi: 10.1104/PP.107.108480

Hemmerlin, A., Harwood, J. L., and Bach, T. J. (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.* 51 (2), 95–148. doi: 10.1016/J.PLIPRES.2011.12.001

Holec, S., Lange, H., Kühn, K., Alioua, M., Börner, T., Gagliardi, D., et al. (2006). Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol. Cell. Biol. Taylor Francis* 26 (7), 2869. doi: 10.1128/MCB.26.7.2869-2876.2006

Huq, E., Al-Sady, B., Hudson, M., Kim, C., Apel, K., Quail, P. H., Al-Sady, B., Hudson, M., Kim, C., Apel, K., and Quail, P. H. (2004). Phytochrome-interacting factor 1 is a critical bHLH: Regulator of chlorophyll biosynthesis. *Sci. Am. Assoc. Advancement Sci.* 305 (5692), 1937–1941. doi: 10.1126/SCIENCE.1099728/SUPPL_FILE/HUQ.SOM.PDF

Ichikawa, T., Nakazawa, M., Kawashima, M., Iizumi, H., Kuroda, H., Kondou, Y., et al. (2006). The FOX hunting system: an alternative gain-of-function gene hunting technique. *Plant J. John Wiley Sons Ltd.* 48 (6), 974–985. doi: 10.1111/J.1365-313X.2006.02924.X

Iorizzo, M., Ellison, S., Senalik, D., Zeng, P., Satapoomin, P., Huang, J., et al. (2016). A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* 48, 6. doi: 10.1038/ng.3565

Jang, I. C., Yang, J. Y., Seo, H. S., and Chua, N. H. (2005). HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signaling. *Genes Dev. Cold Spring Harbor Lab. Press* 19 (5), 593–602. doi: 10.1101/GAD.1247205

Jing, Y., and Lin, R. (2020). Transcriptional regulatory network of the light signaling pathways. *New Phytologist. John Wiley Sons Ltd.* 227 (3), 683–697. doi: 10.1111/NPH.16602

Job, N., and Datta, S. (2021). PIF3/HY5 module regulates BBX11 to suppress protochlorophyllide levels in dark and promote photomorphogenesis in light. *New Phytologist. John Wiley Sons Ltd.* 230 (1), 190–204. doi: 10.1111/NPH.17149

Kami, C., Lorrain, S., Hornitschek, P., and Fankhauser, C. (2010). Light-regulated plant growth and development. *Curr. Topics Dev. Biol. Acad. Press Inc.* 91 (C), 29–66. doi: 10.1016/S0070-2153(10)91002-8

Keller, Y., Bouvier, F., d'Harlingue, A., and Camara, B.. (1998). Metabolic compartmentation of plastid prenyllipid biosynthesis. Evidence for the involvement of a multifunctional geranylgeranyl reductase. *Eur. J. Biochem. Wiley/Blackwell (10.1111)* 251 (1–2), 413–417. doi: 10.1046/j.1432-1327.1998.2510413.x

Kim, S., Schlicke, H., Van Ree, K., Karvonen, K., Subramaniam, A., Richter, A., et al. (2013). Arabidopsis chlorophyll biosynthesis: an essential balance between the methylerythritol phosphate and tetrapyrrole pathways. *Plant Cell. Oxford Univ. Press* 25 (12), 4984. doi: 10.1105/TPC.113.119172

Kim, J., Kang, H., Park, J., Kim, W., Yoo, J., Lee, N., et al. (2016). PIF1-interacting transcription factors and their binding sequence elements determine the in *vivo* targeting sites of PIF1. *Plant Cell* 28 (6), 1388–1405. doi: 10.1105/tpc.16.00125

Kim, K., Jeong, J., Kim, J., Lee, N., Kim, M. E., Lee, S., et al. (2016). PIF1 regulates plastid development by repressing photosynthetic genes in the endodermis. *Mol. Plant* 9, 1415–1427. doi: 10.1016/j.molp.2016.08.007

Kim, N., Jeong, J., Kim, J., Oh, J., and Choi, G. (2023). Shade represses photosynthetic genes by disrupting the DNA binding of GOLDEN2-LIKE1. *Plant Physiol. Oxford Acad.* 191 (4), 2334–2352. doi: 10.1093/PLPHYS/KIAD029

Kircher, S., Ledger, S., Hayashi, H., Weisshaar, B., Schäfer, E., and Frohnmeyer, H. (1998). CPRF4a, a novel plant bZIP protein of the CPRF family: Comparative analyses of light-dependent expression, post-transcriptional regulation, nuclear import and heterodimerisation. *Mol. Gen. Genet. Springer* 257 (6), 595–605. doi: 10.1007/S004380050687/METRICS

Kishimoto, S., and Ohmiya, A. (2006). Regulation of carotenoid biosynthesis in petals and leaves of chrysanthemum (Chrysanthemum morifolium). *Physiol. Plantarum. John Wiley Sons Ltd.* 128 (3), 436–447. doi: 10.1111/J.1399-3054.2006.00761.X

Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., and Penin, A. A. (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling', *The Plant journal: for cell and molecular biology. Plant J.* 88 (6), 1058–1070. doi: 10.1111/TPJ.13312

Laibach, N., Hillebrand, A., Twyman, R. M., Prüfer, D., and Schulze Gronover, C. (2015). Identification of a *Taraxacum brevicorniculatum* rubber elongation factor protein that is localized on rubber particles and promotes rubber biosynthesis. *Plant J. Wiley/Blackwell (10.1111)* 82 (4), 609–620. doi: 10.1111/tpj.12836

Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., et al. (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell. Oxford Univ. Press* 19 (3), 731. doi: 10.1105/TPC.106.047688

Leivar, P., Tepperman, J. M., Monte, E., Calderon, R. H., Liu, T. L., Quail, P. H., et al. (2009). Definition of early transcriptional circuitry involved in light-induced reversal of PIF-imposed repression of photomorphogenesis in young Arabidopsis seedlings. *Plant Cell. Oxford Acad.* 21 (11), 3535–3553. doi: 10.1105/TPC.109.070672

Lemaire, S. D., Guillont, B., Le Maréchal, P., Keryer, E., Miginiac-Maslow, M., Decottignies, P., et al. (2004). New thioredoxin targets in the unicellular photosynthetic eukaryote Chlamydomonas reinhardtii. *Proc. Natl. Acad. Sci. U. States A. Natl. Acad.*

*Sci.* 101 (19), 7475–7480. doi: 10.1073/PNAS.0402221101/SUPPL_FILE/02221FIG8.PDF

Li, Q. S., Das Gupta, J., and Hunt, A. G. (1998). Polynucleotide phosphorylase is a component of a novel plant Poly(A) polymerase. *J. Biol. Chem. Elsevier* 273 (28), 17539–17543. doi: 10.1074/JBC.273.28.17539

Li, Y., Ren, B., Gao, L., Ding, L., Jiang, D., Xu, X., et al. (2013). Less chlorophyll does not necessarily restrain light capture ability and photosynthesis in a chlorophyll-deficient rice mutant. *J. Agron. Crop Sci. John Wiley Sons Ltd.* 199 (1), 49–56. doi: 10.1111/J.1439-037X.2012.00519.X

Li, X., Liang, T., and Liu, H. (2022). How plants coordinate their development in response to light and temperature signals. *Plant Cell. Oxford Acad.* 34 (3), 955–966. doi: 10.1093/PLCELL/KOAB302

Liao, P., Hemmerlin, A., Bach, T. J., and Chye, M. L. (2016). The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnol. Adv. Elsevier* 34 (5), 697–713. doi: 10.1016/J.BIOTECHADV.2016.03.005

Lichtenthaler, H. K. (1987). Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Methods Enzymol. Acad. Press* 148 (C), 350–382. doi: 10.1016/0076-6879(87)48036-1

Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev.* 50, 47–65. doi: 10.1146/ANNUREV.ARPLANT.50.1.47

Lin, T., Xu, X., Du, H., Fan, X., Chen, Q., Hai, C., et al. (2018). Genome analysis of Taraxacum kok-saghyz Rodin provides new insights into rubber biosynthesis. *Natl. Sci. Rev.* 5 (1), 78–87. doi: 10.1093/nsr/nwx101

Lin, K., Zhao, H., Gan, S., and Li, G.. (2019). Arabidopsis ELF4-like proteins EFL1 and EFL3 influence flowering time. *Gene Elsevier* 700, 131–138. doi: 10.1016/J.GENE.2019.03.047

Lin, T., Xu, X., Ruan, J., Liu, S., Wu, S., Shao, X., et al. (2022). Extensive sequence divergence between the reference genomes of Taraxacum kok-saghyz and Taraxacum mongolicum. *Sci. China Life Sci. Sci. Press (China)* 65 (3), 515–528. doi: 10.1007/S11427-021-2033-2/METRICS

Mangus, D. A., Evans, M. C., and Jacobson, A. (2003). Poly(A)-binding proteins: Multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol. BioMed. Cent.* 4 (7), 1–14. doi: 10.1186/GB-2003-4-7-223/FIGURES/4

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., et al. (2013). CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* 41 (Database issue), D348–D352. doi: 10.1093/NAR/GKS1243

Moon, J., Zhu, L., Shen, H., and Huq, E. (2008). PIF1 directly and indirectly regulates chlorophyll biosynthesis to optimize the greening process in Arabidopsis. *Proc. Natl. Acad. Sci. U. States A. Natl. Acad. Sci.* 105 (27), 9433–9438. doi: 10.1073/PNAS.0803611105/SUPPL_FILE/0803611105SI.PDF

Müller, B., Noll, G. A., Ernst, A. M., Rüping, B., Groscurth, S., Twyman, R. M., et al. (2010). Recombinant artificial forisomes provide ample quantities of smart biomaterials for use in technical devices. *Appl. Microbiol. Biotechnol. Springer-Verlag* 88 (3), 689–698. doi: 10.1007/s00253-010-2771-4

Niephaus, E., Müller, B., van Deenen, N., Lassowskat, I., Bonin, M., Finkemeier, I., et al. (2019). Uncovering mechanisms of rubber biosynthesis in Taraxacum koksaghyz – role of cis-prenyltransferase-like 1 protein. *Plant J.* 100 (3), 591–609. doi: 10.1111/tpj.14471

Nowicka, B., and Kruk, J. (2010). Occurrence, biosynthesis and function of isoprenoid quinones. *Biochim. Biophys. Acta (BBA) Bioenergetics Elsevier* 1797 (9), 1587–1605. doi: 10.1016/J.BBABIO.2010.06.007

Ohmiya, A., Kishimoto, S., Aida, R., Yoshioka, S., and Sumitomo, K. (2006). Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. *Plant Physiol. Oxford Univ. Press* 142 (3), 1193. doi: 10.1104/PP.106.087130

Osterlund, M. T., Ang, L. H., and Deng, X. W. (1999). The role of COP1 in repression of Arabidopsis photomorphogenic development. *Trends Cell Biol. Elsevier Curr. Trends* 9 (3), 113–118. doi: 10.1016/S0962-8924(99)01499-3

Osterlund, M. T., and Deng, X. W. (1998). Multiple photoreceptors mediate the light-induced reduction of GUS-COP1 from Arabidopsis hypocotyl nuclei. *Plant J. John Wiley Sons Ltd.* 16 (2), 201–208. doi: 10.1046/J.1365-313X.1998.00290.X

Panara, F., Fasano, C., Lopez, L., Porceddu, A., Facella, P., Fantini, E., et al. (2022). Genome-wide identification and spatial expression analysis of histone modification gene families in the rubber dandelion Taraxacum kok-saghyz. *Plants Multidiscip. Digital Publishing Institute (MDPI)* 11 (16), 2077. doi: 10.3390/PLANTS11162077

Post, J., van Deenen, N., Fricke, J., Kowalski, N., Wurbs, D., Schaller, H., et al. (2012). Laticifer-specific cis-prenyltransferase silencing affects the rubber, triterpene, and inulin content of Taraxacum brevicorniculatum. *Plant Physiol. Am. Soc. Plant Biologists* 158 (3), 1406–1417. doi: 10.1104/pp.111.187880

Pütter, K. M., van Deenen, N., Müller, B., Fuchs, L., Vorwerk, K., Unland, K., et al (2017). Isoprenoid biosynthesis in dandelion latex is enhanced by the overexpression of three key enzymes involved in the mevalonate pathway', *BMC Plant Biology. BioMed. Cent.* 17 (1), 88. doi: 10.1186/s12870-017-1036-0

Pütter, K. M., van Deenen, N., Unland, K., Prüfer, D., and Schulze Gronover, C. (2019). The enzymes OSC1 and CYP716A263 produce a high variety of triterpenoids in the latex of Taraxacum koksaghyz. *Sci. Rep.* 9 (1), 1–13. doi: 10.1038/s41598-019-42381-w

Ramirez-Cadavid, D. A., Cornish, K., and Michel, F. C. (2017). Taraxacum kok-saghyz (TK): compositional analysis of a feedstock for natural rubber and other bioproducts. *Ind. Crops Prod. Elsevier B.V.* 107, 624–640. doi: 10.1016/J.INDCROP.2017.05.043

Rüdiger, W. (1993)Esterification of chlorophyllide and its implication for thylakoid development. In: *Pigment-protein complexes in plastids: synthesis and assembly* (San Diego: Academic Press). Available at: https://books.google.de/books?hl=de&lr=&id=gDDLBAAAQBAJ&oi=fnd&pg=PA219&dq=Rudiger,+W.++-++Esterification+of++chlorophyllide+and+its+implication+for++thylakoid++development.++In:++C.++Sundqvist++and++M.++Ryberg,++eds.++Pigment-protein+complexes+in++Plast (Accessed 8 October 2018)).

Ruiz-Sola, M. Á., Coman, D., Beck, G., Barja, M. V., Colinas, M., Graf, A., et al. (2016). *Arabidopsis* GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. *New Phytologist. Wiley/Blackwell (10.1111)* 209 (1), 252–264. doi: 10.1111/nph.13580

Saijo, Y., Sullivan, J. A., Wang, H., Yang, J., Shen, Y., Rubio, V., et al. (2003). The COP1–SPA1 interaction defines a critical step in phytochrome A-mediated regulation of HY5 activity. *Genes Dev. Cold Spring Harbor Lab. Press* 17 (21), 2642–2647. doi: 10.1101/GAD.1122903

Sauret-Güeto, S., Botella-Pavía, P., Flores-Pérez, U., Martínez-García, J. F., San Román, C., León, P., et al. (2006). Plastid cues posttranscriptionally regulate the accumulation of key enzymes of the methylerythritol phosphate pathway in Arabidopsis. *Plant Physiol. Oxford Univ. Press* 141 (1), 75. doi: 10.1104/PP.106.079855

Schmid, M., and Jensen, T. H. (2019). The nuclear RNA exosome and its cofactors. *Adv. Exp. Med. Biol. Adv. Exp. Med. Biol.* 1203, 113–132. doi: 10.1007/978-3-030-31434-7_4

Schmidt, F. J., Zimmermann, M. M., Wiedmann, D. R., Lichtenauer, S., Grundmann, L., Muth, J., et al. (2020). The major floral promoter ntFT5 in tobacco (Nicotiana tabacum) is a promising target for crop improvement. *Front. Plant Sci.* 10. doi: 10.3389/FPLS.2019.01666/FULL

Senissar, M., Manav, M. C., and Brodersen, D. E. (2017). Structural conservation of the PIN domain active site across all domains of life. *Protein Sci.: A Publ. Protein Society Wiley-Blackwell* 26 (8), 1474. doi: 10.1002/PRO.3193

Shahmuradov, I. A., and Solovyev, V. V. (2015). Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements. *Bioinf. Oxford Univ. Press* 31 (21), 3544. doi: 10.1093/BIOINFORMATICS/BTV404

Shin, D. H., Choi, M., Kim, K., Bang, G., Cho, M., Choi, S. B., et al. (2013). HY5 regulates anthocyanin biosynthesis by inducing the transcriptional activation of the MYB75/PAP1 transcription factor in Arabidopsis. *FEBS Lett. No Longer Published by Elsevier* 587 (10), 1543–1547. doi: 10.1016/J.FEBSLET.2013.03.037

Shumskaya, M., and Wurtzel, E. T. (2013). THE CAROTENOID BIOSYNTHETIC PATHWAY: THINKING IN ALL DIMENSIONS. *Plant Sci.: an Int. J. Exp. Plant Biol. NIH Public Access* 208, 58. doi: 10.1016/J.PLANTSCI.2013.03.012

Sprenger-Haussels, M., and Weisshaar, B. (2000). Transactivation properties of parsley proline-rich bZIP transcription factors. *Plant J. John Wiley Sons Ltd.* 22 (1), 1–8. doi: 10.1046/J.1365-313X.2000.00687.X

Stolze, A., Wanke, A., van Deenen, N., Geyer, R., Prüfer, D., Schulze Gronover, C., et al. (2017). Development of rubber-enriched dandelion varieties by metabolic engineering of the inulin pathway. *Plant Biotechnol. J. Wiley-Blackwell* 15 (6), 740–753. doi: 10.1111/pbi.12672

Subramanian, C., Kim, B. H., Lyssenko, N. N., Xu, X., Johnson, C. H., and Von Arnim, A. G. (2004). The Arabidopsis repressor of light signaling, COP1, is regulated by nuclear exclusion: Mutational analysis by bioluminescence resonance energy transfer. *Proc. Natl. Acad. Sci. Natl. Acad. Sci.* 101 (17), 6798–6802. doi: 10.1073/PNAS.0307964101

Surmacz, L., and Swiezewska, E. (2011). Polyisoprenoids – Secondary metabolites or physiologically important superlipids? *Biochem. Biophys. Res. Commun. Acad. Press* 407 (4), 627–632. doi: 10.1016/J.BBRC.2011.03.059

Surpin, M., Larkin, R. M., and Chory, J. (2002). Signal transduction between the chloroplast and the nucleus. *Plant Cell. Oxford Acad.* 14 (suppl_1), S327–S338. doi: 10.1105/TPC.010446

Suzuki, M., Kamide, Y., Nagata, N., Seki, H., Ohyama, K., Kato, H., et al. (2004). Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in Arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J.: Cell Mol. Biol. Plant J.* 37 (5), 750–761. doi: 10.1111/J.1365-313X.2004.02003.X

Tang, W., Wang, W., Chen, D., Ji, Q., Jing, Y., Wang, H., et al. (2012). Transposase-derived proteins FHY3/FAR1 interact with PHYTOCHROME-INTERACTING FACTOR1 to regulate chlorophyll biosynthesis by modulating HEMB1 during deetiolation in Arabidopsis. *Plant Cell. Am. Soc. Plant Biologists* 24 (5), 1984–2000. doi: 10.1105/TPC.112.097022/DC1

Thain, S. C., Murtas, G., Lynn, J. R., McGrath, R. B., and Millar, A. J. (2002). The circadian clock that controls gene expression in Arabidopsis is tissue specific. *Plant Physiol. Oxford Univ. Press* 130 (1), 102. doi: 10.1104/PP.005405

Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng./Biotechnol. Springer Sci. Business Media Deutschland GmbH* 148, 63–106. doi: 10.1007/10_2014_295/COVER

Toledo-Ortiz, G., Huq, E., and Quail, P. H. (2014). The HY5-PIF regulatory module coordinates light and temperature control of photosynthetic gene transcription. *PloS Genet. Public Library Sci.* 10 (6), e1004416. doi: 10.1371/JOURNAL.PGEN.1004416

Toledo-Ortiz, G., Huq, E., and Quail, P. H. (2003). The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell. Oxford Acad.* 15 (8), 1749–1770. doi: 10.1105/TPC.013839

Toledo-Ortiz, G., Huq, E., and Rodríguez-Concepción, M. (2010). Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proc. Natl. Acad. Sci. U. States A. Natl. Acad. Sci.* 107 (25), 11626–11631. doi: 10.1073/PNAS.0914428107/SUPPL_FILE/PNAS.200914428SI.PDF

Ulmann, M. (1951) *Wertvolle Kautschukpflanzen des gemässigten Klimas. Akademie-Verlag*. Available at: https://books.google.de/books/about/Wertvolle_Kautschukpflanzen_des_gem%C3%A4ssi.html?id=wih-AAAAIAAJ&redir_esc=y (Accessed 18 April 2023).

Unland, K., Pütter, K. M., Vorwerk, K., van Deenen, N., Twyman, R. M., Prüfer, D., et al. (2018). Functional characterization of squalene synthase and squalene epoxidase in Taraxacum koksaghyz. *Plant Direct. Wiley-Blackwell* 2 (6), 1–15. doi: 10.1002/pld3.63

Vaňáčová, Š., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., et al. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PloS Biol. Public Library Sci.* 3 (6), e189–e189. doi: 10.1371/JOURNAL.PBIO.0030189

van Beilen, J. B., and Poirier, Y. (2007). Guayule and Russian dandelion as alternative sources of natural rubber. *Crit. Rev. Biotechnol. Taylor Francis* 27 (4), 217–231. doi: 10.1080/07388550701775927

van Gelder, K., Rea, K. A., Virta, L. K.A., Whitnell, K. L., Osborn, M., Vatta, M., et al. (2018). Medium-chain polyprenols influence chloroplast membrane dynamics in Solanum lycopersicum. *Plant Cell Physiol. Oxford Acad.* 59 (11), 2350–2365. doi: 10.1093/PCP/PCY157

Vranová, E., COman, D., and Gruissem, W. (2012). Structure and dynamics of the isoprenoid pathway network. *Mol. Plant Elsevier* 5 (2), 318–333. doi: 10.1093/mp/sss015

Walter, M., Kilian, J., and Kudla, J. (2002). 'PNPase activity determines the efficiency of mRNA 3′-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. *EMBO J. John Wiley Sons Ltd.* 21 (24), 6905–6914. doi: 10.1093/EMBOJ/CDF686

Wang, L., Wang, J., Chen, H., and Hu, B.. (2022). Genome-wide identification, characterization, and functional analysis of lncRNAs in Hevea brasiliensis. *Front. Plant Sci. Front. Plant Sci.* 13. doi: 10.3389/FPLS.2022.1012576

Waters, M. T., Wang, P., Korkaric, M., Capper, R. G., Saunders, N. J., and Langdale, J. A. (2009). GLK transcription factors coordinate expression of the photosynthetic apparatus in Arabidopsis. *Plant Cell. Am. Soc. Plant Biologists* 21 (4), 1109–1128. doi: 10.1105/TPC.108.065250

Xie, Q., Ding, G., Zhu, L., Yu, L., Yuan, B., Gao, X., et al. (2019). Proteomic landscape of the mature roots in a rubber-producing grass Taraxacum Kok-saghyz. *Int. J. Mol. Sci. Multidiscip. Digital Publishing Institute (MDPI)* 20 (10), 1*–19. doi: 10.3390/IJMS20102596

Xu, D. (2019). COP1 and BBXs-HY5-mediated light signal transduction in plants. *New Phytologist. John Wiley Sons Ltd.* 228 (6), 1748–1753. doi: 10.1111/NPH.16296

Xu, X., Paik, I., Zhu, L., Bu, Q., Huang, X., Deng, X. W., et al. (2014). PHYTOCHROME INTERACTING FACTOR1 enhances the E3 ligase activity of CONSTITUTIVE PHOTOMORPHOGENIC1 to synergistically repress photomorphogenesis in Arabidopsis. *Plant Cell. Oxford Univ. Press* 26 (5), 1992. doi: 10.1105/TPC.114.125591

Yazaki, K., Arimura, G. I., and Ohnishi, T. (2017). "Hidden" terpenoids in plants: Their biosynthesis, localization and ecological roles. *Plant Cell Physiol.* 58 (10), 1615–1621. doi: 10.1093/pcp/pcx123

Yehudai-Resheff, S., Hirsh, M., and Schuster, G. (2001). Polynucleotide phosphorylase functions as both an exonuclease and a poly(A) polymerase in spinach chloroplasts'. *Mol. Cell. Biol. Taylor Francis* 21 (16), 5408. doi: 10.1128/MCB.21.16.5408-5416.2001

Zhang, Y., Mayba, O., Pfeiffer, A., Shi, H., Tepperman, J. M., Speed, T. P., et al. (2013). A Quartet of PIF bHLH Factors Provides a Transcriptionally Centered Signaling Hub That Regulates Seedling Morphogenesis through Differential Expression-Patterning of Shared Target Genes in Arabidopsis. *PloS Genet. Public Library Sci.* 9 (1), e1003244. doi: 10.1371/JOURNAL.PGEN.1003244

Zhang, C., Zhang, J., Tang, Y., Liu, K., Liu, Y., Tang, J., et al. (2021). DEEP GREEN PANICLE1 suppresses GOLDEN2-LIKE activity to reduce chlorophyll synthesis in rice glumes. *Plant Physiol. Oxford Univ. Press* 185 (2), 469. doi: 10.1093/PLPHYS/KIAA038

Zhang, Z., Shen, G., Yang, Y., Li, C., Chen, X., Yang, X., et al. (2022). 'Metabolic and transcriptomic analyses reveal the effects of ethephon on Taraxacum kok-saghyz Rodin. *Molecules* 27 (11), 3548. doi: 10.3390/MOLECULES27113548

Zuk, M., Działo, M., Richter, D., Dymińska, L., Matuła, J., Kotecki, A., et al. (2016). Chalcone synthase (CHS) gene suppression in flax leads to changes in wall synthesis and sensing genes, cell wall chemistry and stem morphology parameters. *Front. Plant Sci. Front. Res. Foundation* 7. doi: 10.3389/FPLS.2016.00894/BIBTEX

# Frontiers in
# Plant Science

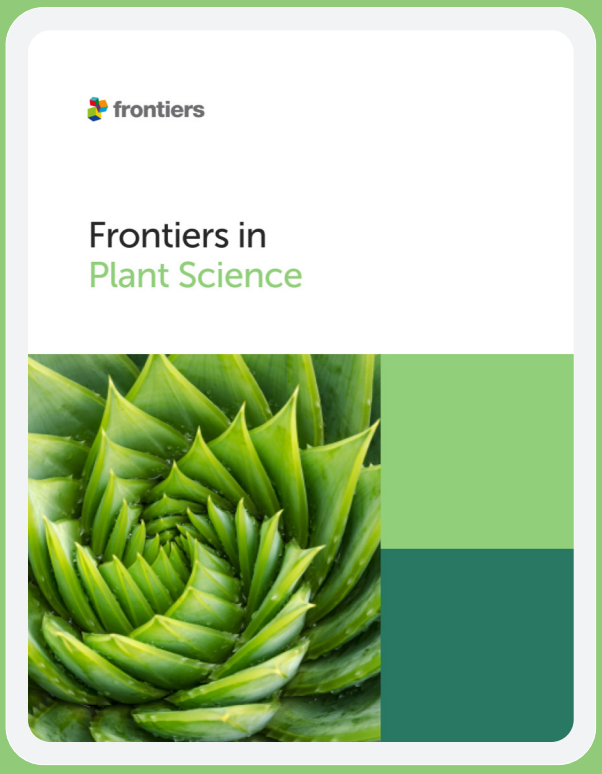
**Cultivates the science of plant biology and its applications**

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