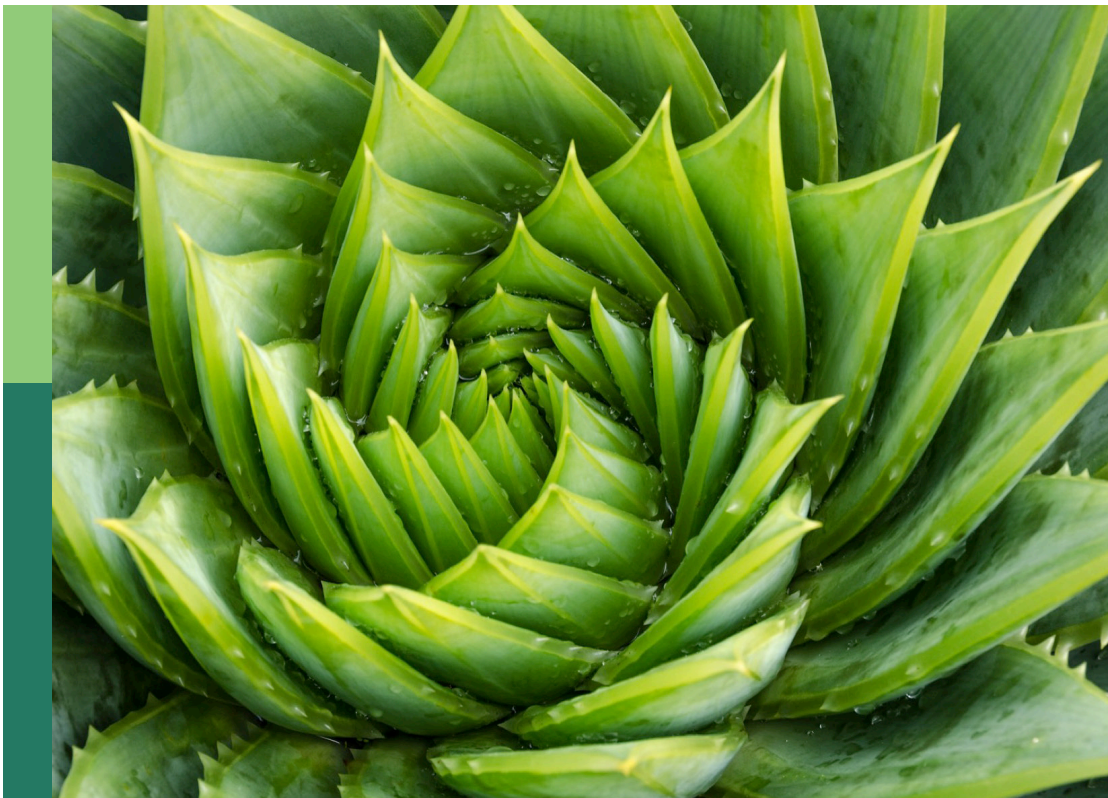
frontiers | Research Topics

Orchid genomics and developmental biology, volume II

Edited by
Jen-Tsung Chen and Katharina Nargar

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-83251-498-6
DOI 10.3389/978-2-83251-498-6

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

Orchid genomics and developmental biology, volume II

Topic editors

Jen-Tsung Chen — National University of Kaohsiung, Taiwan

Katharina Nargar — Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

Citation

Chen, J.-T., Nargar, K., eds. (2023). *Orchid genomics and developmental biology, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-498-6

Table of contents

05 Editorial: Orchid genomics and developmental biology, volume II
Katharina Nargar and Jen-Tsung Chen

08 Comparative Transcriptomic and Metabolic Analyses Reveal the Molecular Mechanism of Ovule Development in the Orchid, *Cymbidium sinense*
Danqi Zeng, Caixia Que, Jaime A. Teixeira da Silva, Shutao Xu and Dongmei Li

25 Organ-Specific Gene Expression Reveals the Role of the *Cymbidium ensifolium-miR396*/*Growth-Regulating Factors* Module in Flower Development of the Orchid Plant *Cymbidium ensifolium*
Fengxi Yang, Chuqiao Lu, Yonglu Wei, Jieqiu Wu, Rui Ren, Jie Gao, Sagheer Ahmad, Jianpeng Jin, Yechun Xv, Gang Liang and Genfa Zhu

40 Anthocyanin and Flavonol Glycoside Metabolic Pathways Underpin Floral Color Mimicry and Contrast in a Sexually Deceptive Orchid
Darren C. J. Wong, James Perkins and Rod Peakall

58 Characterization of Three *SEPALLATA*-Like MADS-Box Genes Associated With Floral Development in *Paphiopedilum henryanum* (Orchidaceae)
Hao Cheng, Xiulan Xie, Maozhi Ren, Shuhua Yang, Xin Zhao, Nasser Mahna, Yi Liu, Yufeng Xu, Yukai Xiang, Hua Chai, Liang Zheng, Hong Ge and Ruidong Jia

72 Orchid Phylotranscriptomics: The Prospects of Repurposing Multi-Tissue Transcriptomes for Phylogenetic Analysis and Beyond
Darren C. J. Wong and Rod Peakall

81 Developmental Characteristics and Auxin Response of Epiphytic Root in *Dendrobium catenatum*
Jili Tian, Weiwei Jiang, Jinping Si, Zhigang Han, Cong Li and Donghong Chen

96 The Transcriptome Profiling of Flavonoids and Bibenzyls Reveals Medicinal Importance of Rare Orchid *Arundina graminifolia*
Sagheer Ahmad, Jie Gao, Yonglu Wei, Chuqiao Lu, Genfa Zhu and Fengxi Yang

108 Transcriptional Proposition for Uniquely Developed Protocorm Flowering in Three Orchid Species: Resources for Innovative Breeding
Sagheer Ahmad, Jinliao Chen, Guizhen Chen, Jie Huang, Yang Hao, Xiaoling Shi, Yuying Liu, Song Tu, Yuzhen Zhou, Kai Zhao, Siren Lan, Zhongjian Liu and Donghui Peng

121 Recent Acquisition of Functional m6A RNA Demethylase Domain in Orchid Ty3/Gypsy Elements
Luis Alvarado-Marchena, Mireya Martínez-Pérez, Frederic Aparicio, Vicente Pallas and Florian Maumus

135 Genome-wide identification of *Cymbidium sinense* WRKY gene family and the importance of its Group III members in response to abiotic stress
Yong-Lu Wei, Jian-Peng Jin, Di Liang, Jie Gao, Jie Li, Qi Xie, Chu-Qiao Lu, Feng-Xi Yang and Gen-Fa Zhu

150 Conserved pigment pathways underpin the dark insectiform floral structures of sexually deceptive *Chiloglottis* (Orchidaceae)
Darren C. J. Wong, James Perkins and Rod Peakall

157 Physiological and transcriptomic analysis uncovers salinity stress mechanisms in a facultative crassulacean acid metabolism plant *Dendrobium officinale*
Mingze Zhang, Nan Liu, Jaime A. Teixeira da Silva, Xuncheng Liu, Rufang Deng, Yuxian Yao, Jun Duan and Chunmei He

179 *De novo* transcriptome based insights into secondary metabolite biosynthesis in *Malaxis acuminata* (*Jeevak*)–A therapeutically important orchid
Paromik Bhattacharyya, Tanvi Sharma, Abhinandan Yadav, Lucy Lalthafamkimi, Ritu, Mohit Kumar Swarnkar, Robin Joshi, Ravi Shankar and Sanjay Kumar

194 Genome-wide identification of YABBY genes in three *Cymbidium* species and expression patterns in *C. ensifolium* (Orchidaceae)
Qian-Qian Wang, Yuan-Yuan Li, Jiating Chen, Meng-Jia Zhu, Xuedie Liu, Zhuang Zhou, Diyang Zhang, Zhong-Jian Liu and Siren Lan

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 16 January 2023
DOI 10.3389/fpls.2023.1136350

OPEN ACCESS

EDITED AND REVIEWED BY
Paromik Bhattacharyya,
Institute of Himalayan Bioresource Technology (CSIR), India

*CORRESPONDENCE
Jen-Tsung Chen
jentsung@nuk.edu.tw

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

RECEIVED 03 January 2023
ACCEPTED 09 January 2023
PUBLISHED 16 January 2023

CITATION
Nargar K and Chen J-T (2023) Editorial: Orchid genomics and developmental biology, volume II.
Front. Plant Sci. 14:1136350.
doi: 10.3389/fpls.2023.1136350

Editorial: Orchid genomics and developmental biology, volume II

Katharina Nargar[1,2] and Jen-Tsung Chen[3]*

[1]Australian Tropical Herbarium, James Cook University, Cairns, QLD, Australia, [2]National Research Collections Australia, Commonwealth Industrial and Scientific Research Organisation (CSIRO), Canberra, ACT, Australia, [3]Department of Life Sciences, National University of Kaohsiung, Kaohsiung, Taiwan

KEYWORDS

orchid, biotechnology, functional genomics, developmental biology, omics

Editorial on the Research Topic
Orchid genomics and developmental biology, volume II

Orchidaceae constitute the second-largest flowering plant family worldwide with over 27,000 species found on all continents except Antarctica. Orchids are frequently celebrated for their exceptional morphological and ecological diversity and are highly valued in the horticultural trade. Orchids exhibit distinct floral and physiological features, such as fused male and female flower parts forming the gynostemium, a floral lip often adorned with calli, glands, spurs, and distinctive color patterns, and the crassulacean acid metabolism (CAM), a water-saving physiological pathway which has evolved multiple times independently within the family. Orchids possess highly specialized ecological relationships, such as often species-specific plant-pollinator interactions including food- and sexual deception and dependence on mycorrhizal fungi for germination of their minute seed. The broad range of traits renders orchids prime non-model plants for elucidating the genomic underpinnings and regulatory networks responsible for the generation of this exceptional diversity. To keep exploring such an interesting research field, this volume continues our previous Research Topic entitled "Orchid Genomics and Developmental Biology" (Chen and Nargar) to showcase recent findings and providing novel insights into regulatory mechanisms underpinning orchid biologies, such as reproductive development, responses to biotic and abiotic stresses, visual mimicry of orchid flowers, and bioactive metabolic pathways.

Regulatory mechanisms underpinning reproductive development in orchids

Ovule development plays a vital role in plant reproduction and seed development. However, the regulatory mechanism underpinning ovule development in orchids is poorly understood. Zeng et al. present a comparative transcriptomic and metabolomic study of ovules of different developmental stages in *Cymbidium sinense*. Among 9845 differentially expressed unigenes (DEUs), the team identified candidate genes involved in ovule development, such as homeobox and MADS-box transcription factors, and phytohormones, such as cytokinin, gibberellin, and abscisic acid.

SEPPALATA-like MADS-box genes encode transcription factors which are responsible for floral organ specification. Cheng et al. investigated the role of *SEPPALATA*-like (*SEP*-like) MADS-box genes in the flower development of venus slipper orchids (*Paphiopedilum*). The

team identified three *SEP*-like genes with phylogenetic placement in the core eudicot *SEP3* lineage. Protein-protein interactions indicated that the *SEP*-like genes may interact with B-class and E-class proteins, providing further insights into the role of *SEP*-like genes in the floral development of orchids.

Genes encoding for the YABBY transcription factor family play significant roles in lateral organ development, such as cotyledons and floral organs. Based on genomic sequence data, Wang et al. identified and characterized 24 *YABBY* genes in three ornamental *Cymbidium* species. Expression analysis showed that *YAB2* genes were expressed more strongly in floral organs than in vegetative tissue. The authors identified two *YABBY* genes which were mainly expressed in the gynostemium.

The gynostemium is a remarkably complex structure of fused female and male organs. However, the multi-level regulatory networks in its formation are largely unknown. Yang et al. examined the role of microRNA in flower formation in *Cymbidium ensifolium*. The team found that a distinct microRNA (*Ce*-miRNA396) silenced Growth-Regulating Factors (GRFs) through cleavage. The team showed that GRF transcripts accumulated most in floral tissues where the miR969 concentration was lowest, and found a strong correlation between *Ce*-miRNA396, floral formation, and column specification.

Many orchids of horticultural interest require several years from propagation *via* seeds to flowering, hence shortening this time span is a key desideratum. Ahmad et al. investigated the regulatory mechanism of the curious phenomenon of flower formation in protocorms without prior formation of leaves or roots in three *Cymbidium* species. Through comparative transcriptome analysis, the team identified transcription factor (TF) families and candidate key TFs involved in the regulatory networks that govern the onset of the reproductive stage.

Underpinnings of visual mimicry in sexually-deceptive orchids

Pollination by sexual deception has evolved multiple times independently in Orchidaceae, with floral color playing a key role in visual mimicry. Wong et al. investigated the chemical composition and genetic regulatory networks underpinning floral coloration in a sexually deceptive bird orchid (*Chiloglottis trapeziformis*). The study elucidated the complex tissue-specific regulation of genes and biochemical pathways, in particular of the anthocyanin and flavonol glycoside metabolic pathways, across different stages of flower development.

In a related study, Wong et al. examined the chemical and genetic basis for floral coloration across the genus *Chiloglottis*. Phylogenomic analysis resolved three main evolutionary lineages, with the Formicifera clade sister to the Reflexa clade, and the two in turn sister to the Valida clade. While the biochemical basis of the distinct flower coloration underlying the floral mimicry was found to be conserved within the genus, biochemistry and gene expression levels were more similar among the two more closely related Formicifera and Reflexa lineages compared to the more distantly related Valida clade.

Regulatory mechanisms of bioactive metabolic pathways in orchids

Orchids used in traditional medicine are valued for their bioactive compounds. To alleviate anthropogenic pressure on wild orchids, *in vitro* culture, such as plant tissue culture, is increasingly established. To facilitate the selective enhancement of bioactive metabolites through *in vitro* systems, knowledge about the regulatory mechanism in the biosynthetic pathways is key. Bhattacharyya et al. investigated the metabolic pathways involved in the biosynthesis of key secondary metabolites in *Malaxis acuminata*, a threatened orchid valued in traditional medicine. The study provided insights into the regulatory pathways for the phytosterol β-sitosterol and the phenylpropanoids eugenol and isoeugenol. The study identified leaves as a potential alternate source for the production of bioactive metabolites to facilitate the sustainable use of this threatened species.

In another study on bioactive compounds in orchids, Ahmad et al. investigated the regulatory pathways of flavonoids and bibenzyls in *Arundina graminifolia*. The team identified candidate genes involved in the biosynthesis pathways of these bioactive compounds, including *BIBSY212*, *CYP84A1*, *CYP73A4*, *4CLL7*, *UGT88B1*, *UGT73C3*, *ANS*, *PAL*, *FLS*, and *CHS8*. Most of the candidate genes were expressed highest in leaves and roots. The concentration of phenylpropanoids was found to be highest in leaves, flavonoids in stems, and bibenzyl in leaves.

Mechanisms underlying abiotic stress responses in epiphytic orchids

Zhang et al. investigated salinity stress responses in the epiphytic orchid and facultative CAM plant, *Dendrobium officinale*. The team showed that plants exposed to different levels of salinity stress in their roots responded with changes in the expression of genes related to hormone biosynthesis and response, amino acid and flavonoid metabolism, and the Salt Overly Sensitive (SOS) pathway. Key candidate genes playing a role in salt stress response in *D. officinale* were identified.

To investigate the morphological plasticity of aerial roots and associated regulatory networks in epiphytic orchids, Tian et al. undertook a detailed study of roots under terrestrial, epiphytic, and lithophytic growth conditions and different auxin treatments in *Dendrobium officinale* and *D. catenatum*. Differential expression analysis identified genes associated with the promotion of root elongation growth, which included upregulated auxin transporters and cellulose synthetase genes under low auxin levels. Genes associated with cell proliferation under high auxin levels included transportation and signal transduction pathways and stem cell control and regeneration pathway-related genes. WUSCHEL-related homeobox transcription factor WOX12 from *D. catenatum* was found to confer highly efficient pluripotency acquisition properties, relevant to monocot plant transformation such as in orchids.

Transcription factors of the WRKY family play important roles in plant responses to biotic and abiotic stresses and secondary metabolism. Wei et al. studied WRKY transcription factors in *Cymbidium sinensis*. Among 64 *WRKY* genes, the team identified

key candidate genes, in particular in GROUP III, which were strongly induced in response to hormone treatments, indicating their potentially essential role in hormone signaling. The transcription factor CsWRKY18 was found to be associated with increasing plant tolerance to abiotic stress within the abscisic-acid- (ABA) dependent pathway.

Novel properties in transposable elements in orchids

Transposable elements (TE) can have profound impacts on the host's genomes, e.g., through increasing genome size. During their evolution, TE superfamilies have evolved significant changes in their architecture which can have profound impacts on their interaction with host organisms. Alvarado-Marchena et al. mined angiosperm genomes for new conserved protein domains within long terminal repeat (LTR) retrotransposons. The team discovered an additional open reading frame in Gypsy-type elements in *Phalaenopsis* orchids with similar properties to m6A RNA demethylase in AlkB proteins. The study demonstrated its RNA binding capacity and demethylase activity which may convey LTR retrotransposons increased fitness.

Repurposing genomic resources in orchids

Transcriptome data for non-model organisms such as orchids has seen a massive increase and presents a valuable data resource. Wong and Peakall explored the use of multi-tissue transcriptomes to infer phylogenetic relationships in Orchidaceae at both deep and shallow evolutionary scales. The phylotranscriptomic approach yielded largely consistent findings with other phylogenomic studies. For the phylogenetic placement of mycoheterotrophic species, which can undergo severe plastome degeneration, phylotranscriptomics yielded results consistent with large-scale nuclear phylogenomic studies. The authors discuss other potential uses of genomic resources, including the mining of genomes and transcriptomes for single-copy gene sets of bespoke target capture bait kits.

Author contributions

KN and J-TC drafted the manuscript. Both authors contributed to the article and approved the submitted version.

Acknowledgments

We greatly appreciate the invaluable contributions of all authors, reviewers, and Dr. Neelima Roy Sinha, Specialty Chief Editor for Plant Development and EvoDevo of Frontiers in Plant Science.

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

frontiers in Plant Science

ORIGINAL RESEARCH
published: 21 January 2022
doi: 10.3389/fpls.2021.814275

Comparative Transcriptomic and Metabolic Analyses Reveal the Molecular Mechanism of Ovule Development in the Orchid, *Cymbidium sinense*

Danqi Zeng[1,3†], Caixia Que[2†], Jaime A. Teixeira da Silva[4], Shutao Xu[5] and Dongmei Li[2]*

[1] Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [2] Guangdong Provincial Research Center for Standardization of Production Engineering Technology of Orchids, Shunde Polytechnic, Foshan, China, [3] College of Life Sciences, University of the Chinese Academy of Sciences, Beijing, China, [4] Independent Researcher, Kagawa, Japan, [5] College of Innovative Design, City University of Macau, Taipa, Macao SAR, China

OPEN ACCESS

Edited by:
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

Reviewed by:
Surendra Sarsaiya,
Zunyi Medical University, China
Ali Raza,
Fujian Agriculture and Forestry University, China

***Correspondence:**
Dongmei Li
10125@sdpt.edu.cn

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: 13 November 2021
Accepted: 27 December 2021
Published: 21 January 2022

Citation:
Zeng D, Que C, Teixeira da Silva JA, Xu S and Li D (2022) Comparative Transcriptomic and Metabolic Analyses Reveal the Molecular Mechanism of Ovule Development in the Orchid, Cymbidium sinense. Front. Plant Sci. 12:814275. doi: 10.3389/fpls.2021.814275

Ovule development is pivotal to plant reproduction and seed development. *Cymbidium sinense* (Orchidaceae) has high ornamental value due to its pleasant aroma and elegant floral morphology. The regulatory mechanism underlying ovule development in orchids, especially *C. sinense*, is largely unknown and information on the *C. sinense* genome is very scarce. In this study, a combined analysis was performed on the transcriptome and non-targeted metabolomes of 18 *C. sinense* 'Qi Jian Hei Mo' ovule samples. Transcriptome analysis assembled gene-related information related to six growth stages of *C. sinense* ovules (S1-S6, equivalent to 30, 35, 42, 46, 53, and 60 days after pollination). Illumina sequencing technology was used to obtain the complete set of transcriptome sequences of the 18 samples. A total of 81,585 unigene sequences were obtained after assembly, 24,860 (30.47%) of which were functionally annotated. Using transcriptome sequencing technology, a total of 9845 differentially expressed unigenes (DEUs) were identified in *C. sinense* ovules that were assigned to specific metabolic pathways according to the Kyoto Encyclopedia of Genes and Genomes (KEGG). DEUs associated with transcription factors (TFs) and phytohormones were identified and analyzed. The TFs homeobox and MADS-box were associated with *C. sinense* ovule development. In particular, the phytohormones associated with DEUs such as indole-3-acetic acid (IAA), cytokinin (CK), gibberellin (GA), abscisic acid (ABA), brassinosteroid (BR), and jasmonate (JA), may have important regulatory effects on *C. sinense* ovule development. Metabolomic analysis showed an inconsistent number of KEGG annotations of differential metabolites across comparisons (S2_vs_S4, S2_vs_S5, and S4_vs_S5 contained 23, 26, and 3 annotations, respectively) in *C. sinense* ovules. This study provides a valuable foundation for further understanding the regulation of orchid ovule development and formation, and establishes a theoretical background for future practical applications during orchid cultivation.

Keywords: ovule development, differentially expressed unigenes, plant hormones, *Cymbidium sinense*, metabolomics

INTRODUCTION

Cymbidium sinense (Orchidaceae) is an economically important flowering orchid with high ornamental value, elegant color, and delicate scent (Huang and Dai, 1998; Kim et al., 2016). Members of the Orchidaceae not only account for a high proportion of all flowering plants, they also have highly diversified and specialized flower morphology. Extensive research has been conducted on the molecular regulation and genetic regulation of floral development in orchids (Xu et al., 2006; Aceto and Gaudio, 2011; Zhang et al., 2013; Teixeira da Silva et al., 2014). The Orchidaceae has a unique flower type with a characteristic column (Rudall and Bateman, 2002). Pollinia are found at the top of the column while the ovary, which develops post-pollination, is found at the bottom of it (Yu and Goh, 2001). The ovule develops into a seed after fertilization and ensures normal reproduction and the production of offspring (Schneitz, 1999), which is regarded as an important biological process during plant development.

Although the ovule in animals is sometimes considered as the ovum or egg cell, the ovule of angiosperms (seed plants) can have a variety of roles throughout development given its complex structure, including integument(s), the micropyle and nucellus (Rudall, 2021). In Arabidopsis (*Arabidopsis thaliana*), ovules are initiated from the carpels during flower development (Smyth et al., 1990). There are also numerous reports on the regulation of ovule growth and development in Arabidopsis (Gross-Hardt et al., 2002; Sieber et al., 2004; Colombo et al., 2008; Liu et al., 2019), such as the INO transcription factor (TF), which regulates the distal-proximal mode of the ovule (Villanueva et al., 1999). Ovule development is also very complex in rice (*Oryza sativa*): research focused on the morphology and cellular characteristics of ovules (Lopez-Dee et al., 1999; Locato and De Gara, 2018) and on the regulatory mechanism of development, noting that plant hormones play a vital role in embryonic development (Bencivenga et al., 2011). Separately, sequencing revealed the developmental process of rice ovules at the gene expression level (Kubo et al., 2013; Wu et al., 2015, 2017). In the Bromeliaceae, ovule-specific features have been described in detail (Fagundes and de Araujo Mariath, 2014; Mendes et al., 2014), as well as embryological characteristics of the ovule (Sajo et al., 2004), ovule morphology (Kuhn et al., 2016), and other aspects related to the ovule (Novikoff and Odintsova, 2008; Nogueira et al., 2015).

Some research on ovule development has been conducted in the Orchidaceae. Early studies identified genes expressed predominantly in the ovules of *Phalaenopsis* (Nadeau et al., 1996). The functional genes related to ovule development in *Phalaenopsis* and *Dendrobium* have been characterized, with C- and D-class MADS-box genes, such as *PeMADS1* and *DOAG2*, respectively, playing an important role in ovule development in these two orchid genera (Chen et al., 2012; Wang et al., 2020). Other studies found an increasingly important role of plant hormones in ovule development in the Orchidaceae (Nadeau et al., 1993; Ketsa and Rugkong, 2000; Tsai et al., 2008), in particular, critical regulatory role of ethylene in ovary maturation and differentiation. Compared with other orchids, very little research has been conducted on the regulation of ovule development in *Cymbidium*. Only a few ovule-related genes have been cloned from *Cymbidium*, especially *C. sinense*, with some reports focusing on the morphology and development of the *C. sinense* ovule (Swamy, 1942; Yeung et al., 1994; Yeung and Law, 1997). The molecular mechanisms in *C. sinense* ovules is still poorly understood.

Transcriptomic data can be used to glean a global understanding of gene expression in individual organisms at a certain point in time or in a specific tissue (Denoeud et al., 2008; Alagna et al., 2009; Dassanayake et al., 2009). Some studies on orchids have reported the key role of cDNA sequencing in discovering functional candidate genes in floral development, flowering time, and other developmental events, for example in *Phalaenopsis* and *Dendrobium* (Hsiao et al., 2006; Xu et al., 2006), as well as *Oncidium* and *Vanda* (Tan et al., 2005; Teh et al., 2011). Despite these studies, the molecular mechanism underlying ovule development in the Orchidaceae has received limited attention. In particular, for *C. sinense*, a comprehensive description of the complete complement of expressed genes is still unavailable, and candidate genes related to ovule development are poorly understood (Zhang et al., 2013). A critical in-depth study of the *C. sinensis* ovule is required to unravel its structural and developmental aspects as well as the underlying mechanism as a way to further *Cymbidium* breeding programs and develop a cultivation mechanism with high-yielding cultivars.

In this study, we performed an integrated transcriptomic and metabolomic analysis to investigate global changes in gene expression and metabolites during the development of *C. sinense* ovules. Our results shed light on the molecular mechanism of ovule development in this orchid.

MATERIALS AND METHODS

Plant Materials and Experimental Design

The flowers of *C. sinense* 'Qi Jian Hei Mo,' which was used in this study, were manually pollinated. Plants were grown and maintained in pots in a greenhouse at the Jiu Wan Orchid Field, in Foshan, China. The greenhouse was well ventilated, relative air humidity was 75–80%, day/night temperatures were 28/25°C with a 12-h photoperiod. Six developmental stages (S1–S6) of *C. sinense* ovules were collected from 12 fruits, frozen rapidly in liquid nitrogen and kept at −80°C until RNA extraction. Biological triplicates for each stage were used for RNA sequence (RNA-seq) analysis. Ovules at S2, S4, and S5 were collected to analyze metabolites, with six biological replicates for each stage.

RNA Extraction

Total RNA of each sample was extracted using the Quick RNA extraction kit (Huayueyang Biotechnology Co. Ltd., Beijing, China). Tissues were ground to a fine powder using a mortar and pestle after freezing them in liquid nitrogen. Powder was transferred into an RNase-free centrifuge tube, which

contained RNA extraction buffer, and vortexed rapidly for 30 s, performed according to the manufacturer's protocol. The integrity and content of total RNA was determined by a NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific, Wilmington, NC, United States) and an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA, United States). Three microgram of total RNA of each sample with an RNA integrity number (RIN) value ≥7.5 and concentration ≥ 100 ng/μL were used for cDNA library preparation and sequencing by Biomarker Technologies Inc. (Beijing, China).

cDNA Library Preparation and Sequencing

mRNA enrichment from total RNA was performer using oligo $d(T)_{25}$ magnetic beads (New England Biolabs Inc., Ipswich, MA, United States). mRNA was used for library preparation with the NEBNext® Ultra™ RNA Library Prep Kit (New England Biolabs Inc.). The insert size of libraries was estimated by the Agilent 2100 bioanalyzer, and the effective concentration was assayed by quantitative real-time (RT-qPCR). The libraries were sequencing using an Illumina Novaseq 6000 Sequencing System to produce raw reads. To obtain clean reads from raw reads, reads containing an adapter or ploy-N, as well as the poor-quality reads, were filtered out. These analyses were completed by Biomarker Technologies Inc.

De novo Assembly and Functional Annotation

The total clean reads generated by sequencing the 18 libraries were used for *de novo* full-length transcriptome reconstruction using Trinity (Grabherr et al., 2013) to generate transcripts. The assembled transcripts were clustered by Corset (Davidson and Oshlack, 2014) to remove redundant transcripts. Remaining transcripts were defined as unigenes. The integrity of assembled transcripts was evaluated with Benchmarking Universal Single-Copy Orthologs (BUSCO[1]) (Waterhouse et al., 2018).

Using BLASTx alignment (*E*-value ≤ 10^{-5}), the unigenes previously obtained were annotated with seven functional annotation databases, including the National Center for Biotechnology non-redundant protein sequences database (nr[2]), Clusters of Orthologous Groups of proteins (COG[3]), a manually annotated and reviewed protein sequence database (Swissprot[4]), Protein family (Pfam[5]), Kyoto Encyclopedia of Genes and Genomes (KEGG[6]), evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG[7]) and Gene Ontology (GO[8]). GetOrf (EMBOSS v. 6.3.1) used to predict the amino acid sequences of unigenes. Sequences were aligned in Pfam by HMMER software suite (v. 3.0) (E-value ≤ 10^{-10}). These analyses were completed by the authors in collaboration with Biomarker Technologies Inc.

Identification of Transcription Factor

To identify the TFs in *C. sinense* ovules, a program to identify and classify plant TFs, iTAK (Zheng et al., 2016),[9] and a prediction tool provided by a plant TF database (Tian et al., 2020),[10] were used to identify and classify the TFs. The predicted TFs from the two methods were combined as the total TFs in *C. sinense*.

Identification of Differentially Expressed Unigenes

Based on the *de novo* assembled data, the clean reads produced from each sample were mapped to the reference unigenes with RSEM (Li and Dewey, 2011) to obtain read counts of each sample. Read counts were then converted to fragments per kilobase of transcript sequence per millions (FPKM) to quantify transcript abundance. For differential expression analysis, read counts were used for normalization, estimate fold-change and calculate false discovery rate (FDR) using the DESeq method in the DESeq2 package (Love et al., 2014). A unigene with an absolute value of $\log_2$ (fold change) > 1 and an FDR < 0.05 between the six stages was defined as a differentially expressed unigene (DEU). These analyses were completed by the authors in collaboration with Biomarker Technologies Inc.

Metabolic Profiling

To screen out metabolites with important biological significance and with significant changes during the development of *C. sinense* ovules, three biological replicates of S1–S6 were analyzed, giving a total of 18 samples. We used a mixer mill (MM 400, Retsch, Haan, Germany) to homogenize the three freeze-dried biological samples of each stage, by adding zirconia beads (Biospec, BioSpec Products Inc., Bartlesville, OK, United States), and applying 30 Hz for 90 s. Samples of 20 ± 1 mg were transferred into 2 mL EP tubes (Axygen Scientific Inc., Silicon Valley, CA, United States), 500 μL of precooled extract (methanol (CNW Technologies, Duesseldorf, Germany; 67-56-1, HPLC): chloroform (Adamas, Basel, Switzerland; 67-66-3, HPLC = 3:1, v/v) was added, then 10 μL of adonitol (Sigma-Aldrich, Burlingtonm MA, United States; 488-81-3, ≥99%) was added. Samples were homogenized for 30 s using a vortex oscillator (Vortex-Genie 2, Scientific Industries, New York, NY, United States). Samples were immediately ground by a grinding mill (JXFSTPRP-24, Jingxin Industrial Development Co. Ltd., Shanghai, China) for 4 min at 35 Hz. The extraction mixture was treated in an ultrasonic apparatus (YM-080S, Fangao Microelectronics Co. Ltd, Shenzhen, China) for 5 min at 40°C while incubated in ice water, then centrifuged at 4°C for 15 min, at 10,000 rpm.

[1] http://busco.ezlab.org/
[2] http://www.ncbi.nlm.nih.gov
[3] http://www.ncbi.nlm.nih.gov/COG
[4] https://www.expasy.org/
[5] http://pfam.xfam.org/
[6] https://www.kegg.jp/
[7] http://eggnog.embl.de/
[8] http://geneontology.org/
[9] http://bioinfo.bti.cornell.edu/tool/itak
[10] http://planttfdb.gao-lab.org/

The supernatant (100 μL) was carefully collected and transferred to a fresh 1.5 mL EP tube. A 30 μL aliquot of supernatant from each sample was mixed as the quality control (QC) sample. The supernatant was collected and completely dried in a vacuum concentrator (Huamei Biochemical Instrument Factory, Taicang, China), to which 60 μL of methoxyamination hydrochloride (20 mg/mL in pyridine) was added and incubated for 30 min at 80°C. Samples were immediately derivatized with 60 μL of *bis*-(trimethylsilyl)-trifluoroacetamide (including 1% trimethylchlorosilane, v/v) reagent (REGIS Technologies, Chicago, IL, United States) and incubated for 90 min at 70°C. Lastly, 5 μL of fatty acid methyl esters (dissolved in chloroform) reagent (Dr. Ehrenstorfer GmbH, Augsburg, Germany) was added to the QC sample, and progressively cooled in the laboratory to 37°C. Samples were further analyzed by GC-TOF-MS, which was conducted using an Agilent 7890 gas chromatography system (Agilent, Atlanta, GA, United States) coupled with a Pegasus HT time-of-flight mass spectrometer (LECO Corp., St. Joseph, MI, United States). The system utilized a DB-5MS capillary column (30 m × 250 μm inner diameter, 0.25 μm film thickness; J&W Scientific, Folsom, CA, United States). The settings are described next. A 1 μL aliquot of sample was used to run the analysis in splitless mode. Helium was used as the carrier gas, the purge flow of the front inlet was 3 mL min^{-1}, and the gas flow rate across the column was 1 mL min^{-1}. The initial temperature was kept at 50°C for 60 s, before increasing to 310°C at a rate of 10°C min^{-1}, then kept for 8 min at 310°C. The ion source temperatures, transfer line, and injection were 250, 280, and 280°C, respectively. The energy was –70 eV in the electron impact mode. After a solvent delay of 6.35 min, the mass spectrometry data were acquired in full-scan mode with the *m/z* range of 50–500 at a rate of 12.5 spectra per second. The metabolites of 18 *C. sinense* ovule samples were delivered to, and quantitatively analyzed by, Biomarker Technologies Inc.

Identification of Differential Metabolites

The identification of differentially accumulated metabolites (DAMs) was based on an online METLIN batch metabolite search (Progenesis QI[11]) and Biomark's self-built library[12] for metabolite identification. Theoretical fragment identification was simultaneously carried out. The mass number deviations were all within 100 per mil (‰). Details such as data filtering, peak detection, comparison, calculation, and identification of DAMs are described in Tang et al. (2019). Among them, metabolites with FC > 2, *t*-test *P*-value < 0.05 and variable importance in projection (VIP) > 1 were considered as DAMs, which were mapped to the KEGG metabolic pathway for pathway enrichment analysis (FDR ≤ 0.05) (Kanehisa et al., 2008), including the use of the KEGG database,[13] the human metabolome database (HMDB[14]) and clusterProfiler R package[15] to annotate the identified metabolites. These analyses were completed by the authors in collaboration with Biomarker Technologies Inc.

[11]http://non-linear.com/

[12]http://www.biocloud.net/

[13]http://www.genome.jp/kegg/

RESULTS

Sequencing and *de novo* Assembly of *C. sinense* Transcriptome and Functional Annotation

To study the system-wide changes in the *C. sinense* ovule during differential developmental stages (S1–S6), we obtained the transcriptomic data of 18 *C. sinense* ovule samples. The statistical results of transcriptome sequencing are shown in **Table 1**. After removing low-quality reads and all possible contaminations, we obtained a total of 115.15 Gb of clean data; each sample has at least 6.14 Gb of clean data (**Table 1**). For each sample, more than 94.58% of bases had a score of ≥Q30 and the GC content of clean reads was similar (between 45.52 and 46.69%). This indicates that the obtained data are high quality data, and sequencing results could be used for subsequent analyses.

[14]http://www.hmdb.ca/

[15]http://www.bioconductor.org/

TABLE 1 | Sequencing statistics of *Cymbidium sinense* transcriptome during ovary development.

Sample	Number of clean reads	GC content (%)	% ≥ Q30 (%)	Mapped reads (%)	Mapped ratio
S1-1	21,187,529	45.81	95.73	18,189,961	85.85
S1-2	20,558,304	45.95	95.71	17,700,640	86.10
S1-3	20,467,384	46.01	95.65	17,629,023	86.13
S2-1	20,755,996	46.13	95.37	17,763,370	85.58
S2-2	20,235,253	46.12	95.33	17,379,672	85.89
S2-3	22,035,045	45.97	95.48	19,016,783	86.30
S3-1	22,962,814	46.69	95.09	19,984,499	87.03
S3-2	21,612,735	46.40	94.78	18,416,292	85.21
S3-3	21,190,300	46.53	94.94	18,062,959	85.24
S4-1	20,929,960	45.82	95.20	18,070,071	86.34
S4-2	20,900,762	46.14	95.45	18,154,223	86.86
S4-3	19,595,266	45.72	95.07	16,814,367	85.81
S5-1	22,375,253	45.68	95.23	19,202,416	85.82
S5-2	20,440,396	45.90	94.58	17,540,777	85.81
S5-3	23,577,643	45.89	95.16	20,235,517	85.83
S6-1	22,860,183	45.52	95.11	19,766,873	86.47
S6-2	20,797,301	45.97	95.40	17,849,805	85.83
S6-3	21,865,298	46.25	95.22	18,968,763	86.75

S1-1, S1-2, and S1-3 indicate the three biological replicates of stage 1 [30 days after pollination (DAP)]. S2-1, S2-2, and S2-3 indicate the three biological replicates of stage 2 (35 DAP). S3-1, S3-2, and S3-3 indicate the three biological replicates of stage 3 (42 DAP). S4-1, S4-2, and S4-3 indicate the three biological replicates of stage 4 (46 DAP). S5-1, S5-2, and S5-3 indicate the three biological replicates of stage 5 (53 DAP). S6-1, S6-2, and S6-3 indicate the three biological replicates of stage 6 (60 DAP).

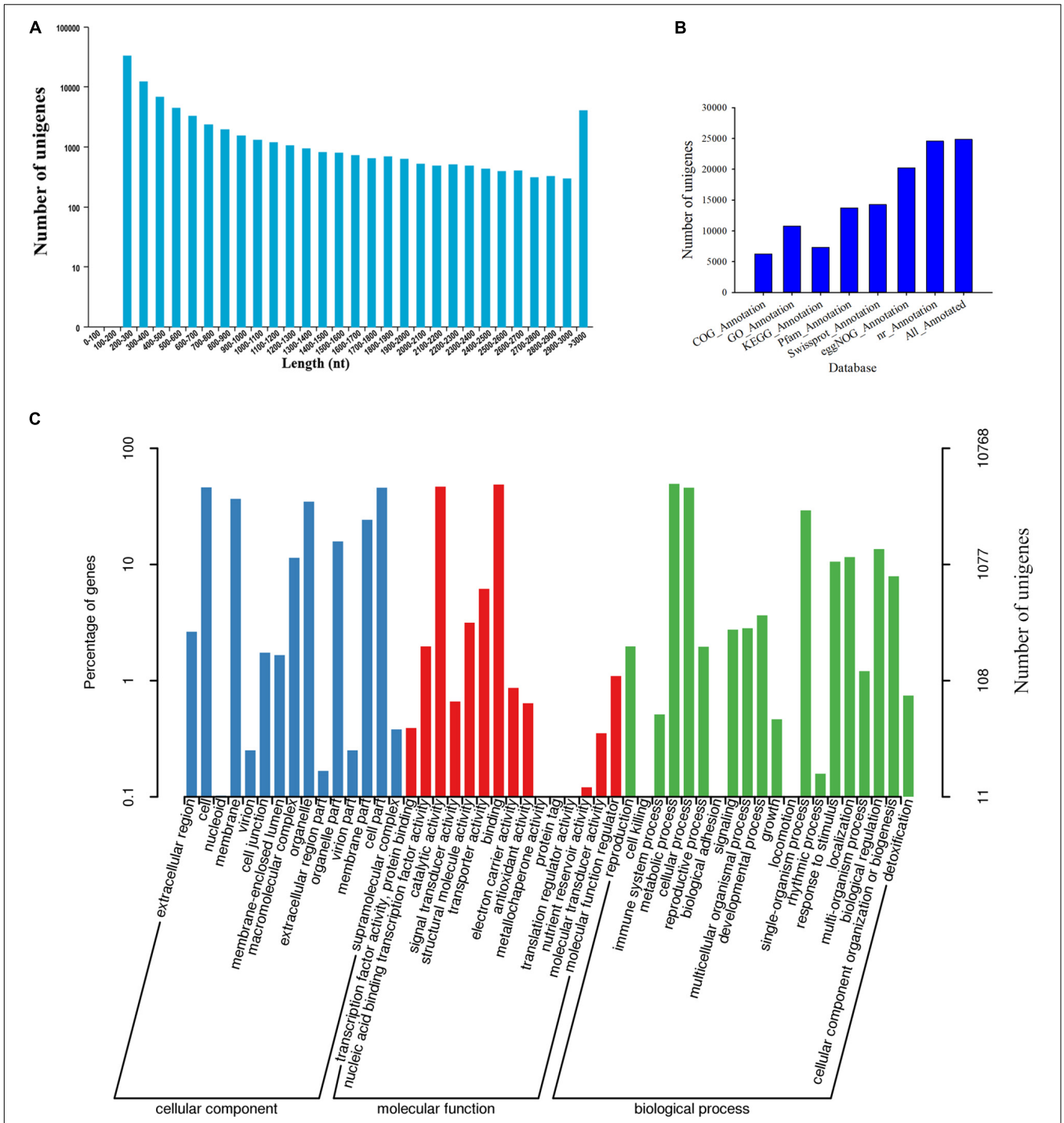


FIGURE 1 | Information of *de novo* assembled unigenes in *C. sinense* gleaned from transcriptomic analysis. **(A)** Length distribution of *de novo* assembled unigenes. **(B)** Annotation of unigenes against seven public databases (COG, GO, KEGG, Pfam, Swissprot, eggNOG, and nr). **(C)** GO functional classification (three subcategories) of the unigenes. The left Y-axis refers to the percentage of a specific category of target genes in that main category, while the right Y-axis refers to the number of target genes expressed in each subcategory.

First, the clean reads generated by sequencing of the transcriptome were aligned and assembled into contigs using Trinity software. Contigs were clustered and partially assembled to obtain transcripts, and then the most important transcript was selected as the unigene sequence. After further clustering and assembly, we obtained 81,585 unigenes in total (**Supplementary Table 1**). The N50 of the unigenes is 1,605, and 16,826 unigenes were longer than 1,000 bp

(**Figure 1A** and **Supplementary Table 1**). As shown in **Figure 1A**, the specific distribution of unigenes indicates that the sequence assembly of the *C. sinense* ovule is relatively complete.

Second, the 24,860 unigenes found in the *C. sinense* ovule were functionally annotated using seven public databases (COG, GO, KEGG, Pfam, SwissProt, eggnog, and nr) (**Figure 1B**). Of all unigenes, 30.47% were similar to known gene sequences in all databases (**Supplementary Table 2**).

To evaluate the functional role of target genes, their functional categories were classified according to their GO enrichment. A total of 10,768 genes (**Supplementary Table 2**) were assigned to at least one GO term, accounting for 43.31% of the functional categories of all 24,860 unigenes that were successfully annotated. These GO terms were categorized into 50 functional groups which were divided into three categories, including biological processes, cellular components and molecular functions (**Figure 1C**), which might play important roles during ovule development. In the cellular component category, the unigenes were chiefly related to "cell," "cell part," "membrane," and "organelle." The categories "binding" and "catalytic activity" were the predominant terms among the GO molecular function category. Within the biological processes category, "metabolic process," "cellular process," and "single-organism process" were the most abundant terms. In particular, the highest percentage of "metabolic process" among the biological processes suggests that a large number of new genes might be involved in secondary metabolite synthesis in *C. sinense* ovules.

Identification of Transcription Factors in *C. sinense*

TFs can recognize specific DNA sequences and regulate the expression of target genes, and are key regulators for regulating various biological processes (Mitsis et al., 2020). In order to promote functional gene research in *C. sinense*, a transcriptome-wide identification of TFs in *C. sinense* was performed. The MYB TF family accounted for the largest proportion (20.67%) of these genes, followed by 148 members, as well as zf-CCCH and bHLH with 92 and 69 members, respectively (**Figure 2** and **Supplementary Table 3**). Of particular note, two TF families, homeobox and MADS, which play various roles in plant growth and development, also accounted for a relatively large proportion, containing 52 and 48 members, respectively (**Supplementary Table 3**). These growth-related TFs might be involved in the development of *C. sinense* ovules.

Identification and Clustering Analysis of Differentially Expressed Unigenes During *C. sinense* Ovule Development

To clarify the key DEUs that regulate ovule development, a threshold with a $\log_2$ ratio >1 and $P < 0.05$ were used in pairwise comparisons to identify DEUs. Gene expression analysis showed that the expression of 9,845 unigenes was differentially altered in paired comparison of the six growth stages of *C. sinense* ovules (**Supplementary Table 4**). As shown in **Figure 3**, the pairwise

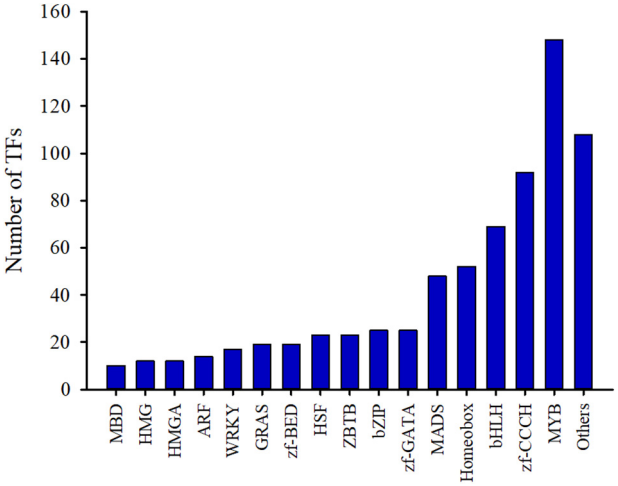


FIGURE 2 | Identification of transcription factors (TFs) from *C. sinense* ovules.

comparison of DEUs during ovule growth and development (except for S1_vs_S2 and S5_vs_S6 comparisons) indicated that a total of 5,065 unigenes were up-regulated while 4,780 unigenes were down-regulated. Among them, S3_vs_S6 had the largest number of DEUs, followed by S3_vs_S5, S2_vs_S6 and S1_vs_S6 comparisons, while no DEUs were detected in S1_vs_S2 and S5_vs_S6 comparisons. This implies that the S3 period of *C. sinense* ovule development might be a key node of drastic changes in gene expression, and that the periods when unigenes displayed obvious differential expression lay between S3 and S5 or S6.

To determine significant changes in KEGG pathways, we used DEU co-expression analysis. Furthermore, we performed *k*-means cluster analysis on 9,845 DEUs to decipher the overall trend of gene expression profiles. The clustering analysis indicated that key transcriptional activation or repression occurred during *C. sinense* ovule development. The 9,845 DEUs were grouped into four clusters during *C. sinense* ovule development (**Figure 4**). In cluster 1, S3 was the peak value (highest point), and the most enriched KEGG pathway is "starch and sucrose metabolism". This suggests that changes to genes in this pathway might be due to energy demands of the developing ovule. Clusters 2 and 3 displayed similar expression trends: S3 was the lowest point; overall, changes were relatively flat in S1–S3 and S4–S6. The most enriched KEGG pathway was "protein processing in endoplasmic reticulum". This suggests that protein processing was active during later stages of ovule development (S4–S6). In cluster 4, like cluster 1, the peak value was observed in S3 with the most enriched KEGG pathway being "photosynthesis". These results indicate that S3 is a key point of transcriptional changes, in which several key functional unigenes involved in "starch and sucrose metabolism," "protein processing in endoplasmic reticulum," "photosynthesis," and "phenylpropanoid biosynthesis" were highly enriched from among the six growth and development stages of the *C. sinense* ovule.

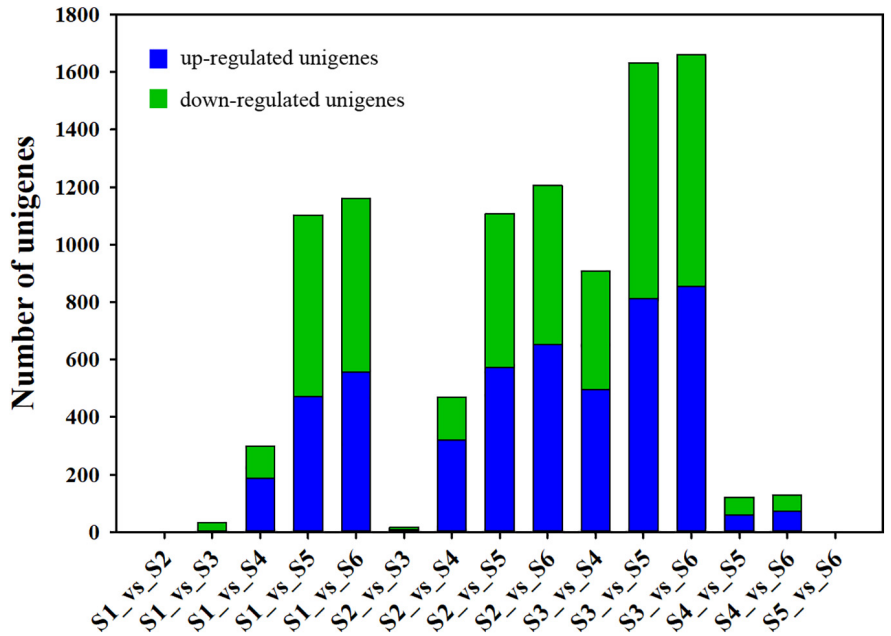


FIGURE 3 | Identification of differential expressed unigenes by pairwise comparisons. "S1_vs_S2" represents "the comparison between S1 and S2" or "S1 compared to S2". The same applies to all other comparisons between S1 and S6.

Characterization of Functional Genes Involved in Starch and Sucrose Metabolism, and Phenylpropanoid Biosynthesis

To obtain a more comprehensive understanding of transcript expression in *C. sinense* ovules at different developmental stages, we identified DEUs involved in starch and sucrose metabolism and phenylpropanoid biosynthesis during S1–S6 (**Tables 2**, **3**).

Starch and sucrose are important compounds in the growth and development of plants. Based on bioinformatics analysis of the *C. sinense* transcriptome, DEUs assigned to the "starch and sucrose metabolism" metabolic pathway were highly represented. Two unigenes (c59725.graph_c0 and c63927.graph_c0), encoding beta (β)-glucosidase, were up-regulated in S4–S6, with the highest expression in S4 (**Table 2**). The up-regulation of these two β-glucosidase unigenes at a late development stage suggests that related substances catalyzed by β-glucosidase increased in this stage of ovule development. The expression of an endoglucanase (c60343.graph_c0) unigene peaked in S3, but the overall expression of another endoglucanase unigene (c62140.graph_c0) was up-regulated during ovule development. Endoglucanase and β-glucosidase act synergistically to hydrolyze cellulose (Singhania et al., 2013). These results suggest that the hydrolysis of cellulose might occur during the development of the *C. sinense* ovule. Two trehalose 6-phosphate phosphatase unigenes (c59506.graph_c0 and c61840.graph_c0) were differentially expressed during ovule development. The expression of one trehalose 6-phosphate phosphatase unigene (c61840.graph_c0) unigene in S1, S2, and S3, peaking in S3, was significantly higher than in S4, S5, and S6, dipping in S6, while the expression of another (c59506.graph_c0) unigene peaked in S5, and its overall expression was higher in the later stages of growth and development of the *C. sinense* ovule (i.e., S4–S6) than in the early growth period (S1–S3). This indicates that these two trehalose 6-phosphate phosphatase unigenes may play different roles in *C. sinense* ovule development. Overall, the expression of two β-amylase (c48295.graph_c0 and c34186.graph_c0) unigenes peaked in S3. β-Amylase mediates maltose production from starch breakdown (Fulton et al., 2008), suggesting that stored starch was hydrolyzed and used as nutrients by the growing ovule. In addition, more than half of the expression of genes was highest during S3, indicating that this period might be critical for the "starch and sucrose metabolism" metabolic pathway.

The phenylpropanoid metabolic pathway is one of the main pathways for plants to synthesize secondary metabolic compounds. Further analysis revealed that 12 DEUs were identified as homologous sequences of the major genes related to the "phenylpropanoid biosynthesis" metabolic pathway, among which four types of DEUs were annotated (**Table 3**). Cinnamyl alcohol dehydrogenase (CAD) catalyzes the last step in the biosynthesis of monolignols, which are lignin precursors (Baucher et al., 1996). CAD expression level was high in S1–S2, but peaked in S3, suggesting that lignin might accumulate during ovule development. It is worth noting that eight peroxidase (POD) unigenes were discovered among the DEUs assigned to the "phenylpropanoid biosynthesis" pathway, and the expression levels of four POD unigenes (c60400.graph_c0, c56929.graph_c0, c36064.graph_c0, and c36619.graph_c0) were low while those of another two

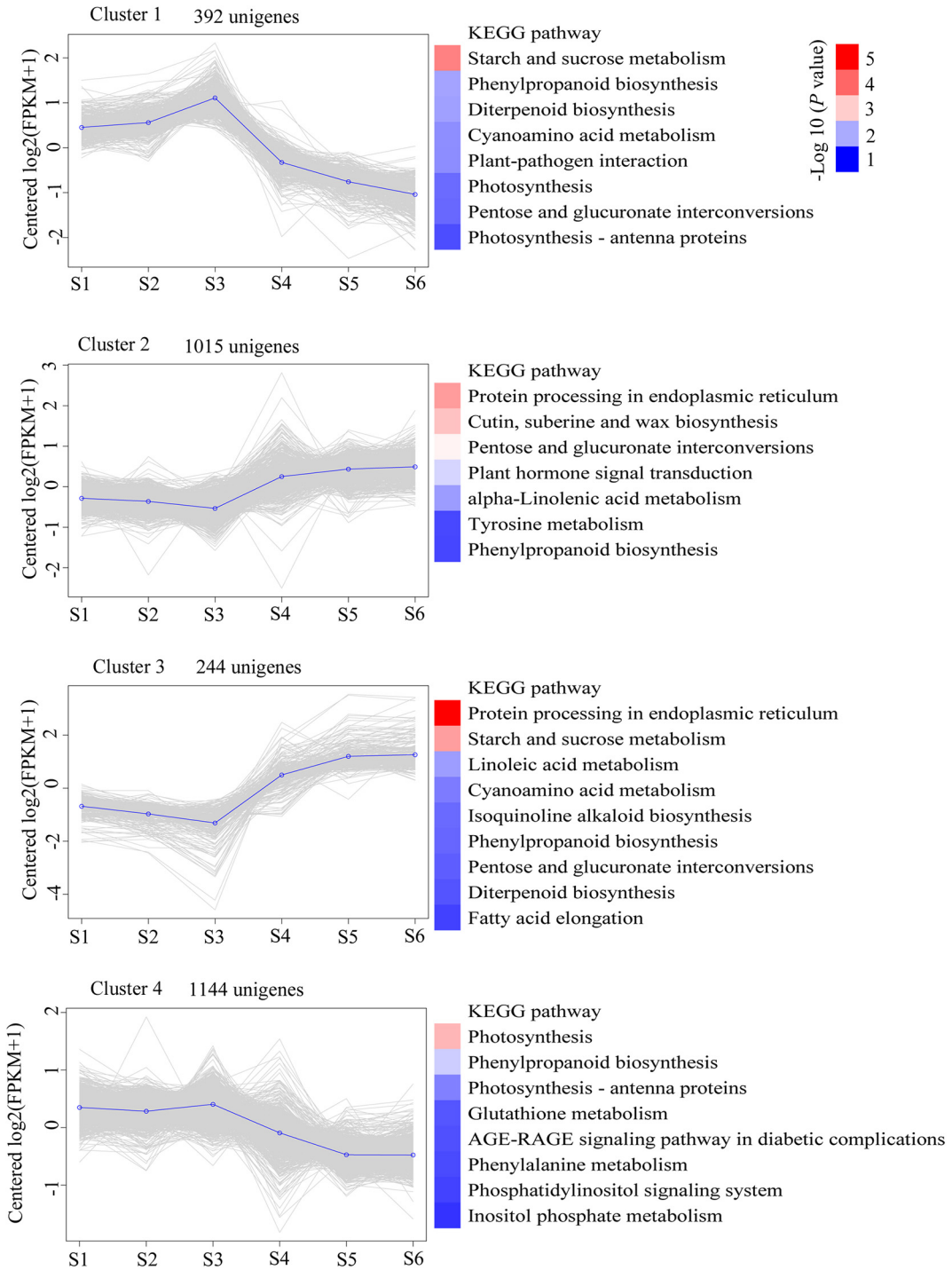


FIGURE 4 | Clustering analysis of time-course RNA-seq revealed transcriptional changes during development of the *C. sinense* ovule. Blue lines highlight the centroid of each cluster. The KEGG pathway categories with a *P*-value < 0.05 are shown. The color codes represent the -log10 (*P* value).

POD unigenes (c65092. graph_c0 and c56253.graph_c0) were very high, ranging from 84.05 to 224.89 FPKM. Of the remaining two PODs, one (c36841.graph_c0) had a higher expression level in the late period of germinal development i.e., S4–S6, while the other (c67154.graph_c0) had a higher expression level in the early period of growth and development i.e., S1–S3. The POD family of enzymes is involved in lignin biosynthesis (Yang et al., 2007; Herrero et al., 2013). Taken together, we speculate that lignin accumulated during ovule development.

TABLE 2 | Homologous sequences of DEUs involved in starch and sucrose metabolism during six growth and developmental stages of *C. sinense* ovules.

Gene ID	Description	S1	S2	S3	S4	S5	S6
c61950.graph_c0	Beta-fructofuranosidase	1.25	0.52	0.57	1.92	1.65	1.68
c63373.graph_c0	Fructokinase	4.06	3.04	7.34	2.22	1.21	1.03
c59725.graph_c0	Beta-glucosidase	1.82	1.09	1.08	6.81	11.85	6.36
c63927.graph_c0	Beta-glucosidase	1.05	0.89	0.13	3.87	12.19	8.44
c60343.graph_c0	Endoglucanase	31.98	17.7	44.66	13.96	14.05	12.79
c62140.graph_c0	Endoglucanase	0	0.03	0	0.07	2.01	3.17
c61840.graph_c0	Trehalase	24.92	28.28	26.87	17.94	12.99	11.02
c62609.graph_c1	Trehalose-phosphate phosphatase	7.17	6.85	9.27	5.86	3.19	2.68
c59506.graph_c0	Trehalose-phosphate phosphatase	7.58	6.38	4.35	15.82	34.67	31.37
c62634.graph_c0	Trehalose-phosphate phosphatase	106.15	133.1	120.87	75.28	35.64	34.49
c48295.graph_c0	Beta-amylase	4	5.11	5.6	2.3	1.49	1.43
c34186.graph_c0	Beta-amylase	8.08	9.95	13.56	3.55	5.17	6.69

The number in each stage indicates the average FPKM. More details about S1–S6 can be found in the materials and methods.

TABLE 3 | Homologous sequences of DEUs involved in phenylpropanoid biosynthesis during six growth and developmental stages of *C. sinense* ovules.

Gene ID	Description	S1	S2	S3	S4	S5	S6
c32911.graph_c0	Shikimate *O*-hydroxycinnamoyltransferase	0.15	0.17	0	0.66	1.06	2.66
c63253.graph_c0	Caffeic acid 3-*O*-methyltransferase	6.67	4.57	1.27	16.6	22.9	21.98
c66522.graph_c5	Caffeic acid 3-*O*-methyltransferase	21.96	27.36	29.73	17	17.43	14.13
c57919.graph_c0	Cinnamyl alcohol dehydrogenase	24.7	27.86	43.26	20.69	12.93	14.84
c60400.graph_c0	Peroxidase	1.62	1.71	2.37	0.8	1.16	0.53
c65092.graph_c0	Peroxidase	112.83	84.05	100.01	224.89	156.8	186.38
c56929.graph_c0	Peroxidase	3.53	3.27	5.94	3.88	1.38	1.56
c36841.graph_c0	Peroxidase	4.24	4.24	2.89	11.58	9.14	11.07
c36064.graph_c0	Peroxidase	2.19	1.32	1.97	0.06	0.18	0.08
c36619.graph_c0	Peroxidase	1.67	0.67	1.35	4.77	1.62	1.89
c56253.graph_c0	Peroxidase	123.79	124.29	111.99	117.35	196.35	208.12
c67154.graph_c0	Peroxidase	67.96	53.94	76.33	33.58	22.13	21

The number in each stage indicates average of FPKM. More details about S1–S6 can be found in the materials and methods.

Differential Metabolite Screening

To explore metabolic networks during *C. sinense* ovule development, an untargeted metabolomics method i.e., gas chromatography coupled with time-of-flight mass spectrometry (GC-TOF-MS), was used to detect the metabolites in S1–S6. In total, 18 samples were analyzed, revealing 556 metabolites. The identified metabolites were demonstrated using volcano plots and Venn diagrams. Generally speaking, gene expression is spatio-temporal (Larochelle et al., 1999; Kudapa et al., 2018), so changes in substances are expected to lag behind gene expression. Jointly observing **Figures 3**, **4**, it can be seen that S3 is a key node of changes in gene expression, so S4, the stage following S3, was selected for the study of metabolites. In addition, a comparison between S1 and S2 (S1_vs_S2) and S5_vs_S6 gene expression changed slowly and no DEUs were detected, so S2 and S5 were randomly selected for metabolomic analysis. Consequently, in order to gain insight into the changes to metabolites among developmental stages, DAMs between groups were identified when VIP was $\geq$1, fold-change was $\geq$2 or $\leq$0.5, and *P* value was $<$0.05. Here, three periods (S2, S4, and S5) were compared to obtained a volcano map for three pairwise comparisons (S2_vs_S4, S2_vs_S5, and S4_vs_S5), to visualize the identified metabolites. Moreover, we identified a total of 52 DAMs (23 in S2_vs_S4, 26 in S2_vs_S5, and three in S4_vs_S5 comparisons). Among the identified DAMs, the S2_vs_S4 comparison showed that four were up-regulated and 19 were down-regulated (**Figure 5A** and **Supplementary Table 5**), seven were up-regulated and 19 were down-regulated in the S2_vs_S5 comparison (**Figure 5B** and **Supplementary Table 5**), and very few metabolites showed significant differences, with only two being up-regulated and one down-regulated in the S4_vs_S5 comparison (**Figure 5C** and **Supplementary Table 5**). Furthermore, in order to compare and analyze the relationships between DAMs between groups, a Venn diagram was drawn based on the number of DAMs obtained from each of the three groups. The results showed that there were overlapping metabolites between groups (**Figure 5D** and **Supplementary Tables 6–8**), such as "analyte 80," "lactic acid," "linoleic acid methyl ester," "galactinol 2," etc. Among these DAMs, 14 were the same in the S2_vs_S4 and S2_vs_S5 comparison (including up- and down-regulation), one was the same in the S2_vs_S4 and S4_vs_S5 comparison, and two were the same in the S2_vs_S5 and S4_vs_S5 comparisons. Additionally, eight and 10 unique DAMs were found in S2_vs_S4 and

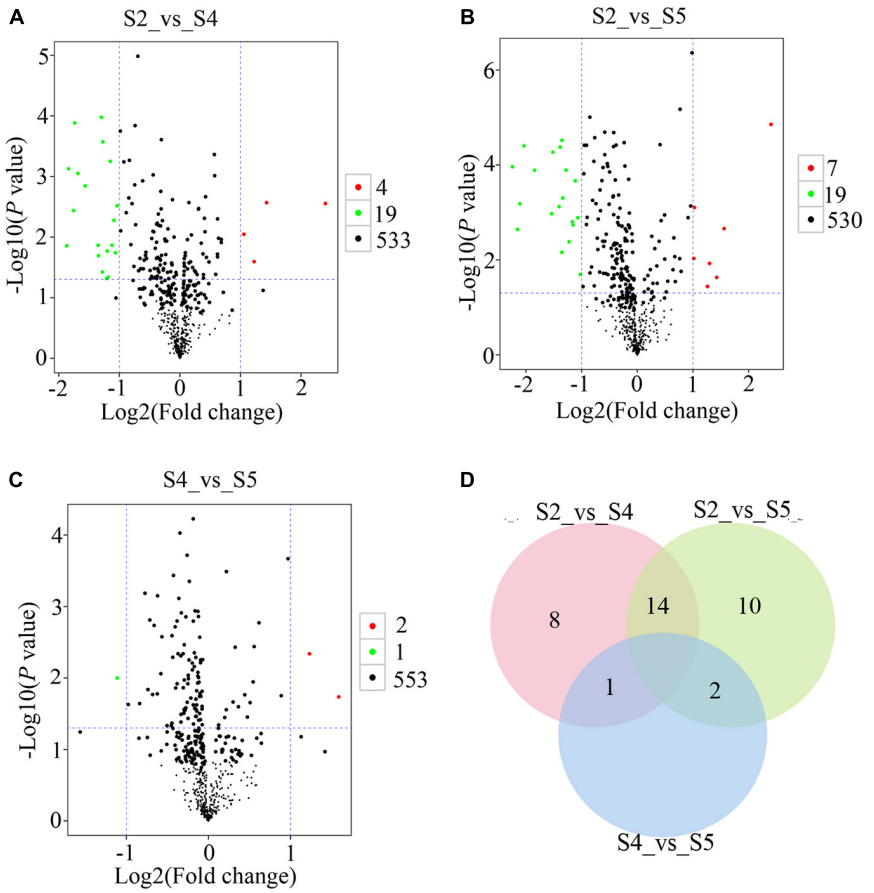


FIGURE 5 | Metabolomic GC-TOF-MS-based analyses of the *C. sinense* ovule at three developmental stages (S2, S4 and S5). Visualization, using volcano plots, of the identified metabolites from S2_vs_S4 **(A)**, S2_vs_S5 **(B)**, and S4_vs_S5 **(C)** comparisons. Green dots in the graph represent down-regulated DAMs, red dots represent up-regulated DAMs, and black represents metabolites that were detected but were not significantly different. **(D)** Venn diagram for the number of unique and overlapping DAMs from S2_vs_S4, S2_vs_S5 and S4_vs_S5 comparisons.

S2_vs_S5 comparisons, respectively. According to the KEGG annotation, five DAMs in the S2_vs_S4 comparison were assigned to 34 KEGG pathways, with four DAMs in "metabolic pathways" and three DAMs in "biosynthesis of secondary metabolites" (**Figures 6A,B**). In the S2_vs_S5 comparison, both "metabolic pathways" and "biosynthesis of secondary metabolites" contained three DAMs (**Figures 6A,C**). However, no DAMs assigned to any KEGG pathway were found in the S4_vs_S5 comparison (**Figure 6A**). These results suggested that the biosynthesis of metabolites was inactive during *C. sinense* ovule development.

Differential Expression Analysis of Homeobox and MADS-Box Unigenes at Different Developmental Stages

A large number of studies found that the TFs of homeobox and MADS-box genes are closely related to plant growth and development (Ramachandran et al., 1994; Jack, 2001; Holland, 2013). Based on the results of TFs identified in *C. sinense* (**Figure 2**), 52 (7.26%) homeobox and 48 (6.70%) MADS-box unigenes were identified (**Supplementary Table 3**). We selected homeobox and MADS-box unigenes for further expression pattern analysis by plotting the FPKM values of unigenes that were differentially expressed at S1-S6 in the *C. sinense* ovule to draw a heat map (**Figure 7**).

A total of 12 homeobox DEUs were identified, and their expression patterns were roughly divided into two clusters (**Figure 7A**). The expression level of seven DEUs (c62370.graph_c0, c57425.graph_c0, c60672.graph_c0, c37324.graph_c0, c57422.graph_c0, c49710.graph_c0, and c63993.graph_c1) in the upper part of the heat map was low in S1-S3, high in S4-S6, and peaked in S6 (five DEUs) or S5 (two DEUs). In contrast, the expression patterns of five DEUs (c51619.graph_c0, c61412.graph_c0, c57812.graph_c0, c33864.graph_c0, and c49736.graph_c0) in the lower part of the heat map were opposite to those of the aforementioned seven DEUs: their expression level in S1–S3 was higher than in S4–S6, peaking in S3 (3 DEUs) or S1 (2 DEUs). It should be emphasized that one DEU (c62370.graph_c0) had the highest overall expression level, the lowest expression at S3, and highest expression at S6.

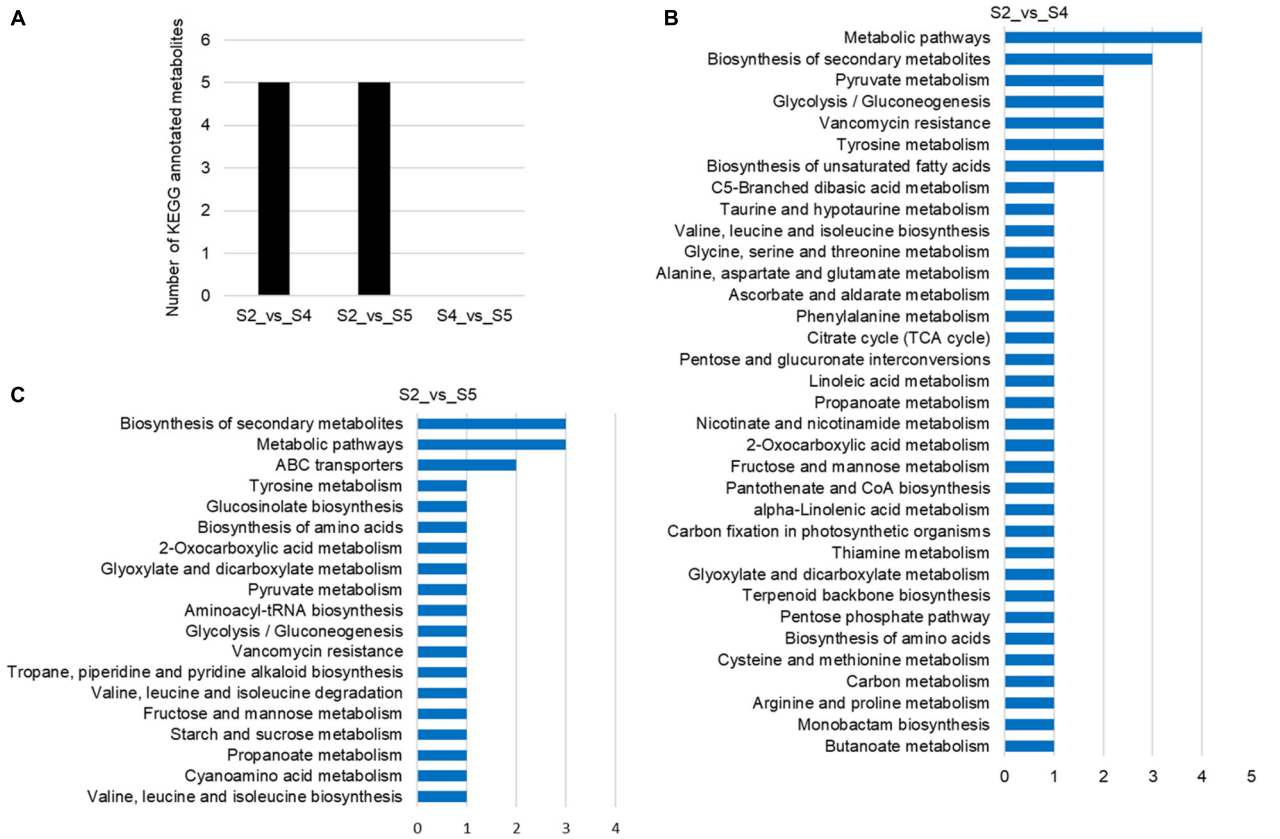


FIGURE 6 | KEGG annotations of differentially accumulated metabolites (DAMs) of the *C. sinense* ovule at three developmental stages (S2, S4, and S5). **(A)** Number of KEGG-annotated metabolites. Visualization, using a bar graph, of the identified metabolites from S2_vs_S4 **(B)**, and S2_vs_S5 **(C)**.

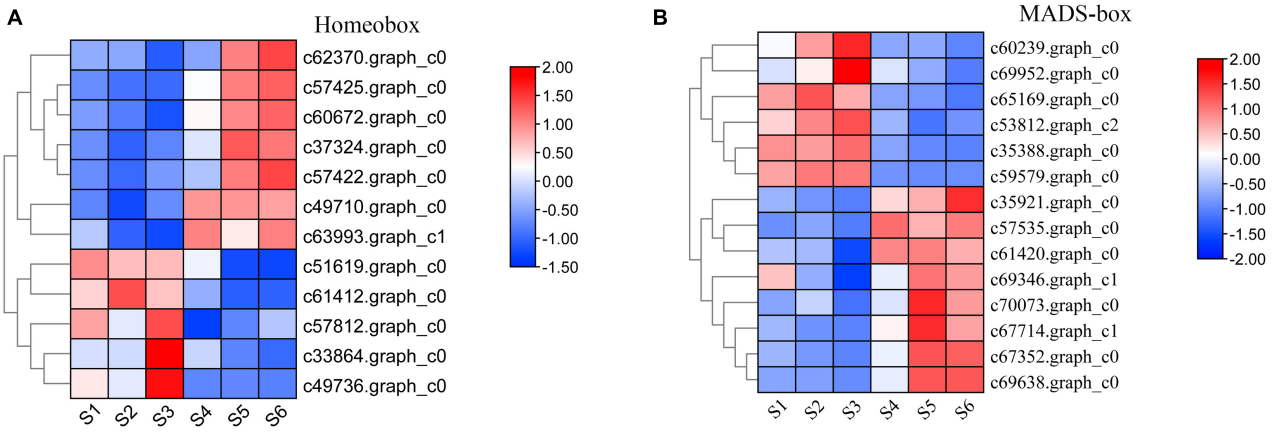


FIGURE 7 | Expression pattern of development-related genes during six growth stages of the *C. sinense* ovule. Expression patterns of DEUs annotated from the homeobox **(A)** and MADS-box **(B)** superfamilies. The color of each box represents log2(FPKM + 1) values, negative values (blue) represent down-regulation and positive values (red) represent up-regulation. Only DEUs from these two superfamilies are shown.

Homeobox genes thus appear to play a role in the growth and development of the *C. sinense* ovule.

Similarly, 14 MADS-box DEUs were identified and their expression patterns were also roughly divided into two clusters (**Figure 7B**). The expression levels of six DEUs (c60239.graph_c0, c69952.graph_c0, c65169.graph_c0, c53812.graph_c2, c35388.graph_c0, and c59579.graph_c0) in the upper part of the heat map, S1–S3 were higher than in S4–S6, except for one DEU (c65169.graph_c0) which showed highest expression in S2, while the remaining five DEUs showed highest

expression in S3. The remaining eight DEUs (C35921.graph_c0, c57535.graph_c0, c61420.graph_c0, c69346.graph_c1, c70073.graph_c0, c67714.graph_c1, c67352.graph_c0, and c69638.graph_c0) in the lower part of the heat map showed an opposite trend to the upper part, and their overall trends are essentially opposite to those of homeobox DEUs. Moreover, six DEUs displayed highest expression in S5. In particular, the overall expression level of one MADS-box DEU (c69638.graph_c0) was much higher than other DEUs, especially in S5 and S6, suggesting that this DEU might play an important role in the growth and development of the *C. sinense* ovule.

Functional Categorization of Deferentially Expressed Genes

To understand the changes in transcript pattern of unigenes involved in plant hormone synthesis and the signal transduction pathway during six stages of development of the *C. sinense* ovule, heat map analysis was used to analyze the relationships between them. Our results indicate that there were six major categories of plant hormones identified, namely IAA (**Figure 8A**), CK (**Figure 8B**), GA (**Figure 8C**), abscisic acid (ABA; **Figure 8D**), BR (**Figure 8E**), and JA (**Figure 8F**). Each category was divided into synthesis unigenes and responsive unigenes. These hormones are closely related to plant vegetative and reproductive growth, fruit ripening senescence, seed dormancy, and other functions.

The IAA synthesis-related unigenes included YUC8, YUC4, and YUC10 (**Figure 8A**). The overall expression level of YUC8 and YUC10 was low. The expression level of YUC4 was high during early stages of growth and development (S1–S3), but showed a decreasing trend in later stages (S4–S6), suggesting that this gene might play a role in immature tissues of *C. sinense* ovules. In addition, five types of unigenes responded to IAA: ARF, AUX/IAA (two homologous unigenes), AUX1 (two unigenes), CH3 (two unigenes), and SAUR (three unigenes). The expression of SAUR increased whereas that of ARF and AUX1 decreased as the ovule grew and developed. The expression patterns of the two homologous unigenes of AUX/IAA and GH3 were inverse during ovule growth and development. These differences in the expression of IAA-responsive unigenes suggest a complex IAA-related response during *C. sinense* ovule development.

The expression of three CK synthesis-related unigenes all peaked in S2 or S3, then decreased in later stages of development (S4–S6, **Figure 8B**). The expression of A-ARR, a CK-responsive unigene, showed a downward trend with ovule development while AHP (two unigenes) showed an upward trend (**Figure 8B**). In addition, there were eight GA synthetic unigenes, including of *GA3*, whose expression level decreased during ovule growth and development (**Figure 8C**), suggesting that it might play a similar or co-regulatory role with the IAA synthesis unigene YUC4 at an early stage of *C. sinense* ovule development. The only GA-responsive unigene, DELLA, had a high overall expression level and few differences in S1–S6 (**Figure 8C**). Moreover, CYP707A7, a ABA synthesis-related unigene, had a high expression level, peaking at S3, and showed a downward trend in S4–S6 (**Figure 8D**). Furthermore, the expression of BR synthesis unigenes CYP85A1 and CYP734A showed a decreasing trend in S4–S6, while CYP90B1 showed an opposite trend. The expression levels of BR-responsive unigenes were slightly different, while the expression trends of the two BKI1 were inconsistent (**Figure 8E**). Finally, from among the five JA synthesis-related unigenes, two (AOS and AOC) were highly expressed whereas the expression levels of the remaining three (OPR8, OPR4, and OPR11) were almost indistinguishable among S1–S6 (**Figure 8F**). In summary, their expression profiles were either up- or down-regulated at different developmental stages, indicating that these unigenes had an important role in the regulation of *C. sinense* ovule development. In particular, several key unigenes such as YUC4, GA3, and CYP707A7, played a special role in regulating the development of the *C. sinense* ovule. These results suggest that plant hormones played a role in *C. sinense* ovule development.

DISCUSSION

Transcriptional Sequencing and *de novo* Assembly of a Non-reference Genome Orchid, *C. sinense*

Ovules are the female sexual organ of higher plants that develop into seeds after fertilization (Shi and Yang, 2011). Ovule development, which is one of the most vital events in high plants because offspring are produced through sexual reproduction, plays a crucial role in plant evolution and the cultivation of new varieties (Yeung et al., 1994; Nadeau et al., 1996; Schneitz, 1999). *C. sinense* is a popular ornamental and economically important flower (potted and cut flowers) in the world and its seedlings are obtained by tissue culture (Xu et al., 2006; Aceto and Gaudio, 2011). However, new varieties of *C. sinense* with novel appearances and quality are annually in demand.

Transcriptional sequencing is an important method to obtain gene sequences and identify key genes involved in developmental events and metabolic pathways. To the best of our knowledge, there are no comprehensive studies on ovule development in *C. sinense*. Notably, *C. sinense* can only produce a massive number of ovules in each ovary after successful pollination (Yeung and Law, 1997), and the ovules of *C. sinense* develop and mature about 45 days after pollination, and can generally be divided into six stages. Consequently, to shed light on this process, transcriptional sequencing and *de novo* assembly of unigenes were performed, while changes in metabolites during ovule development were also estimated. In total, 81,585 unigenes were *de novo* assembled with an average length of 1857.33 bp (**Figure 1A** and **Supplementary Table 1**). In the orchid, *Dendrobium officinale*, 93,881 unigenes with a mean length of 790 bp were obtained using transcriptional sequencing (Zhang et al., 2016). In *Phalaenopsis*, 26,565 unigenes was annotated using transcriptional sequencing (Xu et al., 2015). KEGG pathway annotation showed that DEUs were mainly enriched in several pathways, such as "starch and sucrose metabolism," "phenylpropanoid biosynthesis," "protein processing in endoplasmic reticulum," "photosynthesis," and "plant hormone signal transduction" (**Figure 4**), suggesting that these pathways, particularly the first two, might be associated

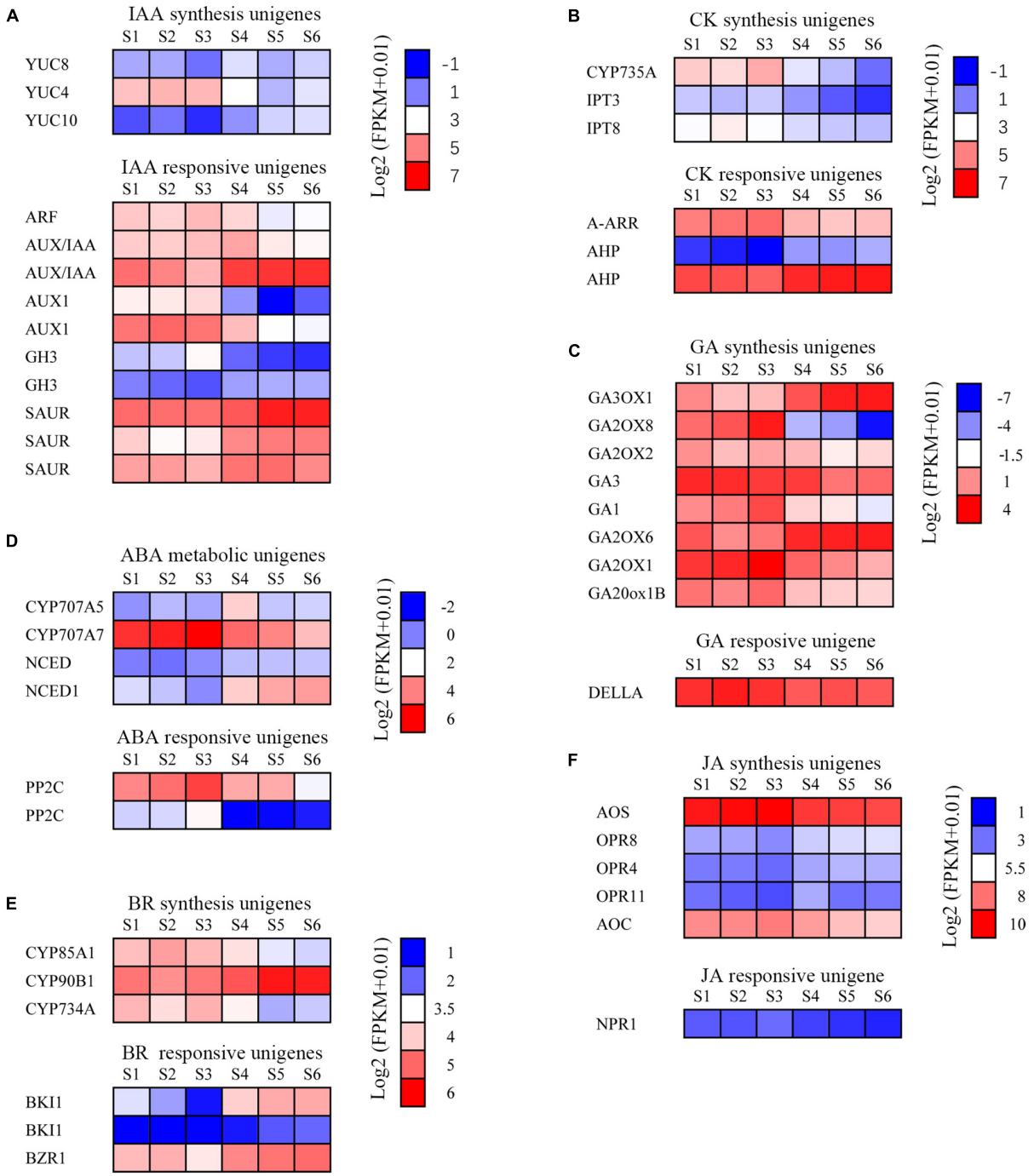


FIGURE 8 | Changes of plant hormone synthesis and signal transduction pathway unigenes during the development of the *C. sinense* ovule. **(A)** Expression analysis of auxin synthesis signal transduction pathway unigenes. ARF, auxin response factor; AUX/IAA, auxin/indole acetic acid protein; GH3, Gretchen Hagen; SAUR, small auxin-up RNAs; YUC, indole-3-pyruvate monooxygenase. **(B)** Expression analysis of cytokinin synthesis and signal transduction pathway unigenes. A-ARR, type-A Arabidopsis response regulator; AHP, histidine-containing phosphotransfer protein; CYP735A, cytokinin hydroxylase; IPT, adenylate isopentenyltransferase. **(C)** Expression analysis of GA synthesis and signal transduction pathway unigenes. GA, Ent-kaurene oxidase; GA2OX, Gibberellin 2-beta-dioxygenase; GA20ox1B, Gibberellin 20 oxidase 1-B; GA3OX, Gibberellin 3-beta-dioxygenase. **(D)** Expression analysis of ABA synthesis and signal transduction pathway unigenes. CYP707A5, CP450; abscisic acid 8'-hydroxylase; NCED, 9-*cis*-epoxycarotenoid dioxygenase; PP2C, protein phosphatase-2C. **(E)** Expression analysis of brassinosteroid synthesis and signal transduction pathway unigenes. BKI1, BRI1 kinase inhibitor 1-like; BZR1, protein BZR1 homolog 1-like; CYP734A, cytochrome P450 734A; CYP85A1, cytochrome P450 85A1; CYP90B1, cytochrome P450 90B1. **(F)** Expression analysis of JA synthesis and signal transduction pathway genes. AOC, Allene oxide cyclase; AOS, Allene oxide synthase; NPR1, BTB/POZ domain and ankyrin repeat-containing protein NPR1-like; OPR, 12-oxophytodienoate reductase. The color of each box represents the $\log_2$(FPKM + 1) value.

with differences among S1–S6 stages in ovule development. In the "phenylpropanoid biosynthesis" pathway, shikimate *O*-hydroxycinnamoyltransferase is involved in the biosynthesis of phenylpropanoid amides (PAs) (Kang et al., 2009). PAs are related to plant growth and development such as flower formation (Facchini et al., 2002). The caffeic acid 3-*O*-methyltransferase (OMT) family members *AtCCOMT2* and *AtCCOMT4-7* were expressed in the inflorescence stems of *Arabidopsis* (Raes et al., 2003). CAD is closely related to fruit development and floral organs (Kim et al., 2007; Singh et al., 2010). These results imply that PAs, OMT, and CAD may be indirectly involved in the regulation of ovule development. Unfortunately, these DEUs have not been reported in other plants, including orchids, in relation to ovule development. Therefore, this not only provides a good basis for discovering the expression of genes related to ovule development in *C. sinense*, but also indicates that transcriptional sequencing is a powerful tool for obtaining novel gene sequences.

Transcription Factors Are Involved in Plant Ovule Development

TFs are involved in developmental progression by regulating the expression of related target genes (Spitz and Furlong, 2012). Development of the ovule, a complex structure and precursor of a seed, requires the participation of a diversity of TFs. TFs related to ovule formation have been identified in *Jatropha curcas* L. (Xu et al., 2019) and *Arabidopsis* (Kelley and Gasser, 2009). Homeobox superfamily TFs are widely involved in plant development, including ovule development. A homeodomain TF, BELL1 (BEL1), is required for altering of both auxin distribution and patterning of the *Arabidopsis* ovule through exogenous cytokinin treatment (Bencivenga et al., 2012). The homeobox superfamily TF WUS gene, belonging to the WOX family, is specifically expressed in ovules and is responsible for ovule development in a range of plants (van der Graaff et al., 2009). Another *WOX* gene *PRETTY FEW SEEDS2* regulates ovule development in *Arabidopsis* (Park et al., 2005). In this study, 52 homeobox superfamily TFs were identified as DEUs (**Figure 2** and **Supplementary Table 3**). These might play a role in orchid ovule development. MADS-box family TFs are well known for their regulation of flower patterning, and increasing studies are revealing that MADS-box TFs are also involved in ovule development. For example, SEEDSTICK (STK; formerly known as AGL11), SHATTERPROOF1 (SHP1; formerly known as AGL1), and SHATTERPROOF 2 (SHP2; formerly known as AGL5) are responsible for ovule development, since *agl11 shp1 shp2* triple mutants displayed a phenotype in which ovules were converted into carpel-like or leaf-like structures in *Arabidopsis* (Pinyopich et al., 2003). The B_{sister} MADS-box gene was found during ovule development in petunia (*Petunia hybrida*) and *Arabidopsis*, and was critical for determining the identity of the endothelial layer within the ovule (de Folter et al., 2006). In our study, 48 DEUs belonging to the MADS-box family were identified (**Figure 2** and **Supplementary Table 3**), suggesting that MADS-box TFs might act as regulators in *C. sinense* ovule development.

Phytohormones May Play a Role in *C. sinense* Ovule Development

Phytohormones are fundamental players in the initiation of ovule formation. In cotton (*Gossypium hirsutum*), the content of endogenous phytohormones such as IAA, CK, ABA, and GA, as well as enzyme activities related with IAA and cytokinin metabolism, changed as the ovule developed (Zhang et al., 2009). In *Arabidopsis*, auxin (IAA) accumulated in ovules after fertilization and acts as an important role in early embryo patterning (Robert et al., 2018). In our study, three indole-3-pyruvate monooxygenase genes were differentially expressed during ovule development in *C. sinense* (**Figure 8**). The genes involved in CK, GA, and ABA biosynthesis and their signal transduction pathway also changed during six different ovule development stages (**Figure 8**). These results indicate that IAA, CK, ABA and GA may play a role in the development of the *C. sinense* ovule. Furthermore, BR can be perceived by membrane-localized kinase, namely BRASSINOSTEROID-INSENSITIVE 1 (Hothorn et al., 2011; She et al., 2011). BR is responsible for ovule integument growth in *A. thaliana* (Jia et al., 2020). Exogenous application of a BR, brassinolide, promoted fiber elongation, while a BR biosynthesis inhibitor repressed fiber development of *in vitro*-cultured cotton ovules (Sun et al., 2005). Our data showed that three genes were involved in BR-biosynthesis and three acted as BR-responsive genes whose expression was altered during ovule development in *C. sinense* (**Figure 8E**). This suggests that BR might play a role in the development of the ovule of this orchid.

CONCLUSION

In this study, we employed transcriptomic and metabolomic analyses to investigate the molecular and metabolic mechanisms controlling ovule development in *C. sinense*. We identified growth- and development-related TFs such as homeobox and MADS-box TFs, as well as plant hormones such as auxin, CK and BR, which may be feasibly involved in *C. sinense* ovule development. These TFs and phytohormones, as important regulators of ovule development, might play various and specific roles during *C. sinense* ovule development, providing comprehensive information to reveal the expression of relevant genes or various biological pathways of ovule development in the Orchidaceae. In addition, DAMs were identified in this paper by pairwise comparisons at six different developmental stages, although they are not directly reported to be associated with ovule development. Therefore, we speculate that the biosynthesis of metabolites was inactive during *C. sinense* ovule development. Overall, our results provide important clues that reveal the molecular mechanism underlying *C. sinense* ovule development, and aim to indicate some significant yet unexplored dynamics. In conclusion, the present results reveal that further research on the molecular mechanism underlying *C. sinense* ovule development can start from two pathways, the first being TFs related to ovule development identified in this study, and the other being hormone metabolism. Understanding the mechanisms of ovule development in orchids could lead toward orchid improvement,

and will ultimately diversify the orchid market. Further genetic and biochemical analyses would help understand the regulatory mechanism of orchid development.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: National Center for Biotechnology Information (NCBI) BioProject database under accession number PRJNA783745.

AUTHOR CONTRIBUTIONS

DL: conceptualization, resources, supervision, and project administration. DZ and DL: methodology. DZ and SX: software and validation. DZ and JT: formal analysis, writing—review and editing, and visualization. DZ: investigation. DZ and CQ: data curation and writing—original draft preparation. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was supported by the Foundation of the Open Project of Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, South China Botanical Garden, Chinese Academy of Sciences (grant No. KF202008).

ACKNOWLEDGMENTS

The authors are grateful to all the members of our laboratories for helpful discussion and assistance throughout this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.814275/full#supplementary-material

REFERENCES

Aceto, S., and Gaudio, L. (2011). The mads and the beauty: genes involved in the development of orchid flowers. *Curr. Genomics* 12, 342–356. doi: 10.2174/138920211796429754

Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., et al. (2009). Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* 10:399. doi: 10.1186/1471-2164-10-399

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

Bencivenga, S., Colombo, L., and Masiero, S. (2011). Cross talk between the sporophyte and the megagametophyte during ovule development. *Sex. Plant Reprod.* 24, 113–121. doi: 10.1007/s00497-011-0162-3

Bencivenga, S., Simonini, S., Benkova, E., and Colombo, L. (2012). The transcription factors BEL1 and SPL are required for cytokinin and auxin signaling during ovule development in *Arabidopsis*. *Plant Cell* 24, 2886–2897. doi: 10.1105/tpc.112.100164

Chen, Y.-Y., Lee, P.-F., Hsiao, Y.-Y., Wu, W.-L., Pan, Z.-J., Lee, Y.-I., et al. (2012). C- and D-class MADS-box genes from *Phalaenopsis equestris* (Orchidaceae) display functions in gynostemium and ovule development. *Plant Cell Physiol.* 53, 1053–1067. doi: 10.1093/pcp/pcs048

Colombo, L., Battaglia, R., and Kater, M. M. (2008). *Arabidopsis* ovule development and its evolutionary conservation. *Trends Plant Sci.* 13, 444–450. doi: 10.1016/j.tplants.2008.04.011

Dassanayake, M., Haas, J. S., Bohnert, H. J., and Cheeseman, J. M. (2009). Shedding light on an extremophile lifestyle through transcriptomics. *New Phytol.* 183, 764–775. doi: 10.1111/j.1469-8137.2009.02913.x

Davidson, N. M., and Oshlack, A. (2014). Corset: enabling differential gene expression analysis for *de novo* assembled transcriptomes. *Genome Biol.* 15:410. doi: 10.1186/s13059-014-0410-6

de Folter, S., Shchennikova, A. V., Franken, J., Busscher, M., Baskar, R., Grossniklaus, U., et al. (2006). A B_{sister} MADS-box gene involved in ovule and seed development in petunia and *Arabidopsis*. *Plant J.* 47, 934–946. doi: 10.1111/j.1365-313X.2006.02846.x

Denoeud, F., Aury, J.-M., Da Silva, C., Noel, B., Rogier, O., Delledonne, M., et al. (2008). Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* 9:R175. doi: 10.1186/gb-2008-9-12-r175

Facchini, P. J., Hagel, J., and Zulak, K. G. (2002). Hydroxycinnamic acid amide metabolism: physiology and biochemistry. *Can. J. Bot.* 80, 577–589. doi: 10.1139/b02-065

Fagundes, N. F., and de Araujo Mariath, J. E. (2014). Ovule ontogeny in *Billbergia nutans* in the evolutionary context of Bromeliaceae (Poales). *Plant Syst. Evol.* 300, 1323–1336. doi: 10.1007/s00606-013-0964-x

Fulton, D. C., Stettler, M., Mettler, T., Vaughan, C. K., Li, J., Francisco, P., et al. (2008). β-AMYLASE4, a noncatalytic protein required for starch breakdown, acts upstream of three active β-amylases in *Arabidopsis* chloroplasts. *Plant Cell* 20, 1040–1058. doi: 10.1105/tpc.107.056507

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., and Amit, I. (2013). Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Gross-Hardt, R., Lenhard, M., and Laux, T. (2002). Wuschel signaling functions in interregional communication during *Arabidopsis* ovule development. *Genes Dev.* 16, 1129–1138. doi: 10.1101/gad.225202

Herrero, J., Esteban-Carrasco, A., and Zapata, J. M. (2013). Looking for *Arabidopsis thaliana* peroxidases involved in lignin biosynthesis. *Plant Physiol. Biochem.* 67, 77–86. doi: 10.1016/j.plaphy.2013.02.019

Holland, P. W. H. (2013). Evolution of homeobox genes. *Wiley Interdiscip. Rev. Dev. Biol.* 2, 31–45. doi: 10.1002/wdev.78

Hothorn, M., Belkhadir, Y., Dreux, M., Dabi, T., Noel, J. P., Wilson, I. A., et al. (2011). Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* 474, 467–490. doi: 10.1038/nature10153

Hsiao, Y.-Y., Tsai, W.-C., Kuoh, C.-S., Huang, T.-H., Wang, H.-C., Wu, T.-S., et al. (2006). Comparison of transcripts in *Phalaenopsis bellina* and *Phalaenopsis equestris* (Orchidaceae) flowers to deduce monoterpene biosynthesis pathway. *BMC Plant Biol.* 6:14. doi: 10.1186/1471-2229-6-14

Huang, J., and Dai, S. (1998). The numerical taxonomy of Chinese *Cymbidium* (in Chinese). *J. Beijing For. Univ.* 20, 38–43.

Jack, T. (2001). Plant development going MADS. *Plant Mol. Biol.* 46, 515–520. doi: 10.1023/a:1010689126632

Jia, D., Chen, L.-G., Yin, G., Yang, X., Gao, Z., Guo, Y., et al. (2020). Brassinosteroids regulate outer ovule integument growth in part *via* the control of *INNER NO OUTER* by BRASSINOZOLE-RESISTANT family transcription factors. *J. Integr. Plant Biol.* 62, 1093–1111. doi: 10.1111/jipb.12915

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

Kang, K., Park, S., Kim, Y. S., Lee, S., and Back, K. (2009). Biosynthesis and biotechnological production of serotonin derivatives. *Appl. Microbiol. Biotechnol.* 83, 27–34. doi: 10.1007/s00253-009-1956-1

Kelley, D. R., and Gasser, C. S. (2009). Ovule development: genetic trends and evolutionary considerations. *Sex. Plant Reprod.* 22, 229–234. doi: 10.1007/s00497-009-0107-2

Ketsa, S., and Rugkong, A. (2000). Ethylene production, senescence and ethylene sensitivity of *Dendrobium* 'Pompadour' flowers following pollination. *J. Hortic. Sci. Biotechnol.* 75, 149–153. doi: 10.1080/14620316.2000.11511214

Kim, S. J., Kim, K. W., Cho, M. H., Franceschi, V. R., Davin, L. B., and Lewis, N. G. (2007). Expression of cinnamyl alcohol dehydrogenases and their putative homologues during *Arabidopsis thaliana* growth and development: lessons for database annotations? *Phytochemistry* 68, 1957–1974. doi: 10.1016/j.phytochem.2007.02.032

Kim, S. M., Jang, E. J., Hong, J. W., Song, S. H., and Pak, C. H. (2016). A comparison of functional fragrant components of *Cymbidium* (*Oriental orchid*) species. *Korean J. Hortic. Sci.* 34, 331–341. doi: 10.12972/kjhst.20160034

Kubo, T., Fujita, M., Takahashi, H., Nakazono, M., Tsutsumi, N., and Kurata, N. (2013). Transcriptome analysis of developing ovules in rice isolated by laser microdissection. *Plant Cell Physiol.* 54, 750–765. doi: 10.1093/pcp/pct029

Kudapa, H., Garg, V., Chitikineni, A., and Varshney, R. K. (2018). The RNA-seq-based high resolution gene expression atlas of chickpea (*Cicer arietinum* L.) reveals dynamic spatio-temporal changes associated with growth and development. *Plant Cell Environ.* 41, 2209–2225. doi: 10.1111/pce.13210

Kuhn, S. A., Nogueira, F. M., Fagundes, N. F., and Mariath, J. E. A. (2016). Morphoanatomy of the ovary and ovule in Bromeliaceae subfamily Tillandsioideae and its systematic relevance. *Bot. J. Linn. Soc.* 181, 343–361. doi: 10.1111/boj.12426

Larochelle, C., Tremblay, M., Bernier, D., Aubin, J., and Jeannotte, L. (1999). Multiple *cis*-acting regulatory regions are required for restricted spatio-temporal *Hoxa5* gene expression. *Dev. Dyn.* 214, 127–140. doi: 10.1002/(SICI)1097-0177(199902)214:2<127::AID-AJA3<3.0.CO;2-F

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Liu, H. H., Xiong, F., Duan, C. Y., Wu, Y. N., Zhang, Y., and Li, S. (2019). Importin β4 mediates nuclear import of GRF-interacting factors to control ovule development in *Arabidopsis*. *Plant Physiol.* 179, 1080–1092. doi: 10.1104/pp.18.01135

Locato, V., and De Gara, L. (2018). Programmed cell death in plants: an overview. *Methods Mol. Biol.* 1743, 1–8. doi: 10.1007/978-1-4939-7668-3_1

Lopez-Dee, Z. P., Wittich, P., Pe, M. E., Rigola, D., Del Buono, I., Gorla, M. S., et al. (1999). *Osmads13*, a novel rice mads-box gene expressed during ovule development. *Dev. Genet.* 25, 237–244. doi: 10.1002/(SICI)1520-6408199925:3<237::AID-DVG6<3.0.CO;2-L

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

Mendes, S. P., Mastroberti, A. A., Mariath, J. E. A., Vieira, R. C., and De Toni, K. L. G. (2014). Ovule and female gametophyte development in the Bromeliaceae: an embryological study of *Pitcairnia encholirioides*. *Botany* 92, 883–894. doi: 10.1139/cjb-2014-0114

Mitsis, T., Efthimiadou, A., Bacopoulou, F., Dimitrios, V., Chrousos, G. P., and Eliopoulos, E. (2020). Transcription factors and evolution: an integral part of gene expression. *World Acad. Sci. J.* 2, 3–8. doi: 10.3892/wasj.2020.32

Nadeau, J. A., Xian, S. Z., Nair, H., and O'Neill, S. D. (1993). Temporal and spatial regulation of 1-aminocyclopropane-1-carboxylate oxidase in the pollination-induced senescence of orchid flowers. *Plant Physiol.* 103, 31–39. doi: 10.1104/pp.103.1.31

Nadeau, J. A., Zhang, X. S., Li, J., and O'Neill, S. D. (1996). Ovule development: identification of stage-specific and tissue-specific cDNAs. *Plant Cell* 8, 213–239. doi: 10.1105/tpc.8.2.213

Nogueira, F. M., Fagundes, N. F., Kuhn, S. A., Fregonezi, J. N., and Mariath, J. E. A. (2015). Ovary and ovule anatomy in the nidularioid complex and its taxonomic utility (Bromelioideae: Bromeliaceae). *Bot. J. Linn. Soc.* 177, 66–77. doi: 10.1111/boj.12227

Novikoff, A. V., and Odintsova, A. (2008). Some aspects of comparative gynoecium morphology in three bromeliad species. *Wulfenia* 15, 13–24.

Park, S. O., Zheng, Z. G., Oppenheimer, D. G., and Hauser, B. A. (2005). The PRETTY FEW SEEDS2 gene encodes an *Arabidopsis* homeodomain protein that regulates ovule development. *Development* 132, 841–849. doi: 10.1242/dev.01654

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E., et al. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424, 85–88. doi: 10.1038/nature01741

Raes, J., Rohde, A., Christensen, J. H., Van de Peer, Y., and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol.* 133, 1051–1071. doi: 10.1104/pp.103.026484

Ramachandran, S., Hiratsuka, K., and Chua, N. H. (1994). Transcription factors in plant growth and development. *Curr. Opin. Genet. Dev.* 4, 642–646. doi: 10.1016/0959-437x(94)90129-q

Robert, H. S., Park, C., Gutièrrez, C. L., Wójcikowska, B., Pěnčík, A., Novák, O., et al. (2018). Maternal auxin supply contributes to early embryo patterning in *Arabidopsis*. *Nat. Plants* 4, 548–553. doi: 10.1038/s41477-018-0204-z

Rudall, P. J. (2021). Evolution and patterning of the ovule in seed plants. *Biol. Rev. Camb. Philos. Soc.* 96, 943–960. doi: 10.1111/brv.12684

Rudall, P. J., and Bateman, R. M. (2002). Roles of synorganisation, zygomorphy and heterotopy in floral evolution: the gynostemium and labellum of orchids and other lilioid monocots. *Biol. Rev. Camb. Philos. Soc.* 77, 403–441. doi: 10.1017/s1464793102005936

Sajo, M. G., Prychid, C., and Rudall, P. (2004). Structure and development of the ovule in Bromeliaceae. *Kew Bull.* 59, 261–267.

Schneitz, K. (1999). The molecular and genetic control of ovule development. *Curr. Opin. Plant Biol.* 2, 13–17. doi: 10.2307/4115859

She, J., Han, Z., Kim, T.-W., Wang, J., Cheng, W., Chang, J., et al. (2011). Structural insight into brassinosteroid perception by BRI1. *Nature* 474, 472–496. doi: 10.1038/nature10178

Shi, D.-Q., and Yang, W.-C. (2011). Ovule development in *Arabidopsis*: progress and challenge. *Curr. Opin. Plant Biol.* 14, 74–80. doi: 10.1016/j.pbi.2010.09.001

Sieber, P., Gheyselinck, J., Gross-Hardt, R., Laux, T., Grossniklaus, U., and Schneitz, K. (2004). Pattern formation during early ovule development in *Arabidopsis thaliana*. *Dev. Biol.* 273, 321–334. doi: 10.1016/j.ydbio.2004.05.037

Singh, R., Rastogi, S., and Dwivedi, U. N. (2010). Phenylpropanoid metabolism in ripening fruits. *Compr. Rev. Food Sci. Food Saf.* 9, 398–416. doi: 10.1111/j.1541-4337.2010.00116.x

Singhania, R. R., Patel, A. K., Sukumaran, R. K., Larroche, C., and Pandey, A. (2013). Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. *Bioresour. Technol.* 127, 500–507. doi: 10.1016/j.biortech.2012.09.012

Smyth, D. R., Bowman, J. L., and Meyerowitz, E. M. (1990). Early flower development in *Arabidopsis*. *Plant Cell* 2, 755–767. doi: 10.1105/tpc.2.8.755

Spitz, F., and Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626. doi: 10.1038/nrg3207

Sun, Y., Veerabomma, S., Abdel-Mageed, H. A., Fokar, M., Asami, T., Yoshida, S., et al. (2005). Brassinosteroid regulates fiber development on cultured cotton ovules. *Plant Cell Physiol.* 46, 1384–1391. doi: 10.1093/pcp/pci150

Swamy, B. (1942). Female gametophyte and embryogeny in *Cymbidium bicolor* Lindl. *Proc. Natl. Acad. Sci. India Sect. B* 15, 194–201. doi: 10.1007/BF03048760

Tan, J., Wang, H. L., and Yeh, K. W. (2005). Analysis of organ-specific, expressed genes in *Oncidium* orchid by subtractive expressed sequence tags library. *Biotechnol. Lett.* 27, 1517–1528. doi: 10.1007/s10529-005-1468-8

Tang, X., You, L., Liu, D., Xia, M., He, L., and Liu, H. (2019). 5-Hydroxyhexanoic acid predicts early renal functional decline in type 2 diabetes patients with microalbuminuria. *Kidney Blood Press. Res.* 44, 245–263. doi: 10.1159/000498962

Teh, S.-L., Chan, W.-S., Abdullah, J. O., and Namasivayam, P. (2011). Development of expressed sequence tag resources for *Vanda* Mimi Palmer and data mining for EST-SSR. *Mol. Biol. Rep.* 38, 3903–3909. doi: 10.1007/s11033-010-0506-3

Teixeira da Silva, J. A., Aceto, S., Liu, W., Yu, H., and Kanno, A. (2014). Genetic control of flower development, color and senescence of *Dendrobium* orchids. *Sci. Hortic.* 175, 74–86. doi: 10.1016/j.scienta.2014.05.008

Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J., and Gao, G. (2020). Plantregmap: charting functional regulatory maps in plants. *Nucleic Acids Res.* 48, D1104–D1113. doi: 10.1093/nar/gkz1020

Tsai, W.-C., Hsiao, Y.-Y., Pan, Z.-J., Kuoh, C.-S., Chen, W.-H., and Chen, H.-H. (2008). The role of ethylene in orchid ovule development. *Plant Sci.* 175, 98–105. doi: 10.1016/j.plantsci.2008.02.011

van der Graaff, E., Laux, T., and Rensing, S. A. (2009). The WUS homeobox-containing (WOX) protein family. *Genome Biol.* 10:248. doi: 10.1186/gb-2009-10-12-248

Villanueva, J. M., Broadhvest, J., Hauser, B. A., Meister, R. J., Schneitz, K., and Gasser, C. S. (1999). *INNER NO OUTER* regulates abaxial-adaxial patterning in *Arabidopsis* ovules. *Genes Dev.* 13, 3160–3169. doi: 10.1101/gad.13.23.3160

Wang, Y., Li, Y., Yan, X., Ding, L., Shen, L., and Yu, H. (2020). Characterization of C- and D-class MADS-box genes in orchids. *Plant Physiol.* 184, 1469–1481. doi: 10.1104/pp.20.00487

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319

Wu, Y., Yang, L., Cao, A., and Wang, J. (2015). Gene expression profiles in rice developing ovules provided evidence for the role of sporophytic tissue in female gametophyte development. *PLoS One* 10:e0141613. doi: 10.1371/journal.pone.0141613

Wu, Y., Yang, L., Yu, M., and Wang, J. (2017). Identification and expression analysis of microRNAs during ovule development in rice (*Oryza sativa*) by deep sequencing. *Plant Cell Rep.* 36, 1815–1827. doi: 10.1007/s00299-017-2196-y

Xu, C., Zeng, B., Huang, J., Huang, W., and Liu, Y. (2015). Genome-wide transcriptome and expression profile analysis of *Phalaenopsis* during explant browning. *PLoS One* 10:e0123356. doi: 10.1371/journal.pone.0123356

Xu, G., Huang, J., Lei, S.-K., Sun, X.-G., and Li, X. (2019). Comparative gene expression profile analysis of ovules provides insights into *Jatropha curcas* L. ovule development. *Sci. Rep.* 9:15973. doi: 10.1038/s41598-019-52421-0

Xu, Y. F., Teo, L. L., Zhou, J., Kumar, P. P., and Yu, H. (2006). Floral organ identity genes in the orchid *Dendrobium crumenatum*. *Plant J.* 46, 54–68. doi: 10.1111/j.1365-313X.2006.02669.x

Yang, Y. J., Cheng, L. M., and Liu, Z. H. (2007). Rapid effect of cadmium on lignin biosynthesis in soybean roots. *Plant Sci.* 172, 632–639. doi: 10.1016/j.plantsci.2006.11.018

Yeung, E. C., and Law, S. K. (1997). "Ovule and megagametophyte development in orchids," in *Orchid Biology: Reviews and Perspectives*, Vol. vii, eds J. Arditti and A. M. Pridgeon (Dordrecht: Springer), 31–73. doi: 10.1007/978-94-017-2498-2_2

Yeung, E. C., Zee, S. Y., and Ye, X. L. (1994). Embryology of *Cymbidium sinense*: ovule development. *Phytomorphology* 44, 55–63.

Yu, H., and Goh, C. J. (2001). Molecular genetics of reproductive biology in orchids. *Plant Physiol.* 127, 1390–1393. doi: 10.1104/pp.010676

Zhang, H., Shao, M., Qiao, Z., Yuan, S., Wang, X., and Hua, S. (2009). Effect of phytohormones on fiber initiation of cotton ovule. *Acta Physiol. Plant* 31, 979–986. doi: 10.1007/s11738-009-0313-4

Zhang, J., He, C., Wu, K., Teixeira da Silva, J. A., Zeng, S., Zhang, X., et al. (2016). Transcriptome analysis of *Dendrobium officinale* and its application to the identification of genes associated with polysaccharide synthesis. *Front. Plant Sci.* 7:5. doi: 10.3389/fpls.2016.00005

Zhang, J., Wu, K., Zeng, S., Teixeira da Silva, J. A., Zhao, X., Tian, C.-E., et al. (2013). Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. *BMC Genomics* 14:279. doi: 10.1186/1471-2164-14-279

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

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 27 January 2022
doi: 10.3389/fpls.2021.799778

Organ-Specific Gene Expression Reveals the Role of the *Cymbidium ensifolium-miR396/Growth-Regulating Factors* Module in Flower Development of the Orchid Plant *Cymbidium ensifolium*

Fengxi Yang[1], Chuqiao Lu[1], Yonglu Wei[1], Jieqiu Wu[1], Rui Ren[1], Jie Gao[1], Sagheer Ahmad[1], Jianpeng Jin[1], Yechun Xv[1], Gang Liang[2] and Genfa Zhu[1]**

[1] Guangdong Key Laboratory of Ornamental Plant Germplasm Innovation and Utilization, Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China, [2] CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Kunming, China

OPEN ACCESS

Edited by:
Katharina Nargar, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

Reviewed by:
Chunmei He, South China Botanical Garden, Chinese Academy of Sciences (CAS), China
Seonghoe Jang, World Vegetable Center Korea Office, South Korea

***Correspondence:**
Genfa Zhu
genfazhu@163.com
Yechun Xv
xuyechun@gdaas.cn

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: *22 October 2021*
Accepted: *27 December 2021*
Published: *27 January 2022*

Citation:
Yang F, Lu C, Wei Y, Wu J, Ren R, Gao J, Ahmad S, Jin J, Xv Y, Liang G and Zhu G (2022) Organ-Specific Gene Expression Reveals the Role of the Cymbidium ensifolium-miR396/Growth-Regulating Factors Module in Flower Development of the Orchid Plant Cymbidium ensifolium. Front. Plant Sci. 12:799778. doi: 10.3389/fpls.2021.799778

Orchids are some of the most popular ornamental plants worldwide. Orchid floral morphology has increasingly attracted horticultural and commercial attention. Although multiple genes have been shown to be involved in the formation of the orchid flower, the underlying multi-level regulatory networks are largely unknown. In this study, we analyzed the ontogeny of flower development in *Cymbidium ensifolium*, a traditional orchid in the tropical and subtropical regions of Asia, by performing deep sequencing of the transcriptome of individual flower organs to discover organ-specific genes potentially involved in their growth. We identified 3,017 differentially-expressed genes (DEGs) during the development of various flower organs, and observed over-representation of *GROWTH-REGULATING FACTORS* (*GRFs*) specific to flower column (gynostemium). Eleven *C. ensifolium GRFs* (*CeGRFs*) from our transcriptome data clustered into five phylogenetic subgroups. Ten of these *GRFs* shared a region complementary to *C. ensifolium* microRNA396 (*Ce*-miR396), and degradome sequencing confirmed the cleavage of transcripts derived from seven *CeGRFs*. We cloned *Ce*-miR396 and used a protoplast-based transient expression system to overexpress it in *Cymbidium* protoplasts. We observed a significant decrease in the transcripts of several *CeGRFs* in flowers and leaves, indicating a potential role for miR396–GRF module in organ development through the cleavage of distinct *CeGRFs*. Temporal and spatial expression analysis indicated that most *CeGRF* transcripts accumulated in flower buds and column tissues, where *Ce*-miR396 expression was the lowest. Expression dynamics in wild type and floral-defective mutants further confirmed a strong correlation between *Ce*-miR396, *CeGRFs*, and flower organ development and column specification. Moreover,

overexpression of *Ce*-miR396 in *Nicotiana tabacum* resulted in curved pistils and reduced fertility, implying that the conserved role of *Ce*-miR396 in floral development. These results provide tools to better understand the biological roles of *GRFs* in orchid development, and open new avenues for the diversification of orchid floral patterns.

Keywords: orchid, *Cymbidium ensifolium*, floral organ, growth-regulating factor, MiR396

INTRODUCTION

Orchids are a highly valuable floricultural crop. Comprising more than 25,000 species distributed in ∼800 genera, orchids represent one of the largest families of flowering plants (Ramirez et al., 2007; Givnish et al., 2015). Over the past 1,500 years, more than 70,000 orchid hybrids and cultivars have been grown as ornamentals and medicinal plants, as well as food flavoring agents and essential oils (Wang et al., 2015; Li et al., 2017; Su et al., 2018). In China, Japan, South Korea, and Southeast Asia, orchids of genus *Cymbidium* are prized for their beautiful, fragrant flowers and elegant leaves. In particular, *Cymbidium ensifolium*, which belongs to subgenus Jensoa, blossoms many times a year and is thus highly valuable in flower markets of China (Zhang et al., 2015; Yang et al., 2017; Su et al., 2018).

Orchid flower is an important ornamental and industrial material. A standard orchid flower is composed of three petal-like sepals in the first whorl, two lateral petals, a specialized bottom petal (called lip or labellum) in the second whorl, and a column (also called gynostemium) formed by the fusion of pistil, stigma and pollinia in the middle of the flower. Homeotic mutations occur frequently in orchid family, that diversify their spectacular floral morphology (Aceto and Gaudio, 2011; Yang et al., 2017; Xiang et al., 2018). The molecular underpinnings of flower development in terms of floral organ number, arrangement, and initiation timing has been widely studied in the orchid family. A unique mechanism underlying perianth patterning of orchid plants is mediated by MADS-box type transcription factors (TFs) with some modifications of the arabidopsis ABCE model of flower development (Hsu et al., 2015). In addition, many floral meristem identity genes have been functionally characterized, including *FLOWERING LOCUS T* homologs (*Phalaenopsis aphrodite, Oncidium* Gower Ramsey, *Dendrobium nobile*), *CONSTANS-like genes*, *LEAFY* (*Phalaenopsis aphrodite)*, and coregulated transcription factors such as *CINCINNATA-like* (*TCP-like*) and *SQUAMOSA promoter-binding-like* (*SPL-like*) genes (Hou and Yang, 2009; Chou et al., 2013; Lin et al., 2013, 2016; Jang, 2015; Liu et al., 2016). In association with transcription factors, microRNAs play crucial roles in flower development by regulating TF transcript levels. The miRNAs have been detected by deep-sequencing in orchids, such as *Phalaenopsis aphrodite*, *Erycina pusilla,* and *Dendrobium officinale* (An and Chan, 2012; Wang et al., 2013; Yang et al., 2015). However, little is known about how miRNAs function in orchid flower formation.

MiR396 is an evolutionarily conserved miRNA that recognizes a complementary sequence in the mRNA of land plant *GRF*s, encoding a plant-specific family of transcription factors involved in the control of cell proliferation. MiR396/*GRF* regulatory network affects plant growth and responses to environmental changes (Omidbakhshfard et al., 2015). In the model plant arabidopsis, *GRF*s targeted by miR396 regulate leaf and cotyledon growth, embryogenesis, and development of stem, flowers, and roots (Liebsch and Palatnik, 2020). Diverse functions of this module have also been described in non-model plants. For example, the miR396–GRF/GRF-interacting factor (GIF) module influences tomato (*Solanum lycopersicum*) fruit size (Cao D. et al., 2016), *Brassica* sp. root and leaf growth (Hong et al., 2018), maize (*Zea mays*) leaf size and plant height (Wu et al., 2014; Zhang et al., 2018), soybean (*Glycine max*) nematode resistance (Liu et al., 2017), and rice (*Oryza sativa*) plant height, meristem function, flowering time, inflorescence architecture, heading date, seed size and pathogen resistance (Gao et al., 2015; Sun et al., 2016; Shimano et al., 2018; Tang et al., 2018). MiR396–GRF module makes a crucial hub coordinating various growth and physiological responses with endogenous and environmental signals. This makes it a highly promising target for crop breeding and biotechnology. However, considering the extensive functional redundancy or sub-functionalization inherent to GRF family members resulting from sequence divergence and their distinct temporal and spatial expression patterns, the specific nature of underlying mechanisms remain unclear for many species.

The GRF family of TFs remains to be revealed in orchids. We therefore carried out the first comprehensive analysis and molecular dissection of *GRF* gene family in *C. ensifolium*, integrating comparative transcriptome analysis, gene structures, phylogenetic relationships, conserved protein motifs and expression patterns associated with plant development and flower ontogeny. We confirmed the existence of a conserved miR396–GRF module in orchids. We also examined the relationship between *Ce*-miR396 and *Ce*GRFs among different flower patterning varieties and their role in orchid flower development, revealing a large contribution to column specification. Transgenic *Nicotiana tabacum* overexpressing *Ce*-miR396 displayed curved pistil and reduced fertility. This study therefore illustrates that *CeGRFs* are regulated by *Ce*-miR396 during orchid flower

Abbreviations: GRF, growth regulating factor; GIF, GRF-interacting factor; DEGs, differentially expressed genes; GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; TF, transcription factor; PTES, protoplast-based transient expression system; NF-Y, nuclear factor Y; TALE, three-amino-acid-loop-extension; Bzip, basic leucine zipper domain; bHLH, basic helix-loop-helix.

development and provides building blocks for molecular breeding of orchids in the enhancement of their remarkable floral patterns.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Wild-type plants and the natural mutants of *C. ensifolium* were artificially cultivated and collected from the cultivation base of Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, China. All plants were grown and maintained in pots in a greenhouse at day/night temperatures of 26/23°C under 16-h light/8-h dark photoperiod.

Library Construction and Illumina Sequencing

For transcriptome sequencing, we constructed independent cDNA libraries for sepal, petal, labellum, and gynostemium obtained from floral buds at developmental stage 3 of three individual plants, separately. The floral apex resembles an inverted triangle, and the ventral outer sepal and petals grow quickly in this stage. Two replications were included in each sample to create eight multiplexed cDNA libraries. The mRNAs were purified from total RNA using Oligotex mRNA Midi Kit (QIAGEN, Germany) and quantified using Nano-Drop 2000 spectrophotometer (Thermo Scientific, United States) to generate the cDNA library according to Illumina manufacturer's instructions (Yang and Zhu, 2015). The purified mRNAs were fragmented to approximately 200 bp and subjected to first strand and second strand cDNA synthesis, followed by adaptor ligation and low-cycle enrichment using TruSeq®RNA HT Sample Prep Kit (Illumina, United States). The purified library products were evaluated with Agilent 2200 TapeStation and Qubit®2.0 (Life Technologies, United States) and diluted to 10 pM for cluster generation *in situ* on the HiSeq3000 pair-end flow cell, followed by sequencing (2 × 100 bp). An average of more than 75 million reads were generated for each sample.

Degradome Library Construction and Sequencing

The total RNA quantity and purity were analyzed using Bioanalyzer 2100 and RNA 6000 Nano Lab Chip Kit (Agilent, CA, United States) with RIN number > 7.0. Approximately 20 μg of total RNA was used to prepare Degradome library. We followed the method of German et al. (2008) with some modification. Briefly, about 150 ng of poly(A) + RNA was isolated annealed with Biotinylated Random Primers. The biotinylated RNA fragments were captured by streptavidin. The annealed products containing 5'-monophosphates were ligated to 5' adaptor, followed by reverse transcription and PCR. Libraries were sequenced using the 5' adapter only. Then, the single-end sequencing (36 bp) was performed on Illumina Hiseq2500 at the LC-BIO (Hangzhou, China).

Analysis of Differentially Expressed Genes

Gene expression levels were calculated by FPKM values: FPKM = [total transcript fragments/mapped fragments (millions)] × transcript length (kb). Significant differences in gene expression between wild type and mutant were determined using edgeR. The false discovery rate (FDR) was applied to identify the threshold of *P*-value in multiple tests. An FDR < 0.05 and $|\log_2 \text{ratio}| > 1$ (two-fold change) were set as the threshold of significant difference in gene expression. Differentially expressed genes (DEGs) were annotated using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses (Yang and Zhu, 2015; Yang et al., 2017). All the DEGs were mapped to GO terms (or KEGG pathways) in the databases[1,2] and gene numbers were calculated for each term (or pathway) (Kanehisa and Goto, 2000; Conesa et al., 2005). Then a hypergeometric test was applied to find significantly enriched terms in DEGs compared to the genomic background. For this purpose, we used the corrected *P*-value ≤ 0.05 as a threshold for significantly enriched GO terms and *Q*-value ≤ 0.05 as a threshold for KEGG pathways.

Protoplast Isolation and Transfection

The plasmid pAN580-*GFP* containing dual cauliflower mosaic virus 35S promoter and *GFP* gene was used as the empty control. The 124-bp sequence of *Ce*-miR396 precursor was cloned into vector pAN580-*GFP*, and replaced with the *GFP* gene, resulting in the plasmid pAN580-pre-*Ce*-miR396. The young leaves and flower buds at developmental stage 3 of *C. ensifolium* were collected for protoplast isolation. The protoplast isolation was conducted following the past protocols (Negrutiu et al., 1987). Petals were cut into 0.5–1.0 mm strips and transferred into the freshly prepared enzyme-solution [1.0% (weight/volume, *w*/*v*) Cellulase R10, 0.5% (*w*/*v*), Macerozyme R10, 500-mM D-mannitol, 20-mM KCl, and 20-mM MES (pH = 5.7), 10-mM $CaCl_2$, 0.1% (*w*/*v*) BSA]. The released protoplasts were harvested after incubation at 28°C in the darkness with rotations of 30 rpm for 5–6 h. Protoplast transfection was carried out using PEG-mediated protocol with slight modifications. Briefly, an equal volume of freshly prepared PEG solution (40% (*w*/*v*) PEG 4000, 0.2 M mannitol and 0.1 M $CaCl_2$) was gently mixed with plasmid DNA in MMG solution [15 mM $MgCl_2$, 0.4 M mannitol and 4 mM MES (PH = 5.7)]. Transfected protoplasts were incubated at 23°C for 6–36 h in the darkness. Transfection efficiency was measured according to the expression of the GFP reporter of transient expression vector pAN580-GFP (Ren et al., 2020). The transfection efficiency was about 80% in both leaf- and flower-derived protoplasts. The GFP fluorescence was observed and 3–5 images were taken in random distribution under an LSM 710 confocal laser scanning microscope. Total RNA was extracted 0, 12, 24, and 36 h after transfection for reverse transcription followed by quantitative PCR (RT-qPCR), to analyze transcript abundance of *CeGRFs* and *Ce*-miR396 *in vivo*.

[1] http://www.geneontology.org/
[2] http://www.genome.ad.jp/

Stem-Loop Reverse Transcription-PCR of miR396 and Reverse Transcription-Quantitative PCR of Target Growth-Regulating Factor Genes

The cDNA of mature miRNA was prepared using miRNA reverse transcription kit M-MLV (Takara, China), and the reverse-transcribed products were used as template for RT-qPCR with gene-specific primers. The miRNA specific stem-loop primers and gene-specific RT-qPCR primers were designed according to the protocol described previously (Chen et al., 2005). For target gene validation, the cleavage site-spanning fragments of GRF genes were detected. Total RNA extracted from different tissue types was reverse-transcribed by oligo (dT) primed cDNA synthesis protocol (Fermentas). The resulting cDNA was subjected to quantitative PCR using Bio-Rad CFX-96 RealTime PCR System (Bio-Rad, United States) in a final volume of 20 μl containing 2 μl of cDNA and 10 μl of SYBR premix Ex-taq^{TM} (Takara, Japan). Ubiquitin was used as an internal control for normalization to compare gene expression level between the accessions. For each reported result, at least three independent biological samples were subjected to a minimum of three technical replicates. The primers designed with Primer 7.0 software are listed in **Supplementary Table 1**.

Multiple Sequence Alignment and Phylogenetic Analysis

Cymbidium ensifolium growth-regulating factor coding sequences were identified in our transcriptome dataset using a Basic Local Alignment Tool (BLAST) analysis for proteins. A total of 92 complete protein sequences were aligned using MUSCLE, including 11 CeGRFs from *Cymbidium ensifolium*, 9 AtGRFs from *Arabidopsis thaliana*, 12 OsGRFs from *Oryza sativa*, 17 BrGRFs from *Brassica rapa*, 14 ZmGRFs from *Zea mays*, 10 DcGRFs from *Dendrobium catenatum*, 10 AsGRFs from *Apostasia shenzhenica,* and 9 PeGRFs from *Phalaenopsis equestris.* Based on alignment, a Maximum likelihood (ML) phylogenetic tree was constructed using MEGAX with 1,000 bootstrap replicates and Jones-Taylor-Thornton method (Kumar et al., 2016). The online software iTOL was applied to edit the phylogenetic tree.[3] The conserved motifs of CeGRF proteins were analyzed online using Multiple Expectation Maximization for Motif Elicitation (MEME) with default parameters.[4]

Plasmid Construction and Tobacco Transformation

We generated cDNAs for the precursor fold-back structure of *Ce*-miR396 by reverse transcription. After verification by sequencing, we inserted pre-*Ce*-miR396 into the pCambia1300 vector downstream of the constitutive Cauliflower Mosaic Virus 35S promoter. This construct was introduced into Agrobacterium tumefaciens strain GV3101 for transformation of tobacco by the leaf-disk method, as described in our previous work (Yang et al., 2009). Transgenic plants were screened on an MS medium containing 10 mg/L hygromycin, and 21 independent lines were obtained.

[3]http://itol.embl.de/

[4]http://meme-suite.org/tools/meme

Scanning Electron Microscopy

Dissected apices of mature flowers were fixed in a solution of 3% glutaraldehyde and 2% formaldehyde for 24 h. Samples were dehydrated in acetone, critical-point dried in liquid CO_2, and mounted on stubs and sputter coated with 25 nm gold. Samples were examined using a JSM-6360LV (JEOL) scanning electron microscope.

RESULTS

Ontogeny of *Cymbidium ensifolium* Flower Development

Cymbidium ensifolium commonly takes 3 years to reach reproductive maturity. Inflorescence meristems (IM) are produced in the peripheral regions of axillary buds and 6 to–11 flowers develop from each IM (**Supplementary Figure 1**). We divided floral development into six stages (from 0 to 5) according to visible changes in flower morphology (**Figure 1**). A crescent bract primordium initiates around the inflorescence meristem (**Figure 1A**), then the flower meristem emerges with a flattened and oval flower meristem (**Figure 1B**, stage 0), which continues to enlarge and form a floret primordium (FP) (**Figure 1C**, stage 1). A central transversal depression can be observed in FP, from which a labellum primordium (LP) initiates in the adaxial region, followed by development of two lateral sepal primordia (SP). The abaxial portion of primordia then enlarges, turning into two lateral petals and median sepal primordia. The column primordium finally emerges from the central region, establishing a typical orchid floral zygomorphy in this phase (**Figures 1D,E**, stage 2). This process is similar in the closely related species *C. sinense*, but develops much faster with a completion time of about 2 days (Su et al., 2018). During stage 3, the floral apex resembles an inverted triangle, with outer sepals overlapping the inner petals (**Figures 1F,G**). The ventral outer sepal and petals grow quickly and cover the column, which consists of an empty locule with no pollinia (**Figure 1H**). At stage 4, two ventral inner petals progressively shelter the labellum and column. The labellum develops a kinked to undulated margin, and the edge of labellum begins to curl toward column (**Figures 1I,J**). The column rapidly elongates at this stage but the pollinia are still not mature (**Figure 1K**). At stage 5, floral organ development completes with three petal-like sepals in the first whorl, two lateral petals and a specialized bottom petal (the lip or labellum) in the second whorl, and a fine-structured column in the central part (**Figures 1L,M**). The column includes carpel and stamens that are differentiated and evolved through the complete fusion of style, stigma, and staminal filament, and has four pollinia on a semi-circular viscidium (**Figure 1N**).

Transcriptome Profiling

To identify key regulators controlling the specification and growth of individual floral organs, we isolated total RNA from

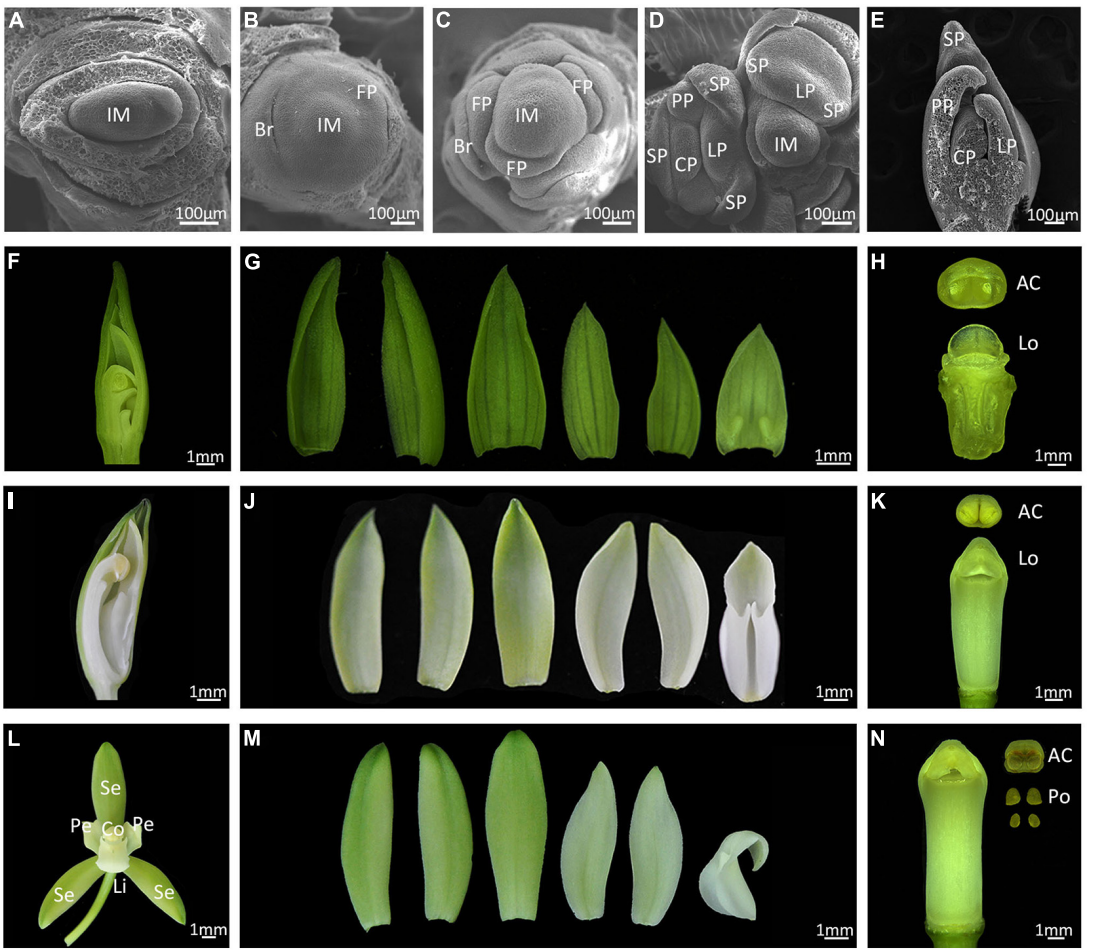


FIGURE 1 | Floral development of *Cymbidium ensifolium*. **(A–E)** Scanning electron micrograph (SEM) of early floral developmental stages. Bar = 100 μm. Im, inflorescence meristem; Br, bract; FP, floret primordium; SP, sepal primordium; PP, petal primordium; LP, labellum primordium; CP, column primordium. **(F–N)** Developing flowers, Bar = 1 mm. The developing flower of stage 3 **(F–H)**, stage 4 **(I–K)**, and mature flowers **(L–N)**. Se, sepal; Pe, petal; Li, lip; Co, column; AC, Anther Cap; Lo, locule; Po, pollinium.

each floral organ during stage 3. At this stage, all flower organ primordia differentiation is completed, enabling a comparative transcriptome analysis. From six pairwise comparisons between sepals, petals, labellum and gynostemium, we identified between 108 and 2,167 differentially-expressed unigenes in individual pairwise comparisons. A total of 3,017 unigenes exhibited significant changes in expression. The number of DEGs between tissues correlated with the degree of their morphological differences. The largest differences occurred between the gynostemium and sepals, with 1,324 up-regulated and 843 down-regulated transcripts. Moreover, 1,143 and 738 transcripts were differentially expressed in the gynostemium compared with petals (**Supplementary Figure 2A**). The smallest number of DEGs (108 transcripts) were found between sepals and petals. The number of DEGs in other pairwise comparisons ranged from 1,091 to 1,881 transcripts (**Supplementary Table 2**).

Gene Ontology (GO) enrichment analysis identified 41 GO terms significantly enriched among different floral organs. The main GO biological processes included oxidation-reduction and metabolism. The cellular components were mainly assigned to mitochondrion, plastids, and plasma membrane. The molecular functions predominantly involved metal ion binding and ATP binding (**Supplementary Figure 2B** and **Supplementary Table 3**). We identified 334 enriched KEGG pathways. The "metabolic and biosynthesis of secondary metabolites" pathway had the most DEGs, followed by "photosynthesis" and "plant hormone signal transduction" (**Supplementary Table 4**). Non-redundant (NR) annotation results indicated that most of the enriched transcripts were connected to plant metabolic processes, oxidation-reduction processes, and establishment of cell structures. We also observed a number of DEGs involved in secondary metabolism and oxidation-reduction pathways, such as genes encoding the enzymes, including acetyltransferase, transketolase, and NADP dependent oxidoreductase. Notably, many TFs among our DGEs; specifically, a large fraction (8 out of 11) of *GRF* family members were up-regulated in the inner whorls of flower (the column), suggesting a putative role in flower organ development and/or differentiation (**Supplementary Figure 2C**).

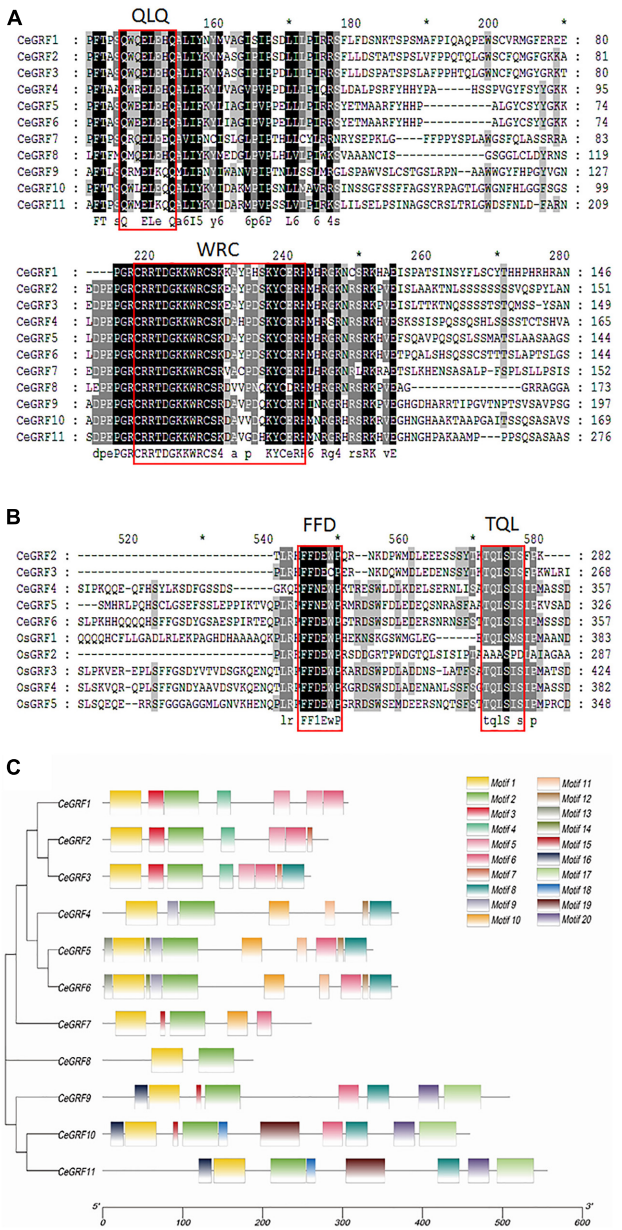


FIGURE 2 | Sequence alignment and functional domain analysis of *Cymbidium ensifolium Growth-Regulating Factor* (CeGRF) proteins. The conserved QLQ and WRC domains in all CeGRFs **(A)** and the FFD and TQL motifs identified in CeGRF2, 3, 4, 5, and 6 **(B)** are indicated by the red box. Identical amino acids are indicated by black and the amino acids with >50% similarity are indicated by gray background. **(C)** Twenty conserved motifs labeled with different colors were identified in the CeGRF sequences using the MEME program. Among them, motif 1 and motif 2 are the QLQ and WRC conserved domains.

Identification and Sequence Analysis of *Cymbidium ensifolium Growth-Regulating Factor* Genes

We identified 11 *CeGRFs* coding sequences in our transcriptome dataset. Considerable variation in length was present in their coding sequences, ranging from 468 to 1,788 bp. The deduced proteins varied between 155 and 595 amino acids (**Supplementary Table 5**). All of the putative CeGRF proteins shared the conserved QLQ and WRC domains in their N-terminal regions (**Figure 2A**). A zinc finger motif (CCCH) was also present within the WRC domains of all CeGRF proteins. Four CeGRFs (CeGRF2, CeGRF3, CeGRF4, CeGRF5, and CeGRF6) shared short stretches of amino acids termed FFD (phenylalanine, phenylalanine, and aspartate) and TQL (threonine, glutamine, and leucine) domains near their C-termini (**Figure 2B**).

The Multiple Em for Motif Elicitation (MEME) program predicted 20 distinct protein motifs in CeGRF proteins (**Figure 2C**). All CeGRFs contained motif 1 (QLQ motif) and motif 2 (WRC motif) at their N-terminus. Each CeGRF possessed 5–10 conserved motifs, except CeGRF8 which only possessed motifs 1 and 2 and showed an overall low sequence similarity with other GRFs. The closely related CeGRF1, CeGRF2, and CeGRF3 shared motifs 1–6; CeGRF2 presented domain 7 as well, while CeGRF3 had both motifs 7 and 8. The more distant CeGRF4, CeGRF5, and CeGRF6 contained the QLQ and WRC motifs, as well as motifs 8–14. The remaining CeGRF9, CeGRF10, and CeGRF11 contained additional motifs 15–20. Detailed information for each motif is listed in **Supplementary Figure 3**.

We conducted a phylogenetic analysis of GRFs from different species to understand their evolutionary relationships. As shown in **Figure 3** and **Supplementary Figure 4**, 92 GRFs clustered into six subgroups (I–VI) by the Maximum Likelihood phylogenetic tree method constructed by MEGA X. Subgroup II contained clusters of GRFs from monocot species only, while subgroup V was specific to dicot species only. CeGRFs were distributed across 5 out of 6 subgroups and were not represented in subgroup V, consistent with their classification as monocots. Subgroups III and V were relatively small, with only eight and nine GRFs each, respectively. By contrast, subgroup IV contained the largest number of GRFs (twenty-eight), followed by subgroups II (nineteen), I (seventeen), and VI (eleven). Notably, we found GRFs from orchid plants (*Cymbidium ensifolium, Dendrobium catenatum*, *Apostasia shenzhenica,* and *Phalaenopsis equestris*) within subgroups I, II, and III, which clustered away from monocot and dicot plants, indicating that GRFs are highly conserved between closely related orchid species and are clearly separated from other genera and more distant species.

Cymbidium ensifolium Growth-Regulating Factors Are Regulated by *Cymbidium ensifolium*-miR396

Arabidopsis *GRFs* are the predicted target genes for 21-nt microRNA miR396. *CeGRF* genes (except *CeGRF7*) share a highly similar sequence over the length of predicted miR396 complementary sequence, with only one mismatch and one bulge in the miR396 binding site (**Figure 4A**). Through degradome sequencing which is a powerful and efficient approach for the validation of miRNA-target genes, we observed that the transcripts of 7 out of 11 *CeGRFs* were cleaved by miR396

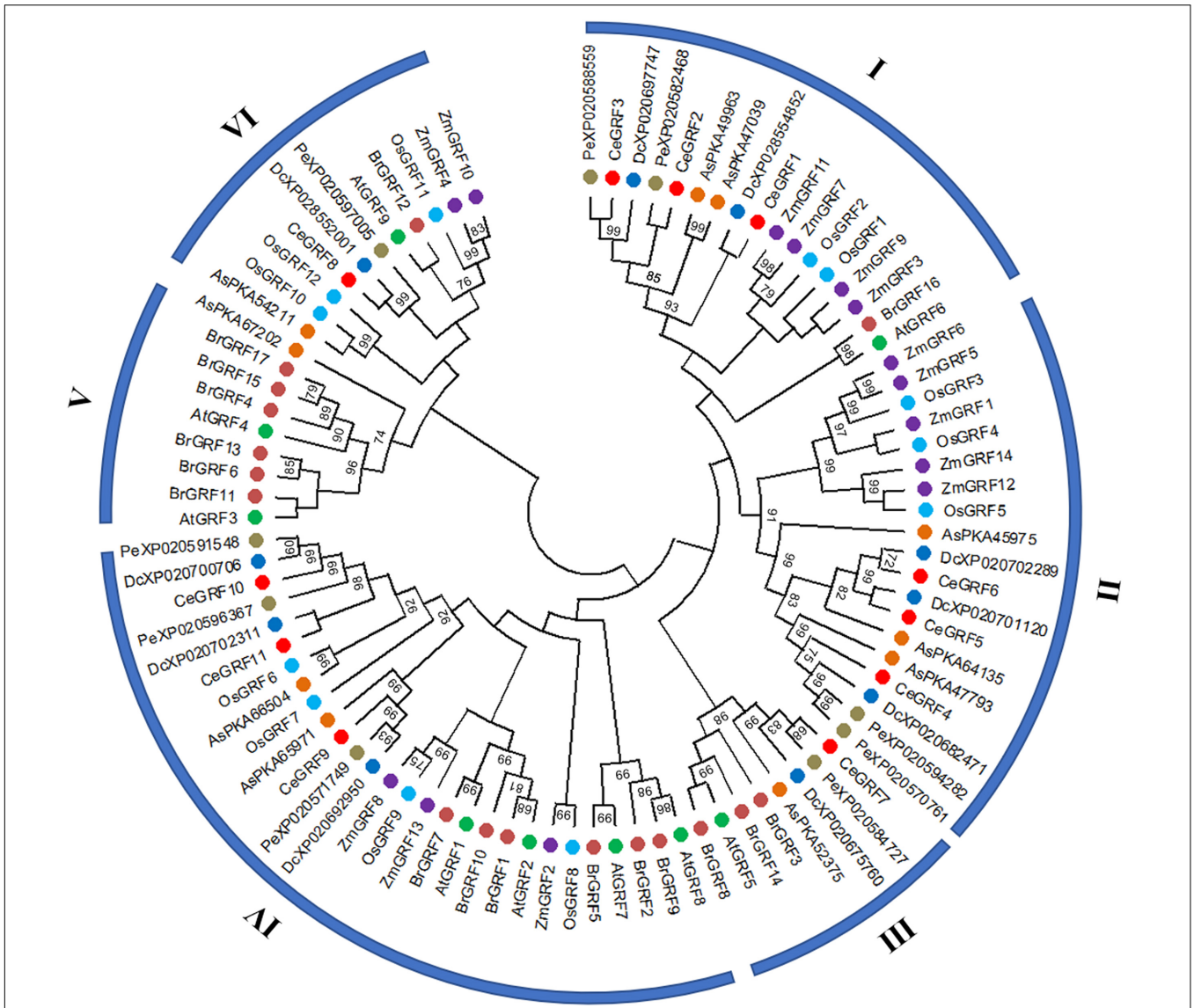
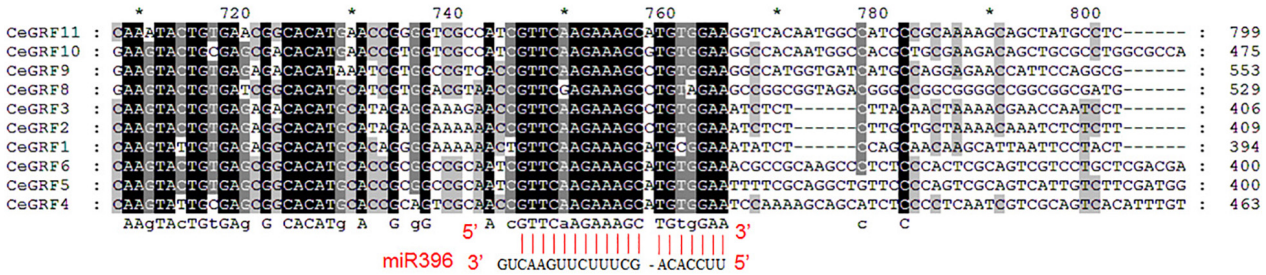


FIGURE 3 | Phylogenetic relationships among GRFs. Neighbor-joining phylogenetic tree was constructed by MEGAX.0 software with the Poisson model and 1,000 bootstrap computations. The GRF proteins were grouped into six distinct types, and represented GRFs from *Arabidopsis thaliana* (At), *Oryza sativa* (Os), *Brassica rapa* (Br), *Zea mays* (Zm), *Dendrobium catenatum* (Dc), *Apostasia shenzhenica* (As) and *Phalaenopsis equestris* (Pe). Each species is represented by a different color.

at the correct predicted position [the 11th nucleotide "A" of the target sequence "UCGUUCAAGAAaGCGUG(A)UGGA"] in *C. ensifolium*, with significant alignment scores between 1 and 3.5 and associated *p*-values ≤ 0.05. *CeGRF6* showed the largest normalized read counts out of all detected *CeGRFs*, followed by *CeGRF10*, indicating a strong *in vivo* signal (**Figure 4B** and **Supplementary Table 6**).

To further confirm the miR396-*GRF* regulatory module in *C. ensifolium*, we searched for *Ce*-miR396 precursor in our transcriptome dataset and identified a 124 bp sequence containing mature *Ce*-miR396 and complementary *Ce*-miR396* (**Figure 4C**). We overexpressed it in *C. ensifolium* using an improved and robust protoplast-based transient expression system (PTES) (Ren et al., 2020; **Figures 5A,B**). Mature *Ce*-miR396 was successfully overexpressed in *Cymbidium* protoplasts, increasing 150 to 350-fold within 36 h after transfection, which was accompanied by the down-regulated expression of *CeGRFs*, in agreement with the results from degradome sequencing. Notably, in leaf protoplasts, *CeGRF2*, *CeGRF3*, *CeGRF4*, *CeGRF5,* and *CeGRF6* transcript levels decreased significantly starting 24 h after transfection, reaching 40–60% in 36 h after transfection compared to time 0, while *CeGRF1*, *CeGRF9*, *CeGRF10,* and *CeGRF11* transcripts were not affected (**Figure 5C**). Contrarily, in flower protoplasts, the cleavage site-spanning fragments of *CeGRF9*, *CeGRF10,* or *CeGRF11* were almost undetectable, while no significant change in *CeGRF2*, *CeGRF3*, *CeGRF4*, *CeGRF5,* or *CeGRF6* expression was observed (**Figure 5D** and **Supplementary Figure 5**).

A



B

Match	Transcript name	Alignment Score	Alignment Range	Cleavage Site	Read Number
target 5' UCGUUCAAGAAaGCGUGUGGA 3' query 3' GUCAAGUUCUUUCGGACACCU 5 '	*CeGRF10*	3.5	411-431	60	145
target 5' UCGUUCAAGAAaGCAUGUGGA 3' query 3' GUCAAGUUCUUUCGGACACCU 5 '	*CeGRF11*	3.5	741-761	197	45
target 5' CCGUUCAAGAAaGCAUGUGGAA 3' query 3' ACCAAGUUCUUUCG-ACACCUU 5 '	*CeGRF3*	3	354-374	1361	45
target 5' UCGUUCAAGAAaGCUUGUGGAA 3' query 3' GUCAAGUUCUUUCG-ACACCUU 5 '	*CeGRF6*	2.5	336-356	594	200
target 5' UCGUUCAAGAAaGCAUGUGGAA 3' query 3' ACCAAGUUCUUUCG-ACACCUU 5 '	*CeGRF5*	3.5	336-356	787	5
target 5' UCGUUCAAGAAaGCAUGUGGAA 3' query 3' GUCAAGUUCUUUCG-ACACCUU 5 '	*CeGRF4*	3.5	399-419	623	2
target 5' CCGUUCAAGAAaGCCUGUGGA 3' query 3' GUCAAGUUCUUUCGGACACCU 5'	*CeGRF2*	1	357-377	940	2

C

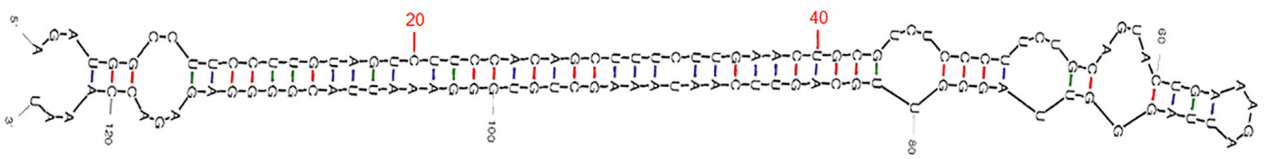


FIGURE 4 | *Cymbidium ensifolium Growth-Regulating Factors* are Targeted by *Ce*-miR396. **(A)** Complementary matching between miR396 and *CeGRFs*. **(B)** Identification of the targeted GRF genes of miR396 by degradome sequencing, read number indicated normalized reads that matched the different positions of the targeted genes. Only the cleavage sites with *P* value < 0.05 were counted. **(C)** Secondary structure of pre-*Ce*-miR396 of *Cymbidium ensifolium*. The mature sequences that match the precursor detected by small RNA-Seq are displayed in red.

Differential regulation of individual genes is likely due to different tissue-specific expression patterns of individual *CeGRF* genes.

Expression Dynamics of *Cymbidium ensifolium-miR396/Cymbidium ensifolium Growth-Regulating Factor* Tightly Correlate With Reproductive Organ Development

To validate the results obtained from DGEs and further explore the role of *Ce*-miR396/*CeGRF* in flower development, we analyzed their expression patterns in different floral developmental stages and organs by stem-loop (for *Ce*-miR396)- or RT (for *CeGRFs*)- qPCR. In general, all *CeGRF* genes were predominantly expressed in young buds (stages 1–4), reaching 15–2,500 fold higher levels than in other tissues, while *Ce*-miR396 was markedly low in these stages (**Figure 6**). The expression levels of *CeGRFs* were in general the lowest in roots, stems, and leaves, except for *CeGRF7* and *CeGRF8*, which showed the lowest expression levels in mature flowers. *CeGRF1*, *CeGRF4*, *CeGRF5,* and *CeGRF11* exhibited similar expression patterns and showed relatively high transcript levels in flowers, and the lowest levels in roots. *CeGRF2*, *CeGRF3*, *CeGRF8*, *CeGRF9*, *CeGRF10,*

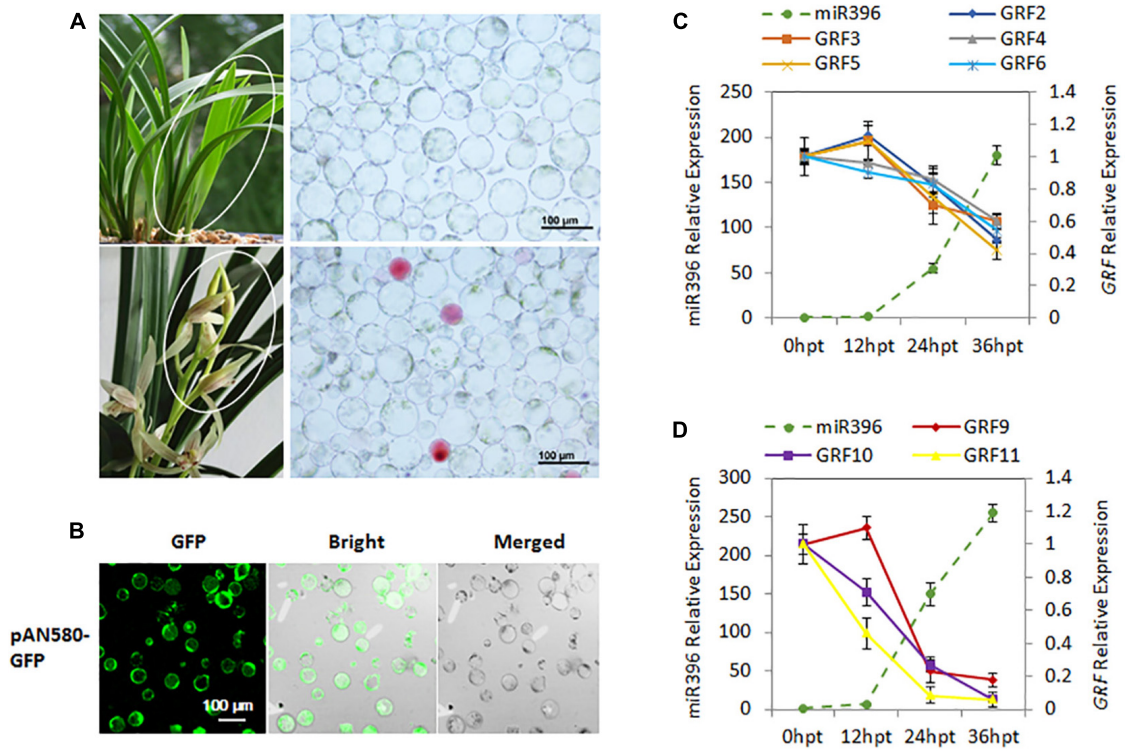


FIGURE 5 | Validation of *Ce*-miR396-Regulated *CeGRFs* expression in a *C. ensifolium* Protoplast-Based Transient Expression System. **(A)** Protoplast isolation from the leaves (top) and flowers (bottom). **(B)** Efficient protoplast-based transient expression system (PTES) in *Cymbidium ensifolium*. The green fluorescence of the GFP reporter is clearly visible in most cells. The maximum transfection efficiency of pPAN-580-GFP (4.71 kb in size) was ∼80%. Bar = 100 μm. **(C,D)** Transcript levels of *Ce*-miR396 (dotted line) and *CeGRF* genes (solid lines) in leaves **(C)** and flowers **(D)**. Mature *Ce*-miR396 was successfully overexpressed in *Cymbidium* protoplasts 24–36 h after transfection, and the expression levels of *CeGRFs* were significantly reduced. The Y-axis indicates fold change in expression at different time points. Expression levels were normalized using the threshold cycle values obtained for the Ubiquitin gene. Error bars indicate the standard deviation of the mean (SD) (n = 3). Three replicates were analyzed, with similar results.

and *CeGRF11* were highly expressed during floral developmental stages 1–3. *CeGRF6* and *CeGRF7* were more strongly expressed in floral developmental stages 2–4 compared to other tissues. By contrast, the expression of mature *Ce*-miR396 was low in developing floral buds and high in roots and leaves, generally showing the opposite expression pattern from *CeGRFs* (**Figure 6**).

Cymbidium ensifolium growth-regulating factors, except *CeGRF7*, showed higher expression in inner whorls than in the other tissues (**Figure 6**). *CeGRF1*, *CeGRF4*, *CeGRF5,* and *CeGRF8* were mainly expressed in the column. *CeGRF3*, *CeGRF6*, *CeGRF9,* and *CeGRF10* were highly expressed in the lip and column, reaching to 5–150 fold increase in expression, consistent with the high transcriptome expression of *CeGRFs* in the inner flower whorls. By contrast, *Ce*-miR396 showed relatively high expression in petal and sepal, and the lowest expression in column (**Figure 7A**). Thus, these results revealed a strong negative correlation between *Ce*-miR396 and *CeGRFs* transcript levels, with antagonistic distribution mainly in the column.

Cymbidium ensifolium evolves a number of natural variation types of flower organs according to floral organ transformation and/or reversion. In this study, three typical natural flower morphology mutants, which produce null-labellum flowers (Var1), stamenoid-like tepals (Var2), or multi-tepals (Var3), were employed to determine the relationship between *CeGRF* gene expression and flower development. When compared with standard flowers (ST), most *CeGRFs* (with the exception of *CeGRF4* and *CeGRF7*) were more strongly expressed in the varieties that developed stamenoid-tepal (Var 2) or multi-tepal (Var 3) flowers, with 2.5–270 fold increase in expression (**Figure 7B**). *Ce*-miR396 displayed the same trend (**Figure 7B**). However, reduced or comparable expression levels were detected in Var 1, which developed similar perianth structures with no lips or column, suggesting a strong correlation between *CeGRF* expression and reproductive organ development. By contrast, *CeGRF4* and *CeGRF7* had comparable expression levels in different floral organs, possibly reflecting functional differences among gene family members.

Regulation of Flower Development by *Cymbidium ensifolium*-miR396 Overexpression in Tobacco

To further dissect the function of *Ce*-miR396, we generated transgenic tobacco plants overexpressing *Ce*-miR396. We obtained 21 independent transgenic plants, a subset of which was analyzed for the presence of *Ce*-miR396. As shown in **Figure 8A**,

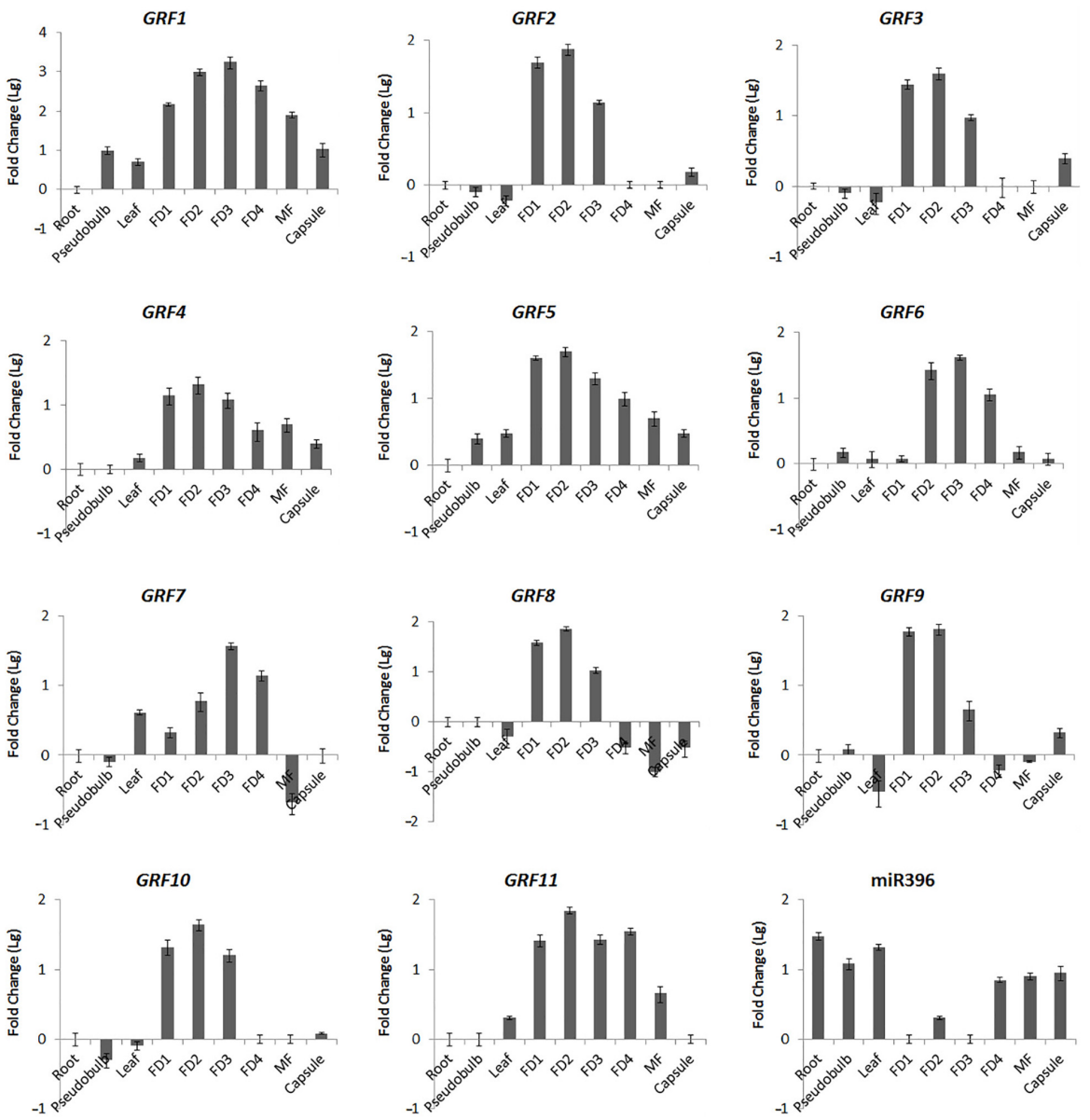


FIGURE 6 | Organ-Specific Expression of *CeGRFs* and *Ce*-miR396. Expression of *CeGRF* genes and *Ce*-miR396 in different organs examined by Real-time RT-qPCR and stem-loop PCR. Tissues analyzed: roots, pseudobulb, leaves, five stages of Floral Development (FD 1, 2, 3, 4, and mature flower (MF). The Y-axis indicates fold change (Log value) in expression among different plant organs and the floral buds at different developmental stages. Expression levels were normalized using the threshold cycle values obtained for the Ubiquitin gene. Root and FD1 were used as a standard in presenting fold change (Lg) of *CeGRFs* and *Ce*-miR396, respectively. Error bars indicate the standard deviation of the mean (SD) (n = 3). Three replicates were analyzed, with similar results.

All T_1 plants had higher levels of *Ce*-miR396 ranging from 3.5 to 245 folds increase compared with wild type and empty construct. We further searched the tobacco genome sequence database and 41 putative *NtGRF* sequences were obtained. Among them, eight sequences with the highest identity with arabidopsis *GRFs* were selected for expression verification in wild-type and transgenic tobacco plants using quantitative RT-PCR (**Supplementary Figure 6A**). By amplifying the cleavage site-spanning fragment, we determined that, compared with WT, levels of eight *NtGRF-like* genes were significantly decreased in three independent transgenics lines 6, 7, and 8 (**Supplementary Figure 6B**), indicating a strong negative regulation of *NtGRF* transcript levels by *Ce*-miR396.

The transgenic plants were shorter and had small and narrow leaves compared to wild type (**Figure 8B**). These phenotypes resembled those seen in the *grf1 grf2 grf3* triple mutant in arabidopsis and transgenic tobacco plants overexpressing arabidopsis miR396 (Kim and Kende, 2004), further supporting a conserved role of *Ce*-miR396 reflected by this narrow-leaf phenotype. The perianth of transgenic plants was similar in

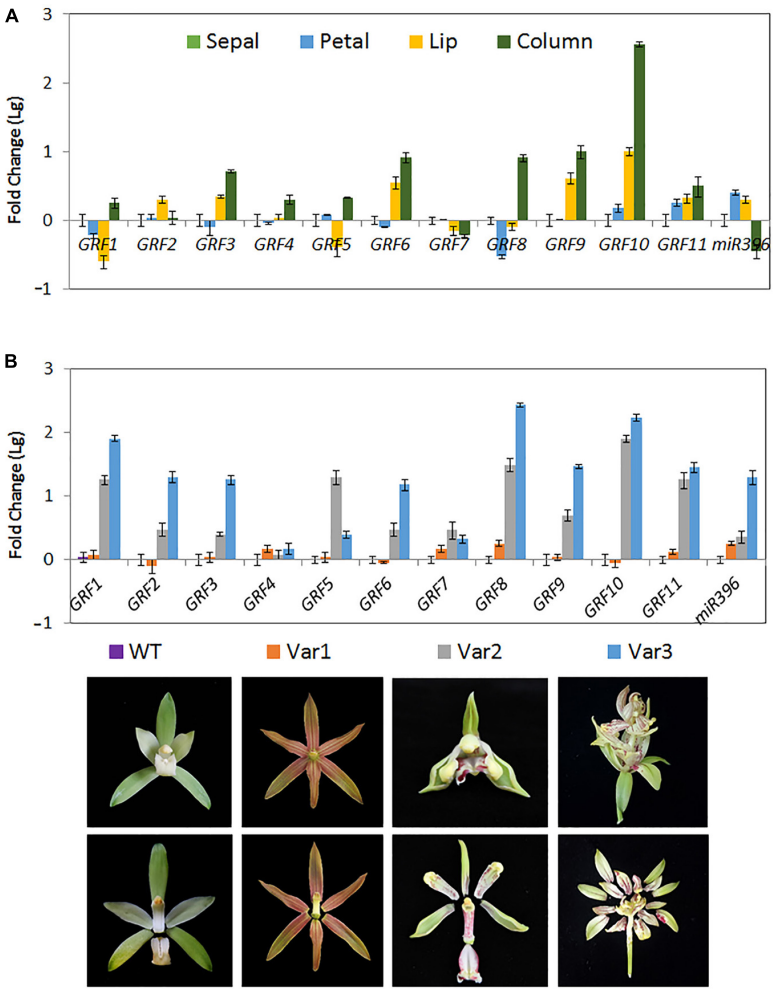


FIGURE 7 | Expression of *Ce*-miR396/*CeGRF* Correlates with Flower Development. Expression of *CeGRF* genes and *Ce*-miR396 in different floral organs examined by RT-qPCR and stem-loop PCR, respectively. We analyzed individual floral organs (sepal, petal, lip, and column) **(A)** and the whole flowers of three floral varieties that develop non-lip (Var1) or stamenoid-tepal (Var2) or multi-tepal flowers (Var 3) **(B)**. The Y-axis indicates fold change (Log value) in expression among different floral organs. Expression levels were normalized using the threshold cycle values obtained for the Ubiquitin gene. The samples from sepal and the wild-type flower (WT) were used as a standard to present fold change (Lg). Error bars indicate the standard deviation of the mean (SD) (n = 3). Three replicates were analyzed, with similar results.

structure with that of wild type but was severely curved (**Figures 8C,D**). The stamens were shorter and some of the anthers developed abnormally with no pollen. Even when the anthers properly released their pollen, stamens were much shorter than pistils, leading to partial sterility, which is similar to the phenotypes associated with arabidopsis lines overexpressing miR396 (**Figures 8E,F**).

DISCUSSION

Expression Patterns of Growth-Regulating Factor Genes in *Cymbidium ensifolium*

Orchid floral development is a hot research topic. We previously published the analysis of 111,892 *C. ensifolium* transcript clusters derived from *de novo* assembly of flower transcriptome (Yang and Zhu, 2015). In the present study, we dissected the *C. ensifolium* flower ontogeny for the first time, and identified 3,017 DEGs by comparative transcriptome analysis among different floral organs. Most DEGs were involved in plant metabolic processes, oxidation-reduction, plant hormone signaling and establishment of cell structures. These are related to the developmental processes underlying flower development, i.e., cell division, membrane-building, and regulation of anabolism. We also identified a number of TFs from several families, including the well-known floral-related MADS-box, NAC, NF-YC, TALE, as well as bZIP, bHLH, FAR1, Zn-finger, and MYB, highlighting the importance of TFs in organ-specific development (**Supplementary Table 7**). Notably, we observed significant expression of most *CeGRF* genes in the

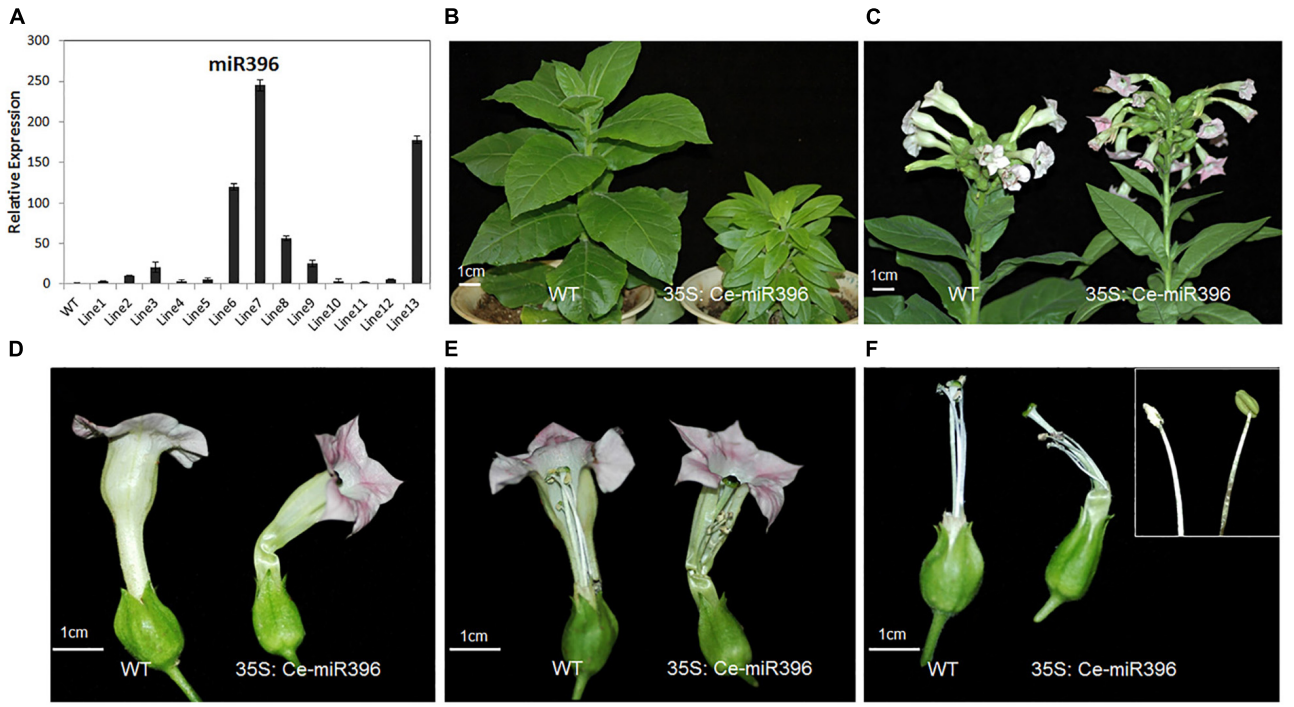


FIGURE 8 | Phenotypes of Flowers Overexpressing Ce-*miR396* in Tobacco Plants. **(A)** Increased expression level of Ce-*miR396* in the transgenic lines. The Y-axis indicates fold change in expression between the wild-type and transgenic plants. Expression levels were normalized using the threshold cycle values obtained for the Ubiquitin gene. Error bars indicate the standard deviation of the mean (SD) (n = 3). Three replicates were analyzed, with similar results. **(B)** Reduced height and leaf size observed in transgenic plants. **(C–F)** Comparison of WT and transgenic flower inflorescence **(C)**, perianth **(D)**, pistils **(E)**, and filaments **(F)**. Bar = 1 cm.

column, signifying the possible role of *CeGRFs* in *C. ensifolium* flower development.

Investigating tissue specificity in expression patterns can provide clues to tissue development. Indeed, *GRF*s are expressed mainly in specific organs and tissues. For example, in *Medicago truncatula*, eight *GRF* genes exhibited lower expression in leaves compared with their expression in other tissues of roots, shoots, and flowers (Bazin et al., 2013). In Chinese cabbage, seven (out of 17) *BrGRF* genes showed higher expression in young leaves compared to other tissues (Wang et al., 2014). In cucumber (*Cucumis sativus*), *CsGRFs* were highly transcribed in ovaries, although *CsGRF4* and *CsGRF6* exhibited the highest expression in leaves, and *CsGRF7* was only detected in roots (Shi et al., 2019). Collectively, these studies highlight the tissue-specific expression diversity of *GRFs* in various plant species (Cao Y. et al., 2016).

Compared to *GRFs* in other plants, our results showed that 10 out of the 11 *CeGRF* genes were predominantly expressed in fast-growing floral buds. In addition, we detected much higher expression of *CeGRF* genes in the column of both wild type and varieties with an over-developed column (Var 2 and Var3, **Figure 7B**), and a converse lower expression in Var1 with no column. These results reveal the important role of *GRFs* in flower development, specifically in column formation and growth. Moreover, it is worth noting that genes belonging to the same phylogenetic groups displayed a similar expression pattern, for example *CeGRF2*/*CeGRF3* in subgroup I, *CeGRF4*/*CeGRF5* in subgroup II, and *CeGRF9*/*CeGRF10* in subgroup IV (**Figure 3**). This suggests that members of same group, sharing the same functional domains, play similar functions in flower development.

Different Spatio-Temporal Activity of *Cymbidium ensifolium-miR396*/*Cymbidium ensifolium Growth-Regulating Factors* in *Cymbidium ensifolium*

MiR396 is an evolutionarily conserved miRNA that recognizes a complementary sequence in *GRF* genes. The components of this miR396–GRF module are conserved in seed plants and work antagonistically to affect the development of various organs in many species. A few examples of development and responses mediated by the miR396–GRF module include floral organ growth, development and fertility (rice and maize), seed and grain size (rice, arabidopsis, *Brassica napus*, and tomato), shoot and inflorescence growth (arabidopsis, tomato, rice, and maize), flowering time (arabidopsis and rice), pathogen resistance (Arabidopsis and rice), photosynthetic activity (arabidopsis and rice), senescence delay arabidopsis drought tolerance (arabidopsis), root growth (rice), leaf angle (rice and maize) and embryogenic transition (arabidopsis) (Kim et al., 2003, 2012; Horiguchi et al., 2005; Luo et al., 2005; Kuijt et al., 2014; Li et al., 2016; Kim, 2019).

This clearly illustrates that miR396 and *GRFs* may have different regulatory roles in different species during plant development, as a consequence of different activity, over time and development, of the downstream target genes. In our study, *CeGRFs* were regulated differently in different organs. When *Ce*-miR396 was overexpressed in a *C. ensifolium*, a subset of *CeGRFs* (*CeGRF2*, *CeGRF3*, *CeGRF4*, *CeGRF5*, and *CeGRF6*) behaved similarly in leaf protoplasts, exhibiting a similar expression pattern and *Ce*-miR396-mediated cleavage, lowering their transcript levels up to 40–60%. The same *CeGRFs* were, however, not affected by *Ce*-miR396 overexpression in flower protoplasts. By contrast, *CeGRF9*, *CeGRF10* and *CeGRF11* exhibited the opposite behavior. The transcript levels of *CeGRF1* and *CeGRF7* showed no significant differences in either case. These results indicate the functional diversity of individual members. In addition, *CeGRF9*, *CeGRF10,* and *CeGRF11* may play important roles in flower development, under the regulation of *Ce*-miR396.

Floral Development Regulation of *Cymbidium ensifolium* by *Cymbidium ensifolium-miR*396/*Cymbidium ensifolium Growth-Regulating Factor*

The miR396/*GRF* module in both eudicots and monocots plays essential roles in the growth and development of floral organs, although the specific mechanism remains unclear for many species. Detailed developmental studies of various higher order loss-of-function mutants like *grf1 grf 2 grf 3* and associated co-activators *gif1 gif 2 gif 3*, as well as *gif* mutants overexpressing miR396 in arabidopsis have uncovered the individual contribution of specific family member in reproductive development (Liang et al., 2014). Arabidopsis GIFs are essential factors for carpel and stamen development, regulating the formation and maintenance of meristematic structures of male and female reproductive organs. In addition, the functionality of the miR396/*GRF* module is conserved in eudicots floral organ development, causing aberrant floral organs reminiscent of *grf* and *gif* mutant phenotypes (Kim and Tsukaya, 2015; Lee et al., 2018).

However, this conservation does not extend to monocots, because the aberrant floral development caused by miR396/GRF in monocots is morphologically distinct from that of eudicots (Luo et al., 2005). Overexpression of rice miR396, or double mutants in the rice *GRF6* and *GRF10* genes produced open husks, long sterile lemmas, and/or anomalous stigma and anther numbers (Duan et al., 2015). The maize *gif1* mutant produced sterile male and female flowers, with multiple silks (pistils) per floret and nucellus protruding out of the carpel (Wu et al., 2014; Zhang et al., 2018). This diversity of phenotypes was explained by the interaction of GIF with tissue-specific GRFs, leading to the regulation of distinct downstream targets. A complementary mechanism may also call upon the regulation of *GRFs* by different miR396 loci. For example, in arabidopsis, overexpression of miR396a causes a more severe floral organ defect phenotype than that of miR396b. Notably, the TF HECATE1 (Gremski et al., 2007), which controls carpel development in arabidopsis, is a direct target gene of miR396/*GRF* module, whereby *GRF* levels influence floral organ growth and development. In rice, *Os*-miRNA396d and its targets *Os*GRF6 and *Os*GRF10, together with *Os*GIF1, are involved in the regulation of floral organ development through Jumonji C domain-containing histone demethylase *Os*JMJ706 and the crinkly4 receptor-like kinase OsCR4 (Liu et al., 2014).

In this study, we conclude that *Ce*-miR396 and *CeGRF9*, *CeGRF10*, and *CeGRF11* are involved in the regulation of flower development based on: the positive correlation between gene expression and phenotypes, target gene cleavage by *Ce*-miR396 in flower resulting in decreased *CeGRF* transcript levels, and the phenotypes caused by *Ce*-miR396 overexpression in tobacco. Considering the preferential expression of *CeGRF9*, *CeGRF10,* and *CeGRF11* in the column of wild type and varieties with over-development of column, and their lower expression in a variety lacking the column structure, we hypothesize that *CeGRF9*, *CeGRF10,* and *CeGRF11* function in column development. Notably, *CeGRF9*, *CeGRF10,* and *CeGRF11* belongs to subgroup V and clustered closely with genes that were reported to function in floral development in other plants, such as arabidopsis *GRF1*, *GRF2,* and *GRF3*, as well as rice *GRF6*. This indicates a conserved role across eudicot and monocot species despite the different consequences of their respective overexpression, such as open husks for *Os*GRF6, or fused or curved flowers for arabidopsis GRF1, GRF 2, and GRF3. However, the orchid floral organs, and especially the column, are unique floral structures that may indicate a shift in protein functions and interactions in floral homeotic genes (Mondragón-Palomino and Theißen, 2008; Hsu et al., 2015). In this sense, screening of target genes downstream of GRFs and their co-activator GIF will further extend our knowledge of the regulation of *Ce*-miR396/*Ce*GRF in orchids flower development.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

FY and GZ designed the experiments and wrote the manuscript with input from all authors. GL and YW analyzed the data. JJ and CL executed the experiments and assembled the figures. JW conducted the qRT-PCR. SA and YX edited the manuscript. All authors read and approved the final manuscript.

FUNDING

This research was funded by grants from National Key R&D Program (2018YFD1000404 and 2019FD1001003), the National Natural Science Foundation of China (31872151 and 31672184), the Natural Science Foundation of Guangdong

province (2017A030312004), Guangzhou Science and Technology Project (201707010307), Innovation Team of Modern Agriculture Industry Technology System in Guangdong Province (2021KJ121), and Guangdong Academy of Agricultural Sciences Discipline Team Construction Project (202127TD and BZ202006).

ACKNOWLEDGMENTS

We are thankful to funding agencies for funding and Plant Editors for providing professional services for language editing and final polishing of our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.799778/full#supplementary-material

Supplementary Figure 1 | Floral development process of *Cymbidium ensifolium.*

Supplementary Figure 2 | Comparative transcriptome analysis of individual floral Organs. **(A)** Transcripts differentially expressed between different floral organs. Up- and down-regulated transcripts were quantified. The results of six comparisons between each two samples are shown. Se, sepal; Pe, petal; Li, Lip; Co, column. **(B)** GO term classification of differentially-expressed unigenes. **(C)** Expression heat map of *CeGRF* genes in sepal, petal, Lip, and column. Expression values from RNA-seq data were $\log_2$-transformed and are displayed as filled blocks in blue to red indicating gene expression intensity from low to high.

Supplementary Figure 3 | Detailed information for protein motifs predicted in CeGRFs.

Supplementary Figure 4 | Phylogenetic analysis of CeGRFs and AtGRFs.

Supplementary Figure 5 | *Cymbidium ensifolium*-miR396-regulated *CeGRFs* expression in a *Cymbidium ensifolium* protoplast-based transient expression system.

Supplementary Figure 6 | *Cymbidium ensifolium*-miR396-regulated *NtGRFs* expression in transgenic tobacco plants.

REFERENCES

Aceto, S., and Gaudio, L. (2011). The MADS and the Beauty: Genes Involved in the Development of Orchid Flowers. *Curr. Genomics* 12, 342–356. doi: 10.2174/138920211796429754

An, F. M., and Chan, M. T. (2012). Transcriptome-wide characterization of miRNA-directed and non-miRNA-directed endonucleolytic cleavage using Degradome analysis under low ambient temperature in *Phalaenopsis aphrodite* subsp. formosana. *Plant Cell Physiol.* 53, 1737–1750. doi: 10.1093/pcp/pcs118

Bazin, J., Khan, G. A., Combier, J. P., Bustos-Sanmamed, P., Debernardi, J. M., Rodriguez, R., et al. (2013). miR396 affects mycorrhization and root meristem activity in the legume *Medicago truncatula*. *Plant J.* 74, 920–934. doi: 10.1111/tpj.12178

Cao, D., Wang, J., Ju, Z., Liu, Q., Li, S., Tian, H., et al. (2016). Regulations on growth and development in tomato cotyledon, flower and fruit via destruction of miR396 with short tandem target mimic. *Plant Sci.* 247, 1–12. doi: 10.1016/j.plantsci.2016.02.012

Cao, Y., Han, Y., Jin, Q., Lin, Y., and Cai, Y. (2016). Comparative Genomic Analysis of the GRF Genes in Chinese Pear (*Pyrus bretschneideri Rehd*), Poplar (*Populous*), Grape (*Vitis vinifera*), Arabidopsis and Rice (*Oryza sativa*). *Front. Plant Sci.* 7:1750. doi: 10.3389/fpls.2016.01750

Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., et al. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 33, e179–e179. doi: 10.1093/nar/gni178

Chou, M. L., Shih, M. C., Chan, M. T., Liao, S. Y., Hsu, C. T., Haung, Y. T., et al. (2013). Global transcriptome analysis and identification of a CONSTANS-like gene family in the orchid *Erycina pusilla*. *Planta* 237, 1425–1441. doi: 10.1089/dna.2014.2469

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

Duan, P., Ni, S., Wang, J., Zhang, B., Xu, R., Wang, Y., et al. (2015). Regulation of OsGRF4 by OsmiR396 controls grain size and yield in rice. *Nat. Plants* 2:15203. doi: 10.1038/nplants.2015.203

Gao, F., Wang, K., Liu, Y., Chen, Y., Chen, P., Shi, Z., et al. (2015). Blocking miR396 increases rice yield by shaping inflorescence architecture. *Nat. Plants* 2:15196. doi: 10.1038/nplants.2015.196

German, M. A., Pillay, M., Jeong, D. H., Hetawal, A., Luo, S., Janardhanan, P., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* 26, 941–946. doi: 10.1038/nbt1417

Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., et al. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. Biol. Sci.* 282:1553. doi: 10.1098/rspb.2015.1553

Gremski, K., Ditta, G., and Yanofsky, M. F. (2007). The HECATE genes regulate female reproductive tract development in *Arabidopsis thaliana*. *Development* 134, 3593–3601. doi: 10.1242/dev.011510

Hong, J. K., Suh, E. J., Lee, S.-B., Yoon, H.-J., and Lee, Y.-H. (2018). Effects of Over expression of *Brassica rapa* GROWTH-REGULATING FACTOR Genes on *B. napus* Organ Size. *Korean J. Breed Sci.* 50, 378–386.

Horiguchi, G., Kim, G. T., and Tsukaya, H. (2005). The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of *Arabidopsis thaliana*. *Plant J.* 43, 68–78. doi: 10.1111/j.1365-313X.2005.02429.x

Hou, C. J., and Yang, C. H. (2009). Functional analysis of FT and TFL1 orthologs from orchid (*Oncidium Gower Ramsey*) that regulate the vegetative to reproductive transition. *Plant Cell Physiol.* 50, 1544–1557. doi: 10.1093/pcp/pcp099

Hsu, H.-F., Hsu, W.-H., Lee, Y.-I., Mao, W.-T., Yang, J.-Y., Li, J.-Y., et al. (2015). Model for perianth formation in orchids. *Nat. Plants* 1:15046. doi: 10.1038/nplants.2015.46

Jang, S. (2015). Functional Characterization of PhapLEAFY, a FLORICAULA/LEAFY Ortholog in *Phalaenopsis aphrodite*. *Plant Cell Physiol.* 56, 2234–2247. doi: 10.1093/pcp/pcv130

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30.

Kim, J. H. (2019). Biological roles and an evolutionary sketch of the GRF-GIF transcriptional complex in plants. *BMB Rep.* 52, 227–238. doi: 10.5483/BMBRep.2019.52.4.051

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

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

Kim, J. H., and Tsukaya, H. (2015). Regulation of plant growth and development by the growth-regulating factor and grf-interacting factor duo. *J. Exp. Bot.* 66, 6093–6107. doi: 10.1093/jxb/erv349

Kim, J. S., Mizoi, J., Kidokoro, S., Maruyama, K., Nakajima, J., Nakashima, K., et al. (2012). Arabidopsis growth-regulating factor7 functions as a transcriptional repressor of abscisic acid- and osmotic stress-responsive genes, including DREB2A. *Plant Cell* 24, 3393–3405. doi: 10.1105/tpc.112.100933

Kuijt, S. J. H., Greco, R., Agalou, A., Shao, J., Hoen, C. C. J., Overnas, E., et al. (2014). Interaction between the GROWTH-REGULATING FACTOR

and KNOTTED1-LIKE HOMEOBOX Families of Transcription Factors. *Plant Physiol.* 164, 1952–1966. doi: 10.1104/pp.113.222836

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lee, S. J., Lee, B. H., Jung, J. H., Park, S. K., Song, J. T., and Kim, J. H. (2018). Growth-regulating factor and grf-interacting factor specify meristematic cells of gynoecia and anthers. *Plant Physiol.* 176, 717–729. doi: 10.1104/pp.17.00960

Li, J., Zhu, G. F., and Wang, Z. H. (2017). Chemical Variation in Essential Oil of *Cymbidium sinense* Flowers from Six Cultivars. *J. Essent Bear Pl.* 20, 385–394. doi: 10.1080/0972060X.2017.1311236

Li, S., Gao, F., Xie, K., Zeng, X., Cao, Y., Zeng, J., et al. (2016). The OsmiR396c-OsGRF4-OsGIF1 regulatory module determines grain size and yield in rice. *Plant Biotechnol. J.* 14, 2134–2146. doi: 10.1111/pbi.12569

Liang, G., He, H., Li, Y., Wang, F., and Yu, D. (2014). Molecular mechanism of microRNA396 mediating pistil development in Arabidopsis. *Plant Physiol.* 164, 249–258. doi: 10.1104/pp.113.225144

Liebsch, D., and Palatnik, J. F. (2020). MicroRNA miR396, GRF transcription factors and GIF co-regulators: a conserved plant growth regulatory module with potential for breeding and biotechnology. *Curr. Opin. Plant Biol.* 53, 31–42. doi: 10.1016/j.pbi.2019.09.008

Lin, C. S., Chen, J. J., Huang, Y. T., Hsu, C. T., Lu, H. C., Chou, M. L., et al. (2013). Catalog of *Erycina pusilla* miRNA and categorization of reproductive phase-related miRNAs and their target gene families. *Plant Mol. Biol.* 82, 193–204. doi: 10.1007/s11103-013-0055-y

Lin, Y. F., Chen, Y. Y., Hsiao, Y. Y., Shen, C. Y., Hsu, J. L., Yeh, C. M., et al. (2016). Genome-wide identification and characterization of TCP genes involved in ovule development of Phalaenopsis equestris. *J. Exp. Bot.* 67, 5051–5066. doi: 10.1093/jxb/erw273

Liu, H., Guo, S., Xu, Y., Li, C., Zhang, Z., Zhang, D., et al. (2014). OsmiR396d-Regulated OsGRFs Function in Floral Organogenesis in Rice through Binding to Their Targets OsJMJ706 and OsCR4. *Plant Physiol.* 165, 160–174. doi: 10.1104/pp.114.235564

Liu, W., Zhou, Y., Li, X., Wang, X., Dong, Y., Wang, N., et al. (2017). Tissue-Specific Regulation of Gma-miR396 Family on Coordinating Development and Low Water Availability Responses. *Front. Plant Sci.* 8:1112. doi: 10.3389/fpls.2017.01112

Liu, X. R., Pan, T., Liang, W. Q., Gao, L., Wang, X. J., Li, H. Q., et al. (2016). Overexpression of an Orchid (*Dendrobium nobile*) SOC1/TM3-Like Ortholog, DnAGL19, in Arabidopsis Regulates HOS1-FT Expression. *Front. Plant Sci.* 7:99. doi: 10.3389/fpls.2016.00099

Luo, A.-D., Liu, L., Tang, Z.-S., Bai, X.-Q., Cao, S.-Y., and Chu, C.-C. (2005). Down-Regulation of OsGRF1 Gene in Rice rhd1 Mutant Results in Reduced Heading Date. *J. Integr. Plant Biol.* 47, 745–752. doi: 10.1111/j.1744-7909.2005.00071.x

Mondragón-Palomino, M., and Theißen, G. (2008). MADS about the evolution of orchid flowers. *Trends Plant Sci.* 13, 51–59. doi: 10.1016/j.tplants.2007.11.007

Negrutiu, I., Shillito, R., Potrykus, I., Biasini, G., and Sala, F. (1987). Hybrid genes in the analysis of transformation conditions : I. Setting up a simple method for direct gene transfer in plant protoplasts. *Plant Mol. Biol.* 8, 363–373. doi: 10.1007/BF00015814

Omidbakhshfard, M. A., Proost, S., Fujikura, U., and Mueller-Roeber, B. (2015). Growth-Regulating Factors (GRFs): A Small Transcription Factor Family with Important Functions in Plant Biology. *Mol. Plant* 8, 998–1010. doi: 10.1016/j.molp.2015.01.013

Ramirez, S. R., Gravendeel, B., Singer, R. B., Marshall, C. R., and Pierce, N. E. (2007). Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* 448, 1042–1045. doi: 10.1038/nature06039

Ren, R., Gao, J., Lu, C., Wei, Y., Jin, J., Wong, S. M., et al. (2020). Highly Efficient Protoplast Isolation and Transient Expression System for Functional Characterization of Flowering Related Genes in *Cymbidium* Orchids. *Int. J. Mol. Sci.* 21:21072264. doi: 10.3390/ijms21072264

Shi, Y., Liu, H., Gao, Y., Wang, Y., Wu, M., and Xiang, Y. (2019). Genome-wide identification of growth-regulating factors in moso bamboo (*Phyllostachys edulis*): in silico and experimental analyses. *Plant Omics* 7:e7510. doi: 10.7717/peerj.7510

Shimano, S., Hibara, K. I., Furuya, T., Arimura, S. I., Tsukaya, H., and Itoh, J. I. (2018). Conserved functional control, but distinct regulation, of cell proliferation in rice and Arabidopsis leaves revealed by comparative analysis of GRF-INTERACTING FACTOR 1 orthologs. *Development* 145:159624. doi: 10.1242/dev.159624

Su, S., Shao, X., Zhu, C., Xu, J., Lu, H., Tang, Y., et al. (2018). Transcriptome-Wide Analysis Reveals the Origin of Peloria in Chinese Cymbidium (*Cymbidium sinense*). *Plant Cell Physiol.* 59, 2064–2074. doi: 10.1093/pcp/pcy130

Sun, P., Zhang, W., Wang, Y., He, Q., Shu, F., Liu, H., et al. (2016). OsGRF4 controls grain shape, panicle length and seed shattering in rice. *J. Integr. Plant Biol.* 58, 836–847. doi: 10.1111/jipb.12473

Tang, Y., Liu, H., Guo, S., Wang, B., Li, Z., Chong, K., et al. (2018). OsmiR396d Affects Gibberellin and Brassinosteroid Signaling to Regulate Plant Architecture in Rice. *Plant Physiol.* 176, 946–959. doi: 10.1104/pp.17.00964

Wang, F., Qiu, N., Ding, Q., Li, J., Zhang, Y., Li, H., et al. (2014). Genome-wide identification and analysis of the growth-regulating factor family in Chinese cabbage (Brassica rapa L. ssp. pekinensis). *BMC Genomics* 15:807. doi: 10.1186/1471-2164-15-807

Wang, H. H., Wonkka, C. L., Treglia, M. L., Grant, W. E., Smeins, F. E., and Rogers, W. E. (2015). Species distribution modelling for conservation of an endangered endemic orchid. *AoB Plants* 7:39. doi: 10.1093/aobpla/plv039

Wang, J., Zhang, C., Yan, Y., Wu, W., and Ma, Z. (2013). Identification of conserved MicroRNAs and their targets in Phalaenopsis orchid. *Russ. J. Plant Physiol.* 60, 845–854. doi: 10.1134/S1021443713060150

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

Xiang, L., Chen, Y., Chen, L., Fu, X., Zhao, K., Zhang, J., et al. (2018). B and E MADS-box genes determine the perianth formation in *Cymbidium goeringii* Rchb.f. *Physiol. Plant* 162, 353–369. doi: 10.1111/ppl.12647

Yang, F., Gang, L., Liu, D., Yu, and Di. (2009). Arabidopsis MiR396 Mediates the Development of Leaves and Flowers in Transgenic Tobacco. *J. Plant Biol.* 52, 475–481. doi: 10.1007/s12374-009-9061-7

Yang, F., and Zhu, G. (2015). Digital Gene Expression Analysis Based on De Novo Transcriptome Assembly Reveals New Genes Associated with Floral Organ Differentiation of the Orchid Plant *Cymbidium ensifolium*. *PLoS One* 10:e0142434. doi: 10.1371/journal.pone.0142434

Yang, F., Zhu, G., Wang, Z., Liu, H., Xu, Q., Huang, D., et al. (2017). Integrated mRNA and microRNA transcriptome variations in the multi-tepal mutant provide insights into the floral patterning of the orchid *Cymbidium goeringii*. *BMC Genomics* 18:367. doi: 10.1186/s12864-017-3756-9

Yang, Z., Yang, D., Ding, X., Gao, Y., Li, D., and Xu, T. (2015). MicroRNA expression profiles in conventional and micropropagated *Dendrobium officinale*. *Genes Genomics* 37, 315–325. doi: 10.1007/s13258-014-0257-y

Zhang, D., Sun, W., Singh, R., Zheng, Y., Cao, Z., Li, M., et al. (2018). GRF-interacting factor1 Regulates Shoot Architecture and Meristem Determinacy in Maize. *Plant Cell* 30, 360–374. doi: 10.1105/tpc.17.00791

Zhang, Z., Yan, Y., Tian, Y., Li, J., He, J.-S., and Tang, Z. (2015). Distribution and conservation of orchid species richness in China. *Biolog. Conserv.* 181, 64–72. doi: 10.1016/j.biocon.2014.10.026

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.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 March 2022
doi: 10.3389/fpls.2022.860997

Anthocyanin and Flavonol Glycoside Metabolic Pathways Underpin Floral Color Mimicry and Contrast in a Sexually Deceptive Orchid

Darren C. J. Wong, James Perkins and Rod Peakall*

Ecology and Evolution, Research School of Biology, Australian National University, Canberra, ACT, Australia

Sexually deceptive plants secure pollination by luring specific male insects as pollinators using a combination of olfactory, visual, and morphological mimicry. Flower color is a key component to this attraction, but its chemical and genetic basis remains poorly understood. *Chiloglottis trapeziformis* is a sexually deceptive orchid which has predominantly dull green-red flowers except for the central black callus projecting from the labellum lamina. The callus mimics the female of the pollinator and the stark color contrast between the black callus and dull green or red lamina is thought to enhance the visibility of the mimic. The goal of this study was to investigate the chemical composition and genetic regulation of temporal and spatial color patterns leading to visual mimicry, by integrating targeted metabolite profiling and transcriptomic analysis. Even at the very young bud stage, high levels of anthocyanins were detected in the dark callus, with peak accumulation by the mature bud stage. In contrast, anthocyanin levels in the lamina peaked as the buds opened and became reddish-green. Coordinated upregulation of multiple genes, including dihydroflavonol reductase and leucoanthocyanidin dioxygenase, and the downregulation of flavonol synthase genes (*FLS*) in the callus at the very young bud stage underpins the initial high anthocyanin levels. Conversely, within the lamina, upregulated *FLS* genes promote flavonol glycoside over anthocyanin production, with the downstream upregulation of flavonoid O-methyltransferase genes further contributing to the accumulation of methylated flavonol glycosides, whose levels peaked in the mature bud stage. Finally, the peak anthocyanin content of the reddish-green lamina of the open flower is underpinned by small increases in gene expression levels and/or differential upregulation in the lamina in select anthocyanin genes while *FLS* patterns showed little change. Differential expression of candidate genes involved in specific transport, vacuolar acidification, and photosynthetic pathways may also assist in maintaining the distinct callus and contrasting lamina color from the earliest bud stage through to the mature flower. Our findings highlight that flower color in this sexually deceptive orchid is achieved by complex tissue-specific coordinated regulation of genes and biochemical pathways across multiple developmental stages.

Keywords: *Chiloglottis*, anthocyanin, flavonol glycoside, flower, orchids, transcriptome, sexual deception, mimicry

OPEN ACCESS

Edited by:
Jen-Tsung Chen, National University of Kaohsiung, Taiwan

Reviewed by:
Carlos Martel, Royal Botanic Gardens, Kew, United Kingdom
Surendra Sarsaiya, Zunyi Medical University, China
Roman Tobias Kellenberger, University of Cambridge, United Kingdom

****Correspondence:***
Darren C. J. Wong
darren.wong@anu.edu.au; wongdcj@gmail.com

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: *24 January 2022*
Accepted: *17 February 2022*
Published: *23 March 2022*

Citation:
Wong DCJ, Perkins J and Peakall R (2022) Anthocyanin and Flavonol Glycoside Metabolic Pathways Underpin Floral Color Mimicry and Contrast in a Sexually Deceptive Orchid. Front. Plant Sci. 13:860997. doi: 10.3389/fpls.2022.860997

INTRODUCTION

The majority of angiosperms are pollinated by animals (Ollerton et al., 2011), with floral signaling key to this crucial interaction. Flowers not only contain the reproductive structures, but also convey long-range visual and olfactory signals for pollinator attraction and the simultaneous deterrence of unwanted visitors (Raguso, 2004; Borghi et al., 2017). Brightly colored flowers are the norm in diurnally pollinated plants, ensuring visibility to pollinators against the background, with flower colors spanning the UV and visible spectrum. In addition to bright colors, many flowers display contrasting color markings, spots, and lines, which serve as "nectar guides" for pollinators, with some markings only visible to pollinators that can perceive UV light (Davies et al., 2012; Glover et al., 2013; Nadot and Carrive, 2021).

A diverse range of pigment compounds are responsible for the various colors and markings of flowers. Anthocyanins typically confer the orange, red, violet, and blue colors to flowers, while chalcones, aurones, flavonols, and flavones are often associated with different degrees of pale yellow coloration. Chlorophylls and carotenoids can further extend the color palette to include green and yellow-to-red colors (Grotewold, 2006; Tanaka and Tanaka, 2006; Tanaka et al., 2008). Importantly, the color of flowers is determined not only by the specific pigment compounds present, but also by their interactions with cellular pH, metal ions, and other co-pigments (Tanaka et al., 2008).

The anthocyanin and flavonol glycoside pathway is highly conserved across the flowering plants (Grotewold, 2006; Tanaka et al., 2008; Davies et al., 2012), and thus provides a helpful framework for any new investigation of the molecular basis of flower color. Dramatic floral color changes can be mediated by simple genetic changes affecting the regulation, expression, or function of just a few genes, or via more complex coordinated changes over many genes within the anthocyanin and flavonol pathway (Wessinger and Rausher, 2012; Tanaka and Brugliera, 2013; Zhao and Tao, 2015). Floral color changes can also be due to changes in the downstream genes that modify anthocyanins and flavonol glycosides (Morita et al., 2015). Even further downstream, changes in transport genes (Cheng et al., 2015; Zhao et al., 2020; Lu et al., 2021), and genes associated with the pH of vacuole storage (Fukada-Tanaka et al., 2000; Faraco et al., 2014) can also impact flower color. A case for non-catalytic proteins that can influence anthocyanin production, and so alter flower colors, is also known (Morita et al., 2014).

It is also now well-established that a highly conserved protein complex consisting of MYB–basic-helix-loop-helix (bHLH)–WD40 repeat (WDR) proteins (MBW) are key regulatory determinants of floral pigmentation and patterning in numerous plants (Wong et al., 2016; Allan and Espley, 2018; LaFountain and Yuan, 2021; Rodrigues et al., 2021). However, the precise components of this complex and its transcriptional control are poorly understood in non-model plant species (Davies et al., 2012; Wessinger and Rausher, 2012). Recent work has nonetheless highlighted the complexity of the anthocyanin gene regulatory network, which integrates distinct hierarchical, feedback, and repressor activities, as well as intercellular mobility of some MBW components that execute various developmental stage, tissue-specific, and stress-responsive pigmentation programs (Albert et al., 2014). Due to these complexities, it is impossible to accurately predict the pigments present and the mechanisms involved in pigmentation of a flower based on its color alone. Therefore, chemical and molecular investigations are necessary, particularly for unusually colored flowers.

Uniformly black flowers are very rare in nature, although human interest has resulted in the development of some black-flowered ornamental cultivars of dahlias, lilies, pansies, and tulips (Davies, 2008). Due to the relative rarity of black flowers, the chemical and genetic basis of black coloration in flowers remains poorly understood (Grotewold, 2006; Davies, 2008; Tanaka et al., 2008; Davies et al., 2012; Iwashina, 2015). However, it is clear that an overaccumulation of anthocyanins is largely responsible for the black colors of flowers (Markham et al., 2004; Thomas et al., 2009; Thill et al., 2012; Kellenberger et al., 2019), leaves (Hatier and Gould, 2007; Zheng et al., 2019), grains (Abdel-Aal et al., 2006; Yang et al., 2019), and fruits (Boss et al., 1996; Fan-Chiang and Wrolstad, 2006; Liu et al., 2018).

Examples of naturally black flowers include *Lisianthius nigrescens* where the uniformly dark colored petals contain extraordinary high levels of delphinidin-based anthocyanins (Markham et al., 2004). In the alpine orchid *Gymnadenia rhellicani*, higher amounts of cyanidin-based anthocyanins were observed in the petals of the black compared to red or white floral morphs (Kellenberger et al., 2019). This high anthocyanin content is linked to differential expression of an anthocyanidin synthase gene, *GrANS1* brought by loss-of-function mutation of its regulator, *GrMYB1* (Kellenberger et al., 2019). Local tissue-specific upregulation is also involved in the formation of specific dark markings and spots that generate strong contrasting colors on the flower (Martins et al., 2013, 2017; Li et al., 2014; Zhang et al., 2015). For example, in several subspecies of *Clarkia gracilis* (e.g., the basal-spotted ssp. *albicaulis* and the central-spotted ssp. *sonomensis*), the appearance of reddish-purple petal spots containing cyanidin and peonidin pigments is largely the result of spatially restricted dihydroflavonol reductase (*DFR*) expression in the early stages of flower development (Martins et al., 2013, 2017).

Competition between flavonol glycoside and anthocyanin biosynthesis is predicted to be another common mechanism underpinning variation in the spatial pattern of coloration, although the details remain poorly understood (Sheehan et al., 2016; Yuan et al., 2016). Flavonol synthase (FLS) and DFR use the same dihydroflavanol substrates, but produce flavonol and anthocyanin biosynthetic products, respectively (Tanaka and Brugliera, 2013). Flavonols are uncolored or only faintly colored compared to the vibrant colors of anthocyanins, so tissue- or region-specific coloration can be mediated by the differential expression of these two genes. For example, the absence of anthocyanins in the white areas surrounding the otherwise pink corolla of bumblebee-pollinated *Mimulus lewisii* flowers is due to high *FLS* expression, while low *FLS* expression abolishes this spatial patterning in red hummingbird-pollinated *Mimulus cardinalis* flowers (Yuan et al., 2016). Another study showed that constitutive overexpression of three distinct *FLS* from various plant species (*Rosa rugosa*, *Prunus persica*, and *Petunia*

hybrida) in tobacco resulted in white flowers compared to pale pink flowers in control plants (Luo et al., 2016). Conversely, constitutive overexpression of *DFR* genes in transgenic tobacco plants led to increased anthocyanin accumulation and a deep red flower phenotype.

Sexually deceptive plants secure pollination by luring specific male insects as pollinators using a combination of olfactory, visual, and morphological mimicry (Bohman et al., 2016). Within the orchids, pollination by sexual deception has evolved independently on four continents and is employed by hundreds of species spanning multiple lineages of the Orchidaceae (Peakall et al., 2020). The types of pollinators involved are also diverse, and include ants, bees and wasps (Hymenoptera), fungus gnats and other flies (Diptera), and beetles (Coleoptera) (Gaskett, 2011; Bohman et al., 2016; Peakall et al., 2020; Cohen et al., 2021; Hayashi et al., 2021). In these cases, pollination is achieved during attempted copulation with the labellum, an often highly modified petal of the orchids. Furthermore, in the multiple cases now chemically characterized, sex pheromone mimicry, often involving unusual compounds or unusual blends, is critical for the long range attraction of the pollinator (Bohman et al., 2016; Wong et al., 2017b; Peakall et al., 2020). Rather than being brightly colored, sexually deceptive orchid flowers are most often dull red and green in color, with the labellum often exhibiting starkly contrasting dark structures and markings which are thought to be visual cues that improve the insect mimicry (**Figure 1**, see also photos in Bohman et al., 2016; Johnson and Schiestl, 2016; Peakall et al., 2020). Outside of the Orchidaceae, there are two other known cases of pollination by sexual deception: *Gorteria diffusa* (**Figure 1G**) in the Asteraceae (Ellis and Johnson, 2010) and *Iris paradoxa* (**Figure 1H**) in the Iridaceae (Vereecken et al., 2012) both characterized by flowers that include some very dark pigmentation.

Beyond the potential role of female visual mimicry, where a close match to the color of the female of the pollinator is expected (de Jager and Peakall, 2016), the tissue-specific pigmentation patterns of sexually deceptive flowers may also provide strong color contrasts that improve detectability. For example, *Drakaea livida* shares the same male thynnine wasp pollinator as the unrelated, morphologically distinct, and differently colored *Caladenia decora*, *C. pectinata*, and *C. procera*. Yet, across this set of species, tissue-specific pigmentation provides strong achromatic contrast against a chromatic background, thus, potentially enhancing the detectability by the pollinator (Gaskett et al., 2017).

The very dark, three dimensional "calli" on the dull green and red labellum lamina of sexually deceptive *Chiloglottis* orchids are crucial to their sexual mimicry (**Figure 1**). In the most well-studied case of *C. trapeziformis*, the central black callus structure is the source of the unique volatile compound chiloglottone 1, which is the sex pheromone of their specific thynnine wasp pollinator, *Neozeleboria cryptoides* (Schiestl et al., 2003; Franke et al., 2009). The structure also serves as a visual and tactile mimic of the female, providing a gripping point for male wasps as they attempt to copulate with the tip of the labellum (de Jager and Peakall, 2016). Furthermore, the distance from callus to labellum tip is a close match to the length of the female wasp, and experimentally shortened or elongated labella result in significantly reduced durations of attempted copulation, indicating the importance of morphological mimicry

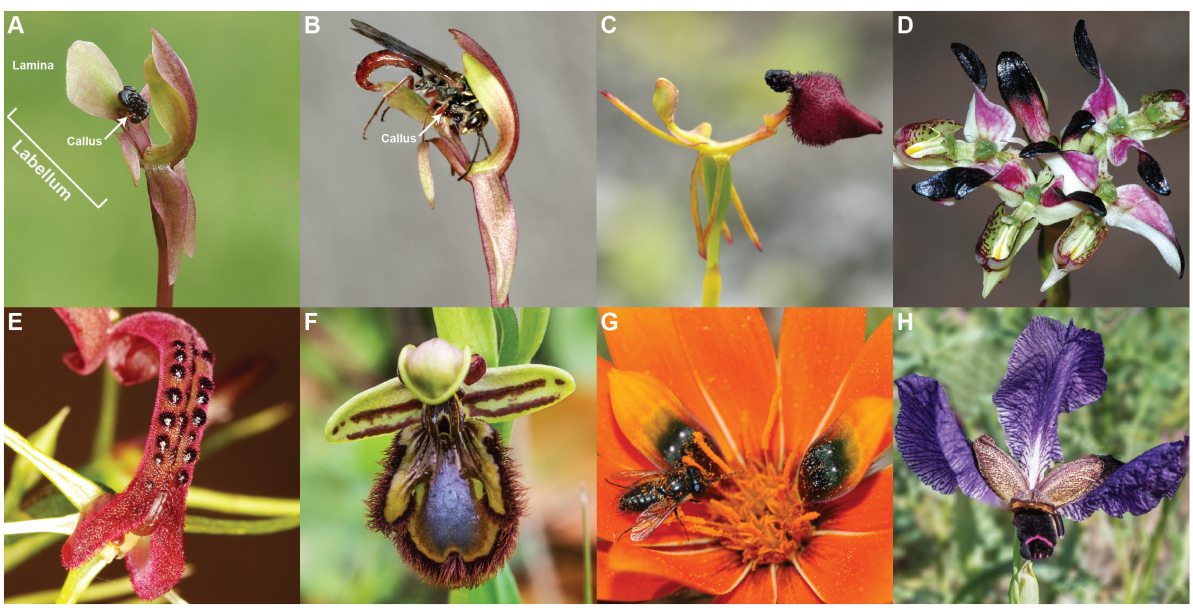


FIGURE 1 | An illustration of the diversity of sexually deceptive orchid and non-orchid species. Orchids: **(A)** The floral structure of *Chiloglottis trapeziformis* with **(B)** its male pollinator *Neozeleboria cryptoides*. The arrows point to the very dark, three dimensional "callus" structure on the dull green and red labellum lamina of *Chiloglottis trapeziformis*, **(C)** *Drakaea glyptodon*, **(D)** *Disa atricapilla*, **(E)** *Cryptostylis leptochila*, and **(F)** *Ophrys speculum*. Non-orchids: **(G)** *Gorteria diffusa*, and **(H)** *Iris paradoxa*. All images have been reproduced with permission from the respective copyright holders. Please refer to the Section "Acknowledgments" for image credits.

(de Jager and Peakall, 2016, 2019). Finally, spectral reflectance measurements revealed strong similarity between the perceived color of the black callus and the female wasp abdomen in hymenopteran visual models. Both female and orchid structures are predicted to be perceived as uncolored or achromatic to male wasps, consistent with the hypothesis that the dark callus structure acts as a visual mimic of the female (de Jager and Peakall, 2016). Conversely, the surrounding lamina of the labellum is chromatic, indicating morphological adaptations to produce within-flower chromatic/achromatic contrast (**Figure 1**).

As the first step toward understanding the chemical and molecular basis of floral color adaptations in *Chiloglottis trapeziformis*, we leveraged targeted metabolite and transcriptome analysis of a flower tissue and developmental series to ask the following questions: (1) What anthocyanins and flavonol glycoside co-pigments are present in the flower? (2) What are the spatiotemporal patterns of anthocyanin and flavonol glycoside levels across flower development and tissue types? (3) Is the unique spatiotemporal distribution of color in the flower explained by differential expression of a few anthocyanin and flavonol glycoside pathway-related genes at specific developmental stages, or coordinated regulation at many genes and multiple stages? We propose possible mechanisms for the tissue-specific pigmentation and its regulation and discuss the evolutionary implications of these findings more broadly for *Chiloglottis* and other sexually deceptive plants.

MATERIALS AND METHODS

Sample Collection

This study builds on a larger effort to understand the biochemistry, biosynthesis and regulation of chiloglottones, the unique semiochemicals used by sexually deceptive *Chiloglottis* orchids (Falara et al., 2013; Amarasinghe et al., 2015; Wong et al., 2017a, 2018, 2019). Here, we turn attention to the chemistry and biosynthesis of flower color in *Chiloglottis trapeziformis*. Flowers at five developmental stages (very young buds, *vyb*; young bud, *yb*; mature buds, *mb*; very mature bud, *vmb*; and mature flowers open to the sun, or sun flowers, *sflw*) were sampled from the Australian National Botanic Gardens (Canberra, ACT, Australia) in September 2014. All tissues were immediately snap-frozen in liquid nitrogen at the point of collection and stored at −80°C until further use. Dissection of floral tissues into callus and labellum remains were performed in liquid N_2.

The classifications of bud and flower stages follows Amarasinghe et al. (2015). Briefly, the *vyb* is characterized by a small and very tightly closed green bud. However, it is noteworthy that dissection of the buds reveals that the callus is already very dark in color at the *vyb* stage. The *vmb* stage is characterized by larger green buds that are about to open, with sepals and petals beginning to separate. By this stage the callus is black, and remains so for the duration of flowering. At the onset of flower opening, the lamina is green. However, within 2 to 3 days of exposure to sunlight, the lamina becomes reddish-green in color. It was at this stage that the *sflw* was sampled (**Figure 2** inset).

Ultra High-Performance Liquid Chromatography Mass Spectrophotometric Analysis

For the targeted metabolite analysis, a total of 54 separate labellum samples were used, encompassing replication for all of the five developmental stages (i.e., *vyb*, *yb*, *mb*, *vmb*, *sflw*). Each sample consisted of either the dissected callus or lamina (labellum remains after the stalked callus was removed) from several individual flowers. Pre-weighed samples were homogenized and extracted in 400 μL of solvent composed of 70:30:1 methanol:water:acetic acid with agitation at 4°C for 24 h. Samples were centrifuged at 13,000 rpm for 10 min and the supernatant was filtered with 0.2 μm PTFE SINGLE StEP filter vials (Thomson) and analyzed using a Thermo Q Extractive Plus UPLC-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA, United States). Samples and standards (5 μL injection volumes) were separated chromatographically on an Agilent reversed-phase Zorbax Eclipse XDB-C18 column (2.1 × 50 mm, 1.8 μm particles) held at 40°C. The mobile phases used were water with 0.1% HPLC grade formic acid (solvent A), and methanol with 0.1% formic acid (solvent B). Samples and standards were eluted with a constant flow rate of 400 μL min^{-1}, with a 25-min gradient program as follows: 0–1.5 min, 6% B; 1.5–2 min, 6–10% B; 2–14 min, 10–60% B; 14–15 min, 60–90% B; 15–19 min, 90% B; 19–20 min, 6% B; 20–25 min, 6% B. Eluted compounds were introduced to the MS via a HESI-II probe (Thermo Fisher Scientific, Waltham, MA, United States). Representative samples were also analyzed on a separate LC-MS instrument fitted with a diode array detector (DAD) operating at 520 nm and at 365 nm and using the same chromatographic parameters, to distinguish anthocyanins from flavonol glycosides.

For putative identification and relative quantification of anthocyanins and flavonols, the HESI was operated in the positive mode. Mass spectra were acquired using full MS and data-dependent MS/MS acquisition (DDA) modes at a scan range of 100 to 1500 m/z. Additionally, representative samples of both tissue types at various developmental stages were analyzed in negative ionization mode. In negative ionization mode, anthocyanins produce a distinctive $[M\text{-}2H + H_2O]^-$ ion in addition to the $[M\text{-}2H]^-$ ion, while isomeric flavonol glycosides do not (Sun et al., 2012a). Thus, the presence of this ion is diagnostic of anthocyanins and can be used to distinguish them from isomeric flavonols with similar fragmentation patterns.

Data were acquired using Thermo Scientific XCALIBUR 4.0 and analyzed using Thermo Scientific FreeStyle software. Anthocyanins and flavonols were distinguished based on (i) accurate masses (4 decimal places) of molecular ions in full MS spectra in both ionization modes, (ii) comparisons of fragmentation patterns in the MS/MS spectra in both ionization modes with those available in online databases such as RIKEN tandem mass spectral database (Sawada et al., 2012) and MassBank of North America[1], (iii) the presence or absence of the diagnostic $[M\text{-}2H + H_2O]^-$ ion in the negative ionization analyses, and (iv) absorbance spectra at 520 and 365 nm. Putative

[1]https://mona.fiehnlab.ucdavis.edu/

anthocyanins and flavonols were quantified by integration of the molecular ion peak in full MS spectra and calculated from linear calibration curves of cyanidin 3-O-glucoside chloride (Sigma-Aldrich) and 7-hydroxycoumarin (Sigma-Aldrich), respectively.

An analysis of variance (ANOVA) and *post hoc* Tukey–Kramer Honestly Significant Difference test were performed to evaluate the differences in total anthocyanin and flavonol glycoside content in the callus and labellum during flower development, following the confirmation of normality by a Shapiro-Wilk test. Statistical analyses were conducted using R[2].

De novo Transcriptome Assembly

RNA extraction, library construction, and RNA sequencing were performed as previously described (Wong et al., 2017a, 2018). However, the assembly and downstream analysis of the *sflw* stage has not been previously reported. Therefore, paired-end reads from the callus and lamina tissues of *vyb*, *vmb* (Wong et al., 2017a, 2018) and *sflw* (used for the first time in this study) stages were first pooled. Next, removal of the adaptor, sliding-window trimming, length (l = 40) and quality filtering, and base correction (*-c*) of the raw PE reads were performed with *fastp* v0.20.0 (Chen et al., 2018) using default settings unless otherwise specified. *De novo* transcriptome construction based on pooled tissue and developmental stages was performed using *Trinity* v2.11.0 (Haas et al., 2013) with default settings except the minimum contig length and *k*-mer size were set to 300 and 31, respectively. Protein-coding prediction was achieved using TransDecoder v5.5.0[3]. To maximize the prediction sensitivity, homology searches against the UniProt Reference Clusters (UniRef90) and protein families (Pfam) databases were performed using DIAMOND (Buchfink et al., 2014) and HMMER (Mistry et al., 2013), and incorporated into the TransDecoder pipeline. Resolution of assembly redundancy, protein-coding region prediction, and identification of accurate gene sets was made using the EvidentialGene tr2aacds4 pipeline (Gilbert, 2013) using default settings.

Transcriptome Analysis of Anthocyanin and Flavonol Glycoside Pathway-Related Genes

Alignment of filtered PE reads toward the assembled transcriptome of *C. trapeziformis* were individually performed with bowtie2 (Langmead and Salzberg, 2012) using the local read alignment mode (*–local*). Read count matrices were obtained using FeatureCounts (Liao et al., 2014) with default parameters except for the *-B* (both ends must be aligned) and *-C* (exclude chimeric fragments) option enabled for each species. Transcripts having sufficiently large counts were retained for further downstream statistical analysis using edgeR with the *filterByExpr* option (Robinson et al., 2009). Differential expression (DE) analysis between groups/treatments of interests were performed using DESeq2 (Love et al., 2014). An absolute $\log_2$ fold change ($|\log2FC| > 0.5$ and a false discovery rate (FDR) threshold < 0.05 determines transcripts that are differentially expressed in each contrast. Transcript expression was expressed as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). The assembled transcriptome was annotated with MapMan categories using Mercator (Lohse et al., 2014). Additional Pfam domains, putative homolog annotations, and transcripts sharing homology and functional domains to known and/or emerging flavonoid pathway components (Grotewold, 2006; Wessinger and Rausher, 2012; Tohge et al., 2017), were obtained via the TransDecoder's prediction pipeline. Hierarchical clustering of expression and DE patterns of the callus and lamina across multiple developmental stages were performed using *hclust* and plotted using the *ggplot2* package in R (see text footnote 2). A summary of developmental stage- and tissue-specific transcriptome analysis of *Chiloglottis trapeziformis* flowers is described in **Supplementary Figure 1** and corresponding gene expression dataset and differential expression result tables are available in **Supplementary Datas 1–3**. New sequence reads obtained for the *sflw* stage has been added to the existing BioProject accession PRJNA390683 and SRA study accession SRP1093281.

RESULTS

The Callus and Labellum Contain a Diverse Array of Anthocyanin and Flavonol Glycosides

A total of five anthocyanins and 14 flavonol glycosides were detected in *Chiloglottis trapeziformis* labella by UHPLC-MS/MS and LC-DAD-MS analysis (**Table 1** and **Supplementary Figure 2**). Cyanidin-based anthocyanins were the dominant pigment, with peonidin- and pelargonidin-based anthocyanins also detected. Kaempferol (K), Quercetin (Q), and their methylated derivatives (Kaempferide, Kde; Isorhamnetin, IR) were the dominant flavonol glycosides. No delphinidin-based anthocyanins, myricetin flavonol glycosides nor common acyl modifications aside from malonyl groups (e.g., acetylation, coumarylation, caffeoylation) were detected (**Table 1**).

Candidate Anthocyanin and Flavonol Glycoside Pathway-Related Genes

Our transcriptome analysis revealed a set of candidate gene homologous to well-known anthocyanin and flavonol glycoside pathway genes (Grotewold, 2006; Tanaka et al., 2008; Davies et al., 2012). This included three chalcone synthase (*CtrCHS1–3*), one chalcone isomerase (*CtrCHI*), two flavone 3-hydroxylase (*CtrF3H1–2*), and four flavonoid 3'-hydroxylase (*CtrF3'H1–4*) encoding enzymes involved in the early biosynthetic steps of general flavonoid biosynthesis. Notably, F3'H catalyzes hydroxylation at the 3'-position of naringenin and dihydrokaempferol to yield eriodictyol and dihydroquercetin, and also converts kaempferol to quercetin (Tohge et al., 2017). Three dedicated anthocyanidin biosynthesis genes (one *CtrDFR1* and two leucoanthocyanidin dioxygenase syn. anthocyanidin synthase) (*CtrLDOX1–2*) and two flavonol glycoside biosynthesis genes (e.g., flavonol synthase, *CtrFLS1–2*) were also identified.

[2]https://www.r-project.org/

[3]https://transdecoder.github.io

TABLE 1 | Putative anthocyanins and flavonols in floral extracts of *Chiloglottis trapeziformis* calli and labellum lamina tissues.

Putative anthocyanin/flavonol glycoside	Putative aglycone	Peak RT (min)	[M]+	[M]+ MSMS transitions (relative intensity)	[M-2H]-, [M-2H + H2O]-
Anthocyanin					
Cyanidin glucoside (CG)	Cyanidin	7.12	449.1076	**287.0547 (100),** 288.0580 (8.5), 71.7628 (0.9), 286.9883 (0.7), 69.7407 (0.3)	447.0931, 465.1034
Cyanidin malonylglucoside (CmG)	Cyanidin	8.93	535.1086	**287.0546 (100),** 288.0579 (7.4), 71.7628 (0.9), 286.9883 (0.7), 258.0482 (0.06)	533.0936, 551.1042
Pelargonidin glucoside (PelG)	Pelargonidin	7.54	433.1131	**271.0596 (100),** 272.0630 (10.3), 67.7641 (1.0), 270.9985 (0.6), 69.7845 (0.4)	431.0979, 449.1090
Peonidin glucoside (PeoG)	Peonidin	8.03	463.1233	**301.0702 (100),** 302.0735 (10.5), 75.2666 (0.9), 300.9990 (0.7), 286.0462 (0.4)	461.1088, 479.1197
Peonidin malonylglucoside (PeomG)	Peonidin	9.76	549.1234	**301.0701 (100),** 302.0734 (8.5), 75.2666 (0.9), 300.9991 (0.7), 286.0456 (0.4)	547.1096, 565.1200
Flavonol glycoside					
Quercetin diglucoside (QGG)	Quercetin	8.91	627.1541	**303.0492 (100),** 85.0287 (6.8), 304.0526 (5.9), 127.0388 (2.0), 97.0285 (1.8), **465.1016 (1.0)**	
Kaempferol diglucoside (KGG)	Kaempferol	8.95	611.1599	**287.0545 (100**), 85.0287 (10.4), 288.0579 (7.1), 97.0286 (2.8), 127.0388 (2.2), **449.1056 (2.0)**	
Quercetin malonyl diglucoside (QmGG)	Quercetin	9.56	713.1547	**303.0493 (100),** 127.0389 (8.7), 85.0287 (8.4), 304.0527 (8.2), 109.0285 (5.3), **465.1026 (2.4)**	
Quercetin glucoside (QG)	Quercetin	10.39	465.1024	**303.0494 (100),** 85.0287 (10.7), 304.0527 (9.5), 97.0286 (2.8), 127.0389 (2.5),	
Unknown Quercetin pentose (QP)	Quercetin	10.48	435.0925	**303.0496 (100),** 73.02881 (15.7), 304.0534 (10.1), 195.0285 (4.6), 57.0341 (4.8)	
Quercetin malonylglucoside (QmG)	Quercetin	10.83	551.1034	**303.0496 (100),** 127.0390 (11.7), 85.0288 (11.6), 109.0286 (9.1), 304.0529 (7.8)	
Kaempferol glucoside (KG)	Kaempferol	11.43	449.1076	**287.0546 (100),** 85.0288 (9.3), 288.0580 (6.9), 97.0286 (3.0), 127.0389 (2.5)	
Isorhamnetin diglucoside (IRGG)	Isorhamnetin	11.81	641.1708	**317.0651 (100),** 318.0686 (6.2), 85.0287 (4.4), 97.0286 (1.7), 79.2659 (1.6), **479.1159 (1.2)**	
Kaempferide diglucoside (KdeGG)	Kaempferide	11.84	625.1761	**301.0701 (100),** 302.0735 (7.6), 85.0287 (7.2), **463.1230 (2.6),** 97.0285 (2.0)	
Kaempferol malonylglucoside (KmG)	Kaempferol	11.86	535.1075	**287.0546 (100),** 127.0390 (11.0), 85.0287 (9.5), 109.0285 (6.9), 159.0285 (4.5)	
Isorhamnetin malonyl diglucoside (IRmGG)	Isorhamnetin	12.21	727.1708	**317.0649 (100**), 318.0683 (9.3), 85.0287 (6.6), 127.0389 (6.0), 109.0284 (4.1), **479.1175 (2.8)**	
Kaempferide malonyl diglucoside (KdemGG)	Kaempferide	12.32	711.1749	**301.0704 (100**), 85.0288 (8.2), 302.0736 (7.7), 127.0390 (7.6), 109.0285 (5.4), **463.1222 (4.6)**	
Isorhamnetin glucoside (IRG)	Isorhamnetin	13.89	479.1182	**317.0650 (100)**, 318.0684 (10.2), 79.2653 (0.9), 316.9880 (0.8), 85.0287 (0.7)	
Isorhamnetin malonylglucoside (IRmG)	Isorhamnetin	14.34	565.1182	**317.0649 (100)**, 127.0389 (10.4), 85.0287 (8.8), 318.0684 (7.6), 109.0285 (6.7)	

The positions of sugar and acyl group attachment to the anthocyanidins or flavonol aglycones were not determined, and glucosides cannot be distinguished from isomeric galactosides using our methodology. Especially informative MSMS transitions are listed in bold.

The discovery of the candidate genes downstream of F3'H is in accord with the anthocyanins (two cyanidin and two peonidin anthocyanins) and flavonol glycosides (five Q and four IR flavonol glycosides) we detected. One pelargonidin anthocyanin, three K and three Kde flavonol glycosides which do not depend on F3'H activity were also found (**Figure 2** and **Table 1**). Many of these genes also fall within the specific BIN9.2.2 category of secondary metabolism.phenolics.flavonoid biosynthesis (e.g., *CtrCHS* in BIN9.2.2.1, *CtrCHI* in BIN9.2.2.2, *CtrF3H* in BIN9.2.2.4, and *CtrDFR* and *CtrLDOX* in BIN9.2.2.9) or were reciprocal best BLAST hits of several plant CYP75B (i.e., *CtrF3'H*) enzymes (**Supplementary Data 1**).

Unlike the conserved early steps in the pathway, the downstream modifications of flavonoids with various glycosyl, methyl, and/or acyl groups are achieved by highly diverse and often family, genera-, and/or even species-specific sets of genes

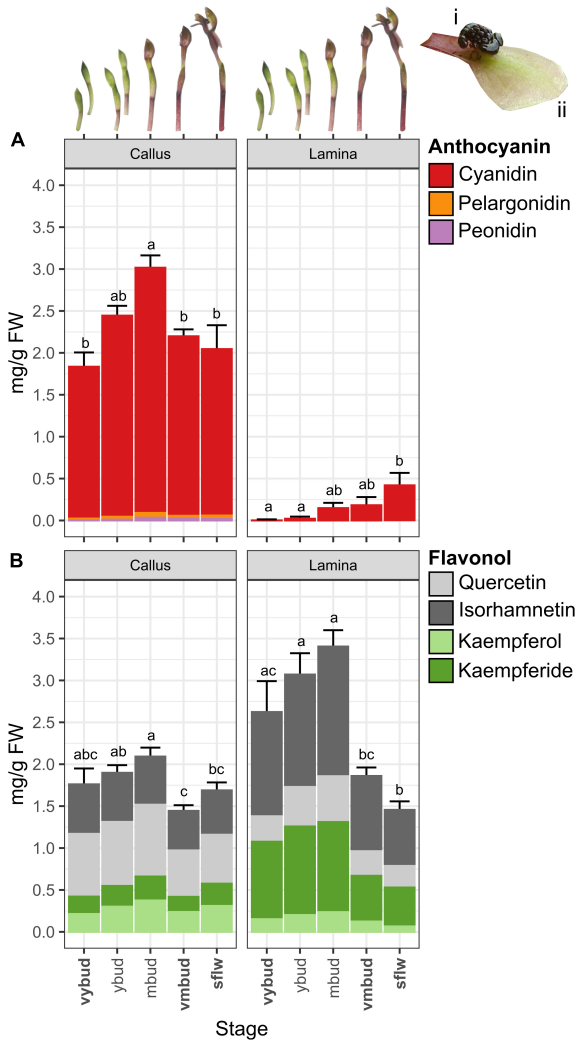


FIGURE 2 | Changes in anthocyanins and flavonol glycoside co-pigments of *Chiloglottis trapeziformis* flowers. Bar graphs depict the total **(A)** anthocyanin and **(B)** flavonol glycoside content (average ± s.e.) in the callus and labellum lamina (see insets i and ii) of *Chiloglottis trapeziformis* flowers at different developmental stages (i.e., very young bud, *vyb*; young bud, *yb*; mature buds, *mb*; very mature bud, *vmb*; and naturally opened flowers in the field, *sflw*). Bold indicates developmental stages with corresponding transcriptome profiled. Labels not connected by the same letter are significantly different at $P < 0.05$ based on ANOVA and Tukey HSD test (see **Supplementary Figure 7** for further information).

(Tanaka et al., 2008). Nonetheless, typically one or more flavonoid glycosyltransferases, methyltransferases, or acyltransferases catalyze these reactions. Three glycosyltransferase homologs potentially relevant to flavonoid metabolism were identified: two transcripts encoding Arabidopsis UDP-glucose:flavonoid 3-O-glucosyltransferase homolog (*CtrUGT78D2a/b*) potentially involved in the glycosylation of anthocyanidins and flavonols at the 3-position (Lim et al., 2004; Tohge et al., 2005), and one encoding an Arabidopsis UGT71B1 (Lim et al., 2004) homolog with specific flavonol 3-O-glucosyltransferase activity (*CtrUGT71B1*).

O-methyltransferases with both broad and narrow flavonoid substrate specificities have been identified in many plants. For example, the grape anthocyanin *O*-methyltransferase VviAOMT that catalyzes the methylation of anthocyanins is also active on flavonol glycosides compounds (Hugueney et al., 2009; Lücker et al., 2010; Provenzano et al., 2014). Here, four transcripts sharing homology toward VviAOMT (*CtrCOMT1–4*) were identified. Many plant flavonoid malonyltransferases characterized to date can catalyze the malonylation of anthocyanins but also flavonol glycosides (Bontpart et al., 2015) and one such enzyme is dahlia anthocyanin 3-glucoside malonyltransferase, Dv3MaT (Suzuki et al., 2002). Four *C. trapeziformis* homologs of Dv3MaT were identified (*Ctr3MAT1–4*).

Subcellular flavonoid transport homologs such as Arabidopsis glutathione-S-transferase, Transparent Testa 19 (TT19) involved in the transport of cyanidin anthocyanins from the cytosol to the tonoplast (Sun et al., 2012b), and various plant ABC and MATE transporters that also facilitate the vacuolar transport of anthocyanins and flavonoids such as grape ABCC1, *A. thaliana* ABCC2, and grape anthoMATEs (Gomez et al., 2009; Francisco et al., 2013; Behrens et al., 2019) were also identified (e.g., *CtrTT19a/b*, *CtrABCC1*, *CtrAM1*, and *CtrABCC2a/b*). These include transcripts encoding homologs of petunia H^+ P-ATPase PH5 and Na^+/H^+ and K^+/H^+ antiporter NHX1 (Faraco et al., 2014) and Japanese morning glory NHX1 (Fukada-Tanaka et al., 2000) involved in the regulation of vacuolar pH were also identified (*CtrPH5* and *CtrNHX1*). Finally, a homolog to the gene encoding CHIL that binds to CHS to serve as a rectifier, thus, facilitating the flux of phenylpropanoid pathway precursors to the flavonoid pathway (Waki et al., 2020), was also found (*CtrCHIL*).

Developmental- and Tissue-Specific Expression of Anthocyanin and Flavonol Glycoside Pathway-Related Genes

In this study, we considered four pairwise developmental stage comparisons between the *vyb*, *vmb* (devC1/L1) and *sflw* (devC2/L2) stages across the callus and lamina: Three tissue-specific comparisons between callus and lamina across the respective *vyb* (ts1), *vmb* (ts2), and *sflw* (ts3) developmental stages corresponding with the samples profiled for anthocyanin and flavonol glycosides. First, we clustered candidate anthocyanin and flavonol glycoside pathway-related genes based on their expression abundance (**Supplementary Figure 3**). Four clusters (cluster A–D) of shared expression intensities were identified. Most notably, *CtrCHS1*, *CtrF3H1/2*, and *CtrLDOX1* were highly expressed (FPKM > 100) especially in the callus and/or lamina of *vyb* (Cluster D) while lowly expressed (FPKM < 1) transcripts such as *CtrF3pH2/3*, *CtrMAT4*, and *CtrCHS2/3* were observed in Cluster C. Furthermore, hierarchical clustering based on the patterns of differential expression revealed clusters containing genes that were differentially expressed between callus and lamina in both specific and multiple developmental stages (**Supplementary Figure 4**). For example, several groups of genes were coordinately downregulated to varying degrees (i.e., cluster 3–5) during the

transition from *vyb* to *vmb* and *sflw* stage regardless of tissue type. In cluster 5, there were many genes encoding enzymes involved in the formation of early anthocyanin and flavonol precursors such as chalcones, flavanones, and dihydroflavonols. These include two *CHS* (*CtrCHS1/3*), one *CHI* (*CtrCHI1*), two *F3H* (*CtrF3H1/2*), and two *F3'H* (*CtrF3pH1/4*) genes. Conversely, cluster 3 contained mostly downstream flavonoid modification genes and displayed stronger and more consistent developmental stage downregulation in both tissues compared to other clusters. Across several clusters in both tissue types over flower development, coordinated regulation was not limited to the upstream biosynthetic pathway but also included downstream modification and/or transport genes. For example, *CtrAM1/2*, *CtrPH5*, and two flavonoid-related glycosyltransferase homologs *CtrUGT71B1* and *CtrUGT78D2b* were present in cluster 5, *CtrTT19b* in cluster 2, and *CtrCOMT1/4* in cluster 4. However, cluster 1 and 3 mostly contained flavonoid modification or transport genes, while the majority of transcripts in cluster 1 encoded downstream transport genes and displayed developmental stage upregulation in both the callus and lamina.

The Chemical Basis and Patterns of Gene Expression in the Black Callus

The total anthocyanin content was substantially higher in callus tissue compared to the lamina across all developmental stages (**Figure 2A**). For example, at the *vyb* stage, total anthocyanin quantities were up to 1300 times higher in the callus than in the lamina. Even after the late peak of anthocyanin production in the *sflw* stage, the callus anthocyanin content remained 5 times greater than in the lamina. Across developmental stages, total anthocyanin content of the callus averaged 1.8 mg g^{-1} FW in *vyb* stage, before peaking in *mb* (3.0 mg g^{-1} FW) and then decreasing in *vmb* and *sflw* to comparable levels observed in *vyb* (2.1–2.2 mg g^{-1} FW). The majority of anthocyanins were cyanidin-based (96–97%) regardless of developmental stage, with relatively low amounts (0.02–0.06 mg g^{-1} FW) of peonidin- and pelargonidin-based anthocyanins (**Supplementary Figure 5**). Across the different developmental stages, the pattern of accumulation for individual anthocyanins were largely consistent with total anthocyanin content (**Supplementary Figure 5**).

Four main types of flavonol glycosides were found in the callus, with the proportions of quercetin (Q)-, isorhamnetin-(IR), kaempferol- (K), and kaempferide- (Kde) based flavonol glycosides ranging between 36–44%, 26–32%, 12–18%, and 12–16%, respectively (**Supplementary Figure 6**). Compared to the anthocyanins, less variation in total flavonol glycoside content was found across the different bud (1.4–2.1 mg g^{-1} FW) and flower (1.7 mg g^{-1} FW) stages (**Figure 2B**). Major flavonol glycosides such as QG and IRG were often most abundant in early developmental stages (i.e., pre *mb* before decreasing slightly in *vmb* and *sflw*) while others such as KG, K(mal)G, and KdeGG showed subtle increments from *vyb* to *mb* before decreasing in *sflw* (**Supplementary Figure 6**). However, the ratios of total anthocyanins to flavonol glycosides in the callus were broadly similar across developmental stages (approximately 1.5 to 1).

One key pattern of gene expression in the callus, especially at the *vyb* stage, was that genes related to anthocyanin and/or flavonol glycoside biosynthesis (e.g., *CtrF3H1/2*, *CtrDFR1*, *CtrLDOX1/2*), transport (e.g., *CtrAM1/2*, *CtrTT19b*), and vacuolar acidification (*CtrPH5*)-related pathways showed significantly higher gene expression when compared to the lamina. Most notably, *CtrDFR1* and *CtrLDOX1/2* transcripts peaked in the callus of *vyb* and were 10–35-fold higher compared to the lamina (**Figure 3** and **Supplementary Figure 4**). Despite evidence for coordinated downregulation, and hence lower gene expression levels for many of these genes from *vmb* onward, some transcripts such as *CtrCHS1*, *CtrFLS1/2*, *CtrMAT1*, *CtrCOMT2/3*, *CtrNHX1*, *CtrAM1/2*, and *CtrABCC2a/b* continued to exhibit higher gene expression levels in the callus compared to the lamina in one or more stages post *vyb*.

The Chemical Basis and Patterns of Gene Expression in the Green Lamina

While the total anthocyanin content was much lower in the lamina than the callus, trace levels were nonetheless detected from the *vyb* stage (0.014 mg g^{-1} FW) with a steady increase to appreciable amounts at the *sflw* (0.42 mg g^{-1} FW) stage. These patterns match the observed transition from green buds to the reddish green mature flower. Like the callus, cyanidin-based anthocyanins were the major anthocyanin constituent of the lamina regardless of developmental stages (99–100%), with very low amounts of peonidin- and pelargonidin-based anthocyanins. Accordingly, CG and C(mal)G content increased steadily from *vyb* to *sflw* in the labellum. Interestingly, total flavonol glycoside content in the lamina closely mirrored the profile of total anthocyanins in the callus (**Figure 1B**). The average total flavonol glycoside content increased from 2.6 to 3.4 mg g^{-1} FW in the *vyb* and *mb*, before decreasing sharply (1.5 mg g^{-1} FW) in *sflw*. This decreasing pattern is concomitant with the sharp rise in total anthocyanins from *vyb* to *sflw*.

When compared with the callus, there was substantial variation in the ratio of total flavonol glycosides to anthocyanins across developmental stages of the lamina with the top two-most extremes observed in *vyb* (1800:1) and *sflw* (3.5:1). Notably, the total (and most individual) flavonol glycoside content in the lamina of *sflw* was the lowest of all developmental stages, directly contrasting with the total anthocyanin content which was greatest in this stage (**Figure 2**). Regardless of developmental stage, isorhamnetin (43–46%) and kaempferide (29–35%) flavonol glycosides were more prevalent in the lamina, while quercetin-(12–19%) and kaempferol-based flavonol glycosides (5–7%) were less prevalent relative to the callus (**Supplementary Figure 6**). Unlike the callus, in the lamina most major flavonol glycosides either displayed a marked increase in content (e.g., QG, IRGG) or remained constant [e.g., IRG, KdeGG, Kde(mal)GG] during the transition from *vyb* to *mb*, followed by a large decrease in the mature stages (**Supplementary Figure 6**).

Coordinated regulation of *CtrFLS1/2* and *CtrCOMT2/3* transcripts (cluster 4) was a striking feature of the lamina. Four transcripts were significantly upregulated in the lamina compared to callus of *vyb*, however, at later stages of *vmb* and *sflw*,

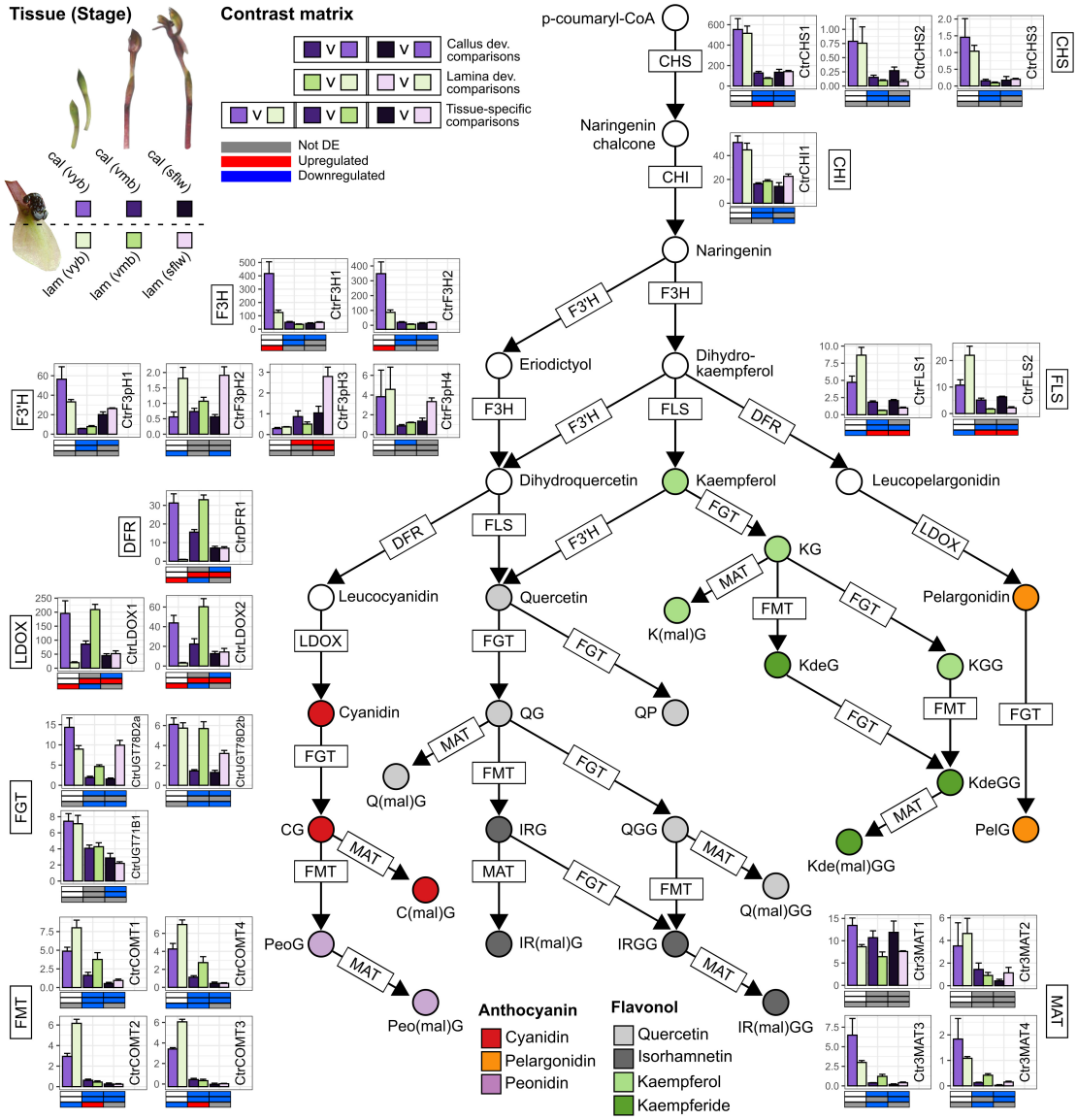


FIGURE 3 | Anthocyanin and flavonol metabolism in the callus and labellum tissues of *Chiloglottis trapeziformis* during flower development. Bars depict the average (±s.e.) normalized expression values (FPKM) of anthocyanin and flavonol biosynthesis and modification pathway genes in the calli and labellum lamina tissues of very young buds (*vyb*), very mature buds (*vmb*), and mature sunflowers (*sflw*). Colored circles indicate putative anthocyanins and flavonols identified in *Chiloglottis trapeziformis* callus (cal) and labellum lamina (lam) tissues (**Table 1**). Various shades of purple and green depict the relevant color callus and labellum lamina tissues at various flower developmental stages. The contrast matrix indicates the differential expression outcomes based on comparisons in **Supplementary Figure 1**. Red, blue, and gray indicate significant upregulation (FDR < 0.05, $\log_2$FC > 0.5), downregulation (FDR < 0.05, $\log_2$FC < –0.5), and no significant differential expression in selected comparisons, respectively. CHS, Chalcone synthase; CHI, Chalcone isomerase; F3H, Flavanone 3-hydroxylase; F3′H, Flavonoid 3′-hydroxylase; F3′5′H, Flavonoid 3′5′-hydroxylase; DFR, Dihydroflavonol 4-reductase; LDOX, Leucoanthocyanidin dioxygenase; FLS, Flavonol synthase; FGT, Flavonoid/anthocyanin glucosyltransferase; FMT, Flavonoid/anthocyanin O-methyltransferase; MAT, Flavonoid/anthocyanin malonyltransferase.

higher callus expression was commonly observed. Coordinated downregulation during the transition from *vyb* to *vmb* and *sflw* stages was also common for many anthocyanin and flavonol glycoside pathway genes in the lamina (i.e., cluster 3–5). Another striking feature of the lamina was a strong (but lagging the callus) developmental upregulation of genes encoding *CtrDFR1*, *CtrLDOX1/2*, *CtrAM1/2*, and *CtrTT19b* (cluster 2). For example, *CtrDFR1* was consistently upregulated by 12- to 55-fold in *vmb* and *sflw* compared to *vyb*. Similarly, *CtrLDOX1/2* were upregulated by four- to 30-fold in *vmb* and flowers compared to *vyb*. Interestingly, these transcripts were only upregulated in the lamina compared to the callus at the *vmb* stage. Other notable transcripts include *CtrUGT78D2a/b*, *CtrMAT3/4*, *CtrCOMT2/3*, and *CtrPH5,* which were more highly expressed in the labellum than the callus in one or more developmental stages post *vyb* (**Figure 3** and **Supplementary Figure 4**).

DISCUSSION

Overview

The goal of this study was to combine targeted metabolite profiling and transcriptomic analysis to investigate the chemical composition and patterns of tissue- and developmental-specific color pathway gene expression in the labellum of *Chiloglottis trapeziformis*. In this sexually deceptive orchid (as in most others), the labellum is the center of the orchid-pollinator interaction, functioning as the olfactory, visual, and tactile mimic of the female of the pollinating species (de Jager and Peakall, 2016, 2019). Furthermore, the color contrast between the dark black 3D callus structure mimicking the female (**Figure 1A**) and the remainder of the reddish green labellum lamina is predicted to aid detectability to the male pollinator. In the discussion that follows, we explore in more detail hypotheses for the molecular basis of the contrasting labellum colors, and consider evolutionary implications and future research directions.

The Diversity of Anthocyanin and Flavonol Co-pigments in *Chiloglottis trapeziformis*

Cyanidin-based anthocyanins are widespread floral pigments across the flowering plants, and are also known from many orchid species (Iwashina, 2015). However, as one of the largest plant families, it is not surprising that anthocyanins have only been investigated in a fraction of the known genera and species of orchids. Pigment chemistry and biosynthesis has been most well studied in orchid genera of horticultural interest, such *Phalaenopsis* (Hsu et al., 2015; Liang et al., 2020), *Vanda* (Khunmuang et al., 2019), *Cattleya* (Li et al., 2020), and *Cymbidium* (Wang et al., 2014) but there have also been several studies of wild species which span a diversity of orchids with both rewarding and deceptive pollination strategies (George et al., 1973; Strack et al., 1989; Fossen and Øvstedal, 2003; Vignolini et al., 2012; Kellenberger et al., 2019; Zhang et al., 2020). Notably, cyanidin-based anthocyanins are major anthocyanin pigments of flowers in the European sexually deceptive genus *Ophrys* (Strack et al., 1989; Vignolini et al., 2012). Unlike anthocyanins, the constituent flavonol glycosides (or any other flavonoid classes) remain poorly understood across the Orchidaceae.

Despite the exceptional diversity (>1850 species) of the Australian orchid flora (Jones, 2021; Peakall et al., 2021), to the best of our knowledge only one previous study of floral anthocyanins has been published. This study, dating back to the 1970s, reported a qualitative survey of floral anthocyanins across some brightly colored and non-sexually deceptive species. The finding indicated the presence of cyanidin-, delphinidin-, petunidin- and malvidin-based anthocyanins (George et al., 1973). The present study is thus the first to investigate in detail the chemical and molecular basis of flower color in an Australian orchid. We have confirmed that cyanidin-based anthocyanins are the major pigment constituents of the labellum of *Chiloglottis trapeziformis* flowers. The labellum lamina also accumulates other mono- and di-hydroxylated anthocyanins and flavonol glycosides, but no trihydroxylated forms were detected. Additionally, among the many possible acyl modifications of the anthocyanin and flavonol glycosides, only the malonylated form was detected (**Table 1**). The composition of flavonol glycosides was more complex than the anthocyanin profile in both tissue types. In the callus, isorhamnetin- and quercetin-based compounds were equally dominant while the lamina accumulated more isorhamnetin- and kaempferide-based flavonol glycosides (**Figure 2B** and **Supplementary Figures 5, 6**).

Patterns of Anthocyanin and Flavonol Glycoside Levels in *Chiloglottis trapeziformis*

Dissection of the tightly closed green bud at the *vyb* stage shows that the callus is already darkly colored. By the *mb* stage the callus appears "black" in color, and it remains so for the lifetime of the flower (Amarasinghe et al., 2015). On the other hand, the color of the lamina changes from uniform green at *vyb* to greenish red at *vmb*, but only reaches its maximum reddish green color in flowers that have been open in sunlight for several days (Amarasinghe et al., 2015). Consistent with these observations, peak anthocyanin accumulation was observed in the *vmb* stage of the callus, compared with the final *sflw* stage of the lamina (**Figure 2A**). Furthermore, the lower levels of lamina anthocyanins are spread thinly across the larger surface area, compared with the smaller callus (**Figure 1A**). Reflectance measurements, projected into hymenopteran vision space, indicate that the dark achromatic callus should be readily distinguished by the pollinator from the contrasting lamina, even at full flower maturity (de Jager and Peakall, 2016).

By contrast with the anthocyanins, total flavonol glycoside levels were broadly similar across callus and lamina tissues and seemingly inversely correlated with the total anthocyanin levels. Nonetheless, some different patterns of composition and changes with development stage were evident (**Figure 2B** and **Supplementary Figure 6**). Critically, when compared to the lamina, lower levels of the methylated forms of Quercetin (Isorhamnetin) and Kaempferol (Kaempferide) were found in the callus (**Figure 2B**).

Patterns of Gene Expression in the Flower Color Pathway Genes of *Chiloglottis trapeziformis*

Figure 3 shows an overlay of the patterns of gene expression in the *C. trapeziformis* labellum partitioned by tissue type and development stage onto the general biosynthetic pathway for anthocyanin and flavonoid production in plants. The gene expression patterns in combination with the outcomes of the hierarchical clustering analysis (**Supplementary Figures 3, 4**) and knowledge of the anthocyanin and flavonol glycoside levels (**Figure 2**), provide insights into the probable genetic basis of the contrasting labellum colors of *C. trapeziformis*.

The two early pathway genes, *CtrCHS1-3* and *CtrF3H1-2*, were equally highly expressed in both callus and lamina tissues of the *vyb* stage, indicating an efficient supply of precursors is available for both anthocyanin and flavonol glycosides biosynthesis in

either tissue. Whereas, the dedicated anthocyanin pathway genes associated with the production of anthocyanins (cyanidin, peonidin, and pelargonidin), *CtrDFR1*, *CtrLDOX1-2* were highly expressed in the *vyb* stage of the callus, but not the lamina. Conversely, the *CtrFLS1-2* associated with the production of the flavanol glycosides, were significantly downregulated in the callus tissue. The methyltransferase, FMT (*CtrCOMT1-4*) genes mirrored the *FLS* gene expression patterns, consistent with the observed increasing accumulation of the methylated flavonol glycosides in the lamina from *vyb* to *mb* stages (**Figure 3** and **Supplementary Figure 6**).

By the *vmb* stage, expression levels of the precursor supply genes, CHS and CHI, were significantly downregulated relative to the *vyb* stage in both tissue types, and remained low across the remaining developmental stages. In the callus, the intermediate anthocyanin pathway genes (F3H, F'3H, DFR and LDOX), generally showed a similar pattern. However, across the remaining developmental stages, there was a strong lag in the peak expression of *CtrDFR1* and *CtrLDOX1-2* between the lamina and callus. This suggests that the combination of upregulation of the anthocyanin genes in the lamina, coupled with a diminished precursor supply is likely linked to the increasing but much lower levels of anthocyanin production in the lamina with development stage.

By the *sflw* stage, the formerly green color of the *vmb* stage, and the freshly opened flower was reddish-green in color. Consistent with this observation, anthocyanin levels in the lamina, while much lower than in the callus, peaked at the *sflw* stage (**Figure 2A**). Subtle gene expression shifts appear to be consistent with these flower color changes. For example, expression of CHS (*CtrCHS1*) and CHI (*CtrCHI1*) increased in both callus and lamina relative to the *vmb* stage. More importantly, callus expression was significantly upregulated relative to the lamina. Thus, it seems likely that precursor supply, while remaining diminished relative to the *vyb* stage, should increase somewhat in this final developmental stage (**Figure 3**).

Patterns of Gene Expression in Genes Potentially Modifying the Flower Color of *Chiloglottis trapeziformis*

Beyond the direct anthocyanin and flavonol metabolism genes (**Figure 3**), other genes such as those involved in modification, transport, and sequestration, as well as genes associated with chlorophyll production may influence the flower color of *C. trapeziformis*. Notably, some genes involved in the modification of anthocyanins and flavonol glycosides showed uniform levels of gene expression across tissue types and developmental stages. This included the flavonoid glucosyltransferases, FGT (*CtrUGT78D2a/b* and *CtrUGT71B1*) and malonyltransferases, MAT (*Ctr3MAT1-4*) (**Figure 3**). Thus, it is likely that these downstream modification steps are not rate-limiting during the active accumulation of peak anthocyanins in the callus and the flavonols in the lamina (*vyb-mb* stages), nor in the final stage of the lamina becoming reddish green (*sflw*). We further predict that one or more of the MAT transcripts (e.g., *Ctr3MAT1-2*) are involved in the production of the malonylated anthocyanins (such as cyanidin malonylglucoside). Such derivatives were abundant (ca. 31–33% of total anthocyanins) in both the callus and labellum (**Figure 2A** and **Supplementary Figure 5**), and their accumulation profiles closely mirrored their non-malonylated anthocyanin precursors. The high abundance of malonylated anthocyanin derivatives in the callus, may be relevant to the preservation of its distinct dark color across developmental stages. Processes downstream of anthocyanin biosynthesis, such as the acylation of anthocyanins (e.g., malonylation) can have important consequences for floral color. For example, acylation has many functional roles including the protection of glycosides from enzymatic degradation, anthocyanin structure stabilization, and assistance in vacuolar uptake and sequestration (Bontpart et al., 2015). Importantly, these various functions may all serve to aid the stabilization of pigments in flowers (Suzuki et al., 2002; Luo et al., 2007).

Among the predicted transporter gene homologs, TT19 (*CtrTT19b*) showed a pattern of gene expression which mirrored *DFR* and *LDOX* genes (cluster 2, in **Supplementary Figure 4**). For example, at the *vyb* stage there was high and differential expression in the callus relative to the lamina, with this pattern reversing at the *vmb* stage. Expression levels in the callus plateaued at the *vmb* stage, while they peaked a second time in the lamina at the *sflw* stage. Thus, high expression levels at this gene appear to correspond with the respective starting points of anthocyanin accumulation and color development which started first in the callus with a peak by *mb*, but lagged in the lamina peaking at *sflw* (**Figure 2A**). In peach, expression of a glutathione S-transferase gene, *PpGST1* strongly correlates with anthocyanin accumulation in fruit tissues. Accordingly, transient overexpression of *PpGST1* increased the anthocyanin content in the fruit of yellow-fleshed nectarine while silencing of *PpGST1* in blood-fleshed (anthocyanin-rich) peach decreased anthocyanin accumulation resulting in white flesh (Zhao et al., 2020). Interestingly, *PpGST1* is also a locus containing functional and non-functional alleles that co-segregated with white and non-white (i.e., pink and red) flowers (Lu et al., 2021). In the purple-flowered cyclamen, expression of an anthocyanin-related GST, *CkmGST3* and upstream core biosynthetic genes such as *CkmF3'5'H* and *CkmDFR2* were tightly co-regulated during flower development. These transcripts were strongly expressed in paler pigmented petals and correlated with the sharp rise in anthocyanin accumulation compared to fully pigmented purple petal (Kitamura et al., 2012).

The expression patterns of a homolog to a gene implicated in the regulation of vacuolar pH, PH5 (*CtrPH5*), also showed some similarity to those of DFR, LDOX and TT19 genes (*CtrDFR1*, *CtrLDOX1/2*, and *CtrTT19b*) (**Supplementary Figures 3, 4**). For example, peak expression occurred in the callus at the *vyb* stage with significant downregulation as the flower developed. By contrast, in the lamina expression peaked in the *vmb* stage. Interestingly, at a homolog of NHX1 (*CtrNHX1*), gene expression levels were high in both tissue types at *vyb*, and tended to remain high in the callus across developmental stages. However, in the lamina, gene expression levels were differentially downregulated relative to the callus. These findings may indicate that vacuolar

pH is lowest in the *vyb* stage of the callus, and in the *vmb* stage of the lamina. Furthermore, we predict that vacuolar acidification in the callus during active anthocyanin accumulation could play a synergistic role in conferring the strong "black" color of the callus. Different levels of vacuolar acidification are known to cause color shifts in flowers. For example, a less acidic vacuolar pH is often associated with blue colored flowers (e.g., *Petunia*, *Ipomea*, *Hydrangea*, and *Phalaenopsis* spp.) that otherwise share similar anthocyanin composition to other non-blue colored flowers (Fukada-Tanaka et al., 2000; Yoshida et al., 2003; Verweij et al., 2008; Liang et al., 2020). Presently, genes implicated in the regulation of vacuolar pH with relevance to flower color have only been established for a few species (i.e., Petunia and Japanese morning glory). Notably, the loss-of-function mutation of petunia *PH5* or the downregulation of *PH5* using RNA interference reduces vacuolar acidification in petals, resulting in "blue-violet" flower color in mutants and transformants compared to "red-violet" wild-type flowers (Faraco et al., 2014). Conversely, a mutation of morning glory *NHX1* impairs vacuolar alkalization, resulting in purple petals in fully bloomed flowers instead of bright blue colors in the wild-type (Fukada-Tanaka et al., 2000).

Chlorophyll is often present in petals during early developmental stages but usually decreases to trace levels in the typical brightly colored flowers of most plants. This loss of green coloration as the flower develops is considered crucial for making flowers more visible to animal pollinators against the green leafy background. However, examples of green colored flowers are nonetheless widespread across the flowering plants. Dull green and red flowers are particularly frequent among sexually deceptive plant species. These include hundreds of examples within the Australian genera *Caladenia*, *Chiloglottis*, *Cryptostylis*, *Drakaea* and *Pterostylis*, among others; tens of European *Ophrys* species, and several sexually deceptive species from South America (Bohman et al., 2016; Peakall et al., 2020).

Despite the widespread occurrence of green colored flowers and a thorough understanding of photosynthesis- and chlorophyll-related pathways (Hörtensteiner, 2006; Tanaka and Tanaka, 2006), little is known about how green colors occur in mature flowers. To date, the best clues are offered in carnations. Insights from carnation (*Dianthus caryophyllus* L.) floral transcriptomes reveal that expression of photosynthesis and chlorophyll biosynthesis genes are tightly associated with chlorophyll content, and thus coloration of green- and white-flowered cultivars (Ohmiya et al., 2014, 2017). Although the types and quantities of chlorophyll were not measured in this study, we predict that retention of chlorophyll pigments is pivotal for the green coloration of *Chiloglottis trapeziformis* flowers. When compared to the callus, the lamina was characterized by lower levels of anthocyanins and higher levels of colorless flavonol glycosides. Furthermore, in the lamina, consistently higher levels of gene expression were detected at photosynthesis- and chlorophyll-related genes across all developmental stages, indicative of enhanced chlorophyll biosynthesis in the lamina (**Supplementary Figure 1** and **Supplementary Data 1**). Thus, the combination of low anthocyanins, high flavonol glycosides and sustained chlorophyll production are likely to be the basis of the predominant green coloration of the lamina from bud to flower opening.

Transcriptional Control of Color Pathways in the Flower

A detailed investigation of the transcriptional control of developmental and tissue-specific of pigmentation is beyond the scope of this study. Nonetheless, we highlight nine R2R3-MYB TF, two bHLH, and one WDR gene(s) potentially implicated in anthocyanin pigmentation in the flower of *Chiloglottis trapeziformis* (**Figure 4** and **Supplementary Data 4**). The striking similarity in differential expression profile between *CtrMYB2a* and *CtrF3H1/2*, *CtrDFR1* and *CtrLDOX1/2* indicate that some shared and dedicated anthocyanin pathway genes are regulatory targets of *CtrMYB2a*. *CtrMYB2a/b* belongs to subgroup 5 of R2R3-MYB TFs that contains several known orchid tissue-specific floral pigmentation regulators (Chiou and Yeh, 2008; Hsu et al., 2015; Li et al., 2020). For example, the *CtrMYB2a* homolog in *Phalaenopsis*, *PeMYB2* is correlated with anthocyanin abundance in the petals and sepals of several cultivars and is highly expressed in full-red pigmented flowers. Transient overexpression of *PeMYB2* in white-flowered *P. aphrodite* ssp. *formosana* induced the expression of endogenous flavonoid and anthocyanin biosynthesis genes (e.g., *F3H*, *F3'H*, *DFR*, or *ANS* genes) and the development of reddish-pink pigmentation (Hsu et al., 2015).

Unlike orchid subgroup 5 R2R3-MYBs, the role of other subgroups (e.g., 6, 7, 20, and 4) in the regulation of flavonoid metabolism and pigmentation is unknown in the Orchidaceae. *CtrMYB12a/b/c* belongs to subgroup 7 R2R3-MYB TFs and is related to Arabidopsis MYB12 which commonly regulate flavonol glycoside accumulation via the activation of *FLS* and shared flavonoid biosynthetic pathway genes (Stracke et al., 2007). In particular, *CtrMYB12a* shares differential expression patterns with *CtrFLS1/2* and *CtrCOMT1-4*. We predict that *CtrMYB12a* regulates flavonol glycoside accumulation by targeting *CtrFLS1/2* and *CtrCOMT1-4*. Conversely, *CtrMYB12c* does not share expression similarities with *CtrFLS1/2*. Instead, *CtrMYB12c* expression coincides with *CtrMYB2a*, suggesting that *CtrMYB12c* is more relevant to the regulation of anthocyanin than flavonol glycoside biosynthesis. In plants, the transcriptional activation strength of *FLS* are known to differ between orthologs of AtMYB12 (Mehrtens et al., 2005) but also associated paralogs of certain species (Shan et al., 2020). In *Freesia*, all four AtMYB12 orthologs FhMYBF1–4 positively regulated several shared flavonoid pathway genes and *FhFLS1* at different magnitudes when overexpressed, however FhMYBF4 lack the capacity to highly induce *FhFLS1* expression (Shan et al., 2020). Some studies have even shown that AtMYB12 orthologs potentially regulate the anthocyanin biosynthetic pathway genes (Czemmel et al., 2009; Zheng et al., 2019; Zhong et al., 2020).

CtrMYB32 and CtrMYB4a/b/c belong to subgroup 4 R2R3-MYB TFs which are active repressors that fine tune flavonoid levels by balancing the inductive effects of the MBW complex (Chen et al., 2019; LaFountain and Yuan, 2021). In particular,

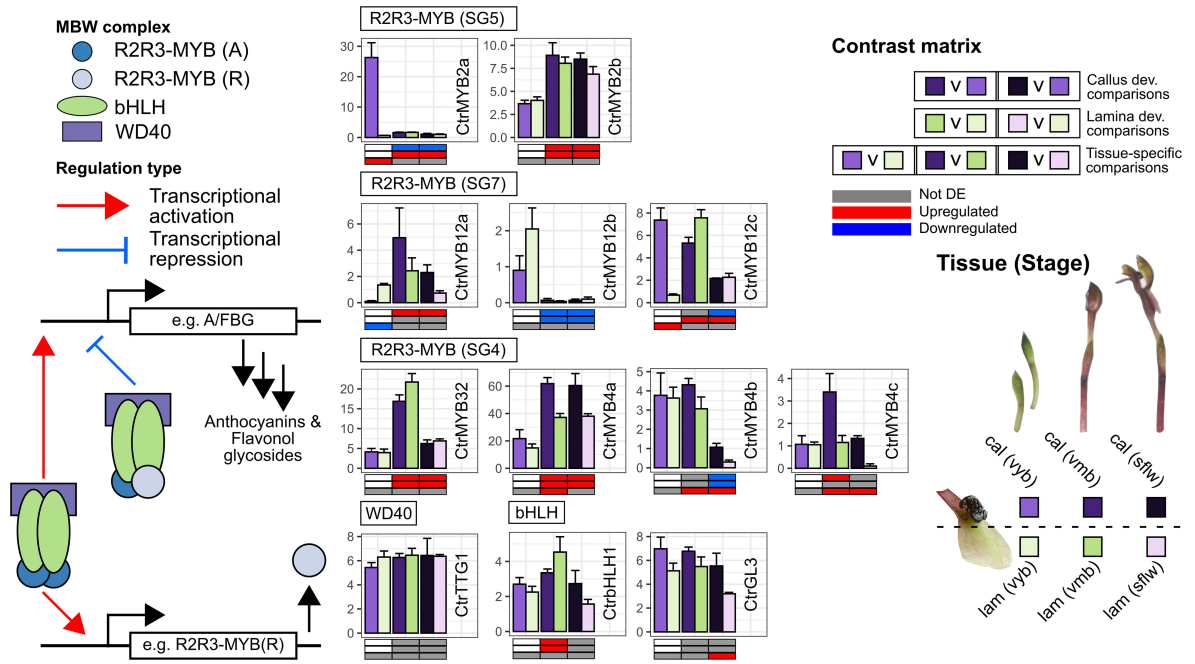


FIGURE 4 | Transcriptional regulation of anthocyanin and flavonol glycoside pathway genes in the callus and labellum tissues of *Chiloglottis trapeziformis* during flower development. Bars depict the average (±s.e.) normalized expression values (FPKM) of putative transcriptional regulatory genes encoding relevant R2R3-MYB, basic-helix-loop-helix (bHLH), and WD40 repeat (WDR) proteins in the calli and labellum lamina tissues of very young buds (*vyb*), very mature buds (*vmb*), and mature sunflowers (*sflw*). A/FBG indicate shared flavonoid as well as anthocyanin/flavonol glycoside biosynthetic pathway genes. Various shades of purple and green depict the relevant color callus and labellum lamina tissues at various flower developmental stages. The contrast matrix indicates the differential expression outcomes based on comparisons in **Supplementary Figure 1**. Red, blue, and gray indicate significant upregulation (FDR < 0.05, $\log_2$FC > 0.5), downregulation (FDR < 0.05, $\log_2$FC < –0.5), and no significant differential expression in selected comparisons, respectively.

CtrMYB32 and *CtrMYB4a* were consistently upregulated in *vmb* and *sflw* while the opposite was true for *CtrMYB4b* compared to *vyb* regardless of tissue type. Furthermore, *CtrMYB4a/b/c* showed significantly higher callus gene expression compared to the lamina in one or more stages post *vyb*. Since further anthocyanin pathway activation by the MBW complex does not benefit the establishment of color or contrast, in the callus, we predict that *CtrMYB4a/b/c* co-operatively repress the inductive effects of the MBW activity in *vmb* and *sflw* once the distinct black color has been established in *vyb*. Conversely, the weaker pathway repression in the labellum compared to the callus post *vyb* may signal the accumulation of anthocyanins.

Implications and Future Directions

In the last decade, considerable progress has been made toward understanding the diversity and distribution (Peakall et al., 2010) and the regulation of the chiloglottones which underpin the sex pheromone mimicry of *Chiloglottis* orchids (Falara et al., 2013; Amarasinghe et al., 2015). However, the chemical and molecular basis of the floral pigmentation in *Chiloglottis,* and indeed of any Australian sexually deceptive orchids, has remained unknown. Here, we demonstrate that the black callus of *Chiloglottis trapeziformis* flowers is characterized by a high abundance of cyanidin-based anthocyanins and diverse flavonol glycoside co-pigments concentrated in a small, highly localized area. Conversely, a high abundance of flavonol glycoside co-pigments and low anthocyanin content is characteristic of the lamina. Differential expression of specific genes controlling the branchpoints into the respective anthocyanin (e.g., *DFR* and *LDOX*) and flavonol glycoside biosynthesis (e.g., *FLS*) and its subsequent modification toward methylated derivatives (e.g., *FMT*) is likely pivotal for generating the contrasting dark color of the callus against the green to reddish green lamina (**Figure 3**).

Beyond *Chiloglottis*, inconspicuous dull colored flowers with tissue-specific dark maroon to black pigmentation is common in sexually deceptive species. In Australia, several orchid genera have some sexually deceptive species which exhibit this tissue-specific contrast (e.g., within *Caladenia*, *Calochilus*), and in some genera, this is exclusively the case (e.g., *Drakaea*, *Cryptostylis*, *Paracaleana*, *Leporella*, *Caleana*) (Gaskett, 2011). Other examples of tissue-specific dark colors in sexually deceptive species outside Australia are the dark labella of many *Ophrys* species (Bradshaw et al., 2010) such as *Ophrys speculum* (**Figure 1F**), the maroon-black callus structure on the pink labellum of *Serapias lingua* (Vereecken et al., 2012; Pellegrino et al., 2017), the large elongated dark maroon labellum of Asian beetle-pollinated *Luisia teres* (Arakaki et al., 2016), the triangular-shaped dark maroon pigmentation on the labellum of bee-pollinated *Mormolyca ringens* (Singer et al., 2004), the central dark maroon labellum and column on an otherwise bright yellow corolla of *Telipogon peruvianus* (Martel et al., 2016, 2019), a combination of dark maroon-colored floral structures including antennae-like petals, anther, and labellum of south African

beetle-pollinated *Disa forficaria* orchids (Cohen et al., 2021), and the raised black spots on the yellow to bright orange ray florets of some fly pollinated *Gorteria diffusa* (Ellis and Johnson, 2010) morphotypes (**Figure 1G**). In these cases, the dark pigmentation likely serves several key purposes important for pollination, including mimicking the color of female insects, providing strong chromatic and achromatic contrast, and aiding the exploitation of pollinator sensory biases.

All of these examples showcase a strong correlation between sexual deception and dark colored flowers across many different evolutionary lineages. As such, understanding the chemical and genetic mechanisms of how such dark coloration is attained is pivotal for understanding how sexual deception evolves. We predict that similar chemical constituents and/or spatiotemporal gene expression of underlying pathway genes (e.g., *DFR*, *LDOX,* and *FLS*) will be pivotal for dark, potentially contrasting tissue-specific pigmentation in other sexually deceptive species. Currently, *Ophrys* is the best-characterized genus to support this prediction. In some confirmed sexually deceptive *Ophrys* orchids such as *O. insectifera*, *O. sphegodes*, and *O. speculum* (Schiestl et al., 1999; Vignolini et al., 2012), cyanidin-based anthocyanins are the major floral anthocyanin pigment present in whole flowers or the labellum (Strack et al., 1989). Furthermore, the dark labellum color of *Ophrys speculum* (**Figure 1F**) was demonstrated to contain a blend of anthocyanin and quercetin flavonol glycoside co-pigments (Vignolini et al., 2012). Transcriptomic surveys of *O. exaltata*, *O. sphegodes*, *O. aymoninii,* and *O. garganica* (Sedeek et al., 2013; Piñeiro Fernández et al., 2019) reveal that anthocyanin and flavonol glycoside biosynthesis pathway genes are active in the dark labella of these species. In *Gorteria diffusa* that possess dark tissue-specific pigmentation flanked by brightly colored orange ray florets, it will be of interest to determine the types of floral pigments, cellular chemistry, and genetic mechanisms involved. In this latter case, the relevance of carotenoid (e.g., xanthophylls) or flavonoid (e.g., chalcones and aurones) pigment pathways are predicted to be the driver of the contrast.

Future studies will also be required to determine the key transcription factors involved and how they regulate the expression of key flavonoid biosynthetic, modification, and transport pathway enzymes in the flowers of *Chiloglottis* and other sexually deceptive plants. Members of the R2R3-MYB, bHLH, and WD-Repeat proteins, including those that we have briefly highlighted in this study, are prime candidates given that they are key determinants of anthocyanin pigmentation intensity and also patterning in many flowers (Davies et al., 2012; Wessinger and Rausher, 2012). However, MADS-box transcription factors should also be investigated as some are involved in establishing proper floral organ identity programs, photosynthesis, chlorophyll metabolism, and pigmentation in orchids (Pan et al., 2014; Hsu et al., 2021). The integration of various omics approaches (e.g., large-scale transcriptomic, metabolomic, ionomics) is also anticipated to provide a holistic view of how structural genes, regulatory factors, and cellular constituents (e.g., pigments, metal ions, pH) interact and mediate the evolutionary color transitions between sexually deceptive systems.

Our findings in *Chiloglottis trapeziformis* indicate that its floral color adaptations for sexual deception require complex tissue-specific coordinated regulation at multiple developmental stages, involving many genes and diverse pigment metabolic and cellular constituent regulating pathways. It is further of interest that the genus *Chiloglottis* belongs to a subtribe (Drakaeinae) in which all genera are exclusively sexually deceptive (with the exception of some rare self-pollinating cases) and all fall within a well-resolved clade on a long branch of the diverse predominantly Australasian tribe the Diurideae (Peakall et al., 2021). Thus, it is likely that the complex chemical and genetic basis of floral color mimicry may have been finely tuned by a long evolutionary process in these sexually deceptive orchids (Weston et al., 2014; Peakall et al., 2021). This is in contrast with some of the well-known flower color polymorphisms found within natural populations (Streisfeld and Rausher, 2011; Kellenberger et al., 2019), and even the flower color changes associated with key pollinator shifts that are often mediated by relatively simple genetic changes restricted to a few genes within the anthocyanin and flavonol glycoside pathways (Wessinger and Rausher, 2012). It will be of interest in future research to determine whether sexually deceptive lineages of more recent evolutionary origin than *Chiloglottis* (e.g., *Caladenia*, Caladeniinae; Diurideae) (Peakall et al., 2021) exhibit floral color adaptions based on simple genetic changes, as one might expect at the early stages in the evolution of sexual deception.

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: New sequence reads obtained for the sflw stage has been added to the existing BioProject accession PRJNA390683 and SRA study accession SRP1093281.

AUTHOR CONTRIBUTIONS

DW and JP performed the experiments. RP and DW secured the funding, designed the study, and coordinated the experiments and data analysis. DW wrote the article with assistance from RP and JP. All authors have read and approved the manuscript.

FUNDING

This work was supported by the Australian Research Council projects DE190100249 to DW and DP150102762 to RP, and Australian Government Research Training Program to JP.

ACKNOWLEDGMENTS

We thank Thy Truong, Adam Carroll, and Joseph Boileau at the Joint Mass Spectrometry Facility for assistance with LC-MS

operation, the Australian National Botanic Gardens, Canberra for permission and permits to take plant material from the gardens, and New South Wales (NSW) National Parks. Photographs were taken by Steven Johnson (**Figures 1D,G**), Jorun Tharaldsen (**Figures 1F,H**), and RP (**Figures 1A–C,E**).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.860997/full#supplementary-material

REFERENCES

Abdel-Aal, E. S. M., Young, J. C., and Rabalski, I. (2006). Anthocyanin composition in black, blue, pink, purple, and red cereal grains. *J. Agric. Food Chem.* 54, 4696–4704. doi: 10.1021/jf0606609

Albert, N. W., Davies, K. M., Lewis, D. H., Zhang, H., Montefiori, M., Brendolise, C., et al. (2014). A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* 26, 962–980. doi: 10.1105/tpc.113.122069

Allan, A. C., and Espley, R. V. (2018). MYBs drive novel consumer traits in fruits and vegetables. *Trends Plant Sci.* 23, 693–705. doi: 10.1016/j.tplants.2018.06.001

Amarasinghe, R., Poldy, J., Matsuba, Y., Barrow, R. A., Hemmi, J. M., Pichersky, E., et al. (2015). UV-B light contributes directly to the synthesis of chiloglottone floral volatiles. *Ann. Bot.* 115, 693–703. doi: 10.1093/aob/mcu262

Arakaki, N., Yasuda, K., Kanayama, S., Jitsuno, S., Oike, M., and Wakamura, S. (2016). Attraction of males of the cupreous polished chafer *Protaetia pryeri pryeri* (Coleoptera: Scarabaeidae) for pollination by an epiphytic orchid *Luisia teres* (Asparagales: Orchidaceae). *Appl. Entomol. Zool.* 51, 241–246. doi: 10.1007/s13355-016-0396-5

Behrens, C. E., Smith, K. E., Iancu, C. V., Choe, J. Y., and Dean, J. V. (2019). Transport of anthocyanins and other flavonoids by the arabidopsis ATP-Binding cassette transporter AtABCC2. *Sci. Rep.* 9, 1–15. doi: 10.1038/s41598-018-37504-8

Bohman, B., Flematti, G. R., Barrow, R. A., Pichersky, E., and Peakall, R. (2016). Pollination by sexual deception - it takes chemistry to work. *Curr. Opin. Plant Biol.* 32, 37–46. doi: 10.1016/j.pbi.2016.06.004

Bontpart, T., Cheynier, V., Ageorges, A., and Terrier, N. (2015). BAHD or SCPL acyltransferase? what a dilemma for acylation in the world of plant phenolic compounds. *New Phytol.* 208, 695–707. doi: 10.1111/nph.13498

Borghi, M., Fernie, A. R., Schiestl, F. P., and Bouwmeester, H. J. (2017). The sexual advantage of looking, smelling, and tasting good: the metabolic network that produces signals for pollinators. *Trends Plant Sci.* 22, 338–350. doi: 10.1016/j.tplants.2016.12.009

Boss, P. K., Davies, C., and Robinson, S. P. (1996). Anthocyanin composition and anthocyanin pathway gene expression in grapevine sports differing in berry skin colour. *Aust. J. Grape Wine Res.* 2, 163–170. doi: 10.1111/j.1755-0238.1996.tb00104.x

Bradshaw, E., Rudall, P. J., Devey, D. S., Thomas, M. M., Glover, B. J., and Bateman, R. M. (2010). Comparative labellum micromorphology of the sexually deceptive temperate orchid genus *Ophrys*: diverse epidermal cell types and multiple origins of structural colour. *Bot. J. Linn. Soc.* 162, 504–540. doi: 10.1111/j.1095-8339.2010.01033.x

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Chen, L., Hu, B., Qin, Y., Hu, G., and Zhao, J. (2019). Advance of the negative regulation of anthocyanin biosynthesis by MYB transcription factors. *Plant Physiol. Biochem.* 136, 178–187. doi: 10.1016/j.plaphy.2019.01.024

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, J., Liao, L., Zhou, H., Gu, C., Wang, L., and Han, Y. (2015). A small indel mutation in an anthocyanin transporter causes variegated colouration of peach flowers. *J. Exp. Bot.* 66, 7227–7239. doi: 10.1093/jxb/erv419

Chiou, C. Y., and Yeh, K. W. (2008). Differential expression of MYB gene (*OgMYB1*) determines color patterning in floral tissue of *Oncidium* gower ramsey. *Plant Mol. Biol.* 66, 379–388. doi: 10.1007/s11103-007-9275-3

Cohen, C., Liltved, W. R., Colville, J. F., Shuttleworth, A., Weissflog, J., Svatoš, A., et al. (2021). Sexual deception of a beetle pollinator through floral mimicry. *Curr. Biol.* 31, 1962–1969. doi: 10.1016/j.cub.2021.03.037

Czemmel, S., Stracke, R., Weisshaar, B., Cordon, N., Harris, N. N., Walker, A. R., et al. (2009). The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiol.* 151, 1513–1530. doi: 10.1104/pp.109.142059

Davies, K. M. (2008). "Modifying anthocyanin production in flowers," in *Anthocyanins*, eds C. Winefield, K. Davies, and K. Gould (New York, NY: Springer), 49–80. doi: 10.1007/978-0-387-77335-3_3

Davies, K. M., Albert, N. W., and Schwinn, K. E. (2012). From landing lights to mimicry: the molecular regulation of flower colouration and mechanisms for pigmentation patterning. *Funct. Plant Biol.* 39, 619–638. doi: 10.1071/FP12195

de Jager, M. L., and Peakall, R. (2016). Does morphology matter? An explicit assessment of floral morphology in sexual deception. *Funct. Ecol.* 30, 537–546. doi: 10.1111/1365-2435.12517

de Jager, M. L., and Peakall, R. (2019). Experimental examination of pollinator-mediated selection in a sexually deceptive orchid. *Ann. Bot.* 123, 347–354. doi: 10.1093/aob/mcy083

Gilbert, D. (2013). *EvidentialGene: Tr2aacds, mRNA Transcript Assembly Software*. Available online at: http://arthropods.eugenes.org/EvidentialGene/trassembly.html (accessed May 1, 2021).

Ellis, A. G., and Johnson, S. D. (2010). Floral mimicry enhances pollen export: the evolution of pollination by sexual deceit outside of the orchidaceae. *Am. Nat.* 176, E143–E151. doi: 10.1086/656487

Falara, V., Amarasinghe, R., Poldy, J., Pichersky, E., Barrow, R. A., and Peakall, R. (2013). The production of a key floral volatile is dependent on UV light in a sexually deceptive orchid. *Ann. Bot.* 111, 21–30. doi: 10.1093/aob/mcs228

Fan-Chiang, H.-J., and Wrolstad, R. E. (2006). Anthocyanin pigment composition of blackberries. *J. Food Sci.* 70, C198–C202. doi: 10.1111/j.1365-2621.2005.tb07125.x

Faraco, M., Spelt, C., Bliek, M., Verweij, W., Hoshino, A., Espen, L., et al. (2014). Hyperacidification of vacuoles by the combined action of two different P-ATPases in the tonoplast determines flower color. *Cell Rep.* 6, 32–43. doi: 10.1016/j.celrep.2013.12.009

Fossen, T., and Øvstedal, D. O. (2003). Anthocyanins from flowers of the orchids *Dracula chimaera* and *D. cordobae*. *Phytochemistry* 63, 783–787. doi: 10.1016/S0031-9422(03)00339-X

Francisco, R. M., Regalado, A., Ageorges, A., Burla, B. J., Bassin, B., Eisenach, C., et al. (2013). ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-glucosides. *Plant Cell* 25, 1840–1854. doi: 10.1105/tpc.112.102152

Franke, S., Ibarra, F., Schulz, C. M., Twele, R., Poldy, J., Barrow, R. A., et al. (2009). The discovery of 2,5-dialkylcyclohexan-1,3-diones as a new class of natural products. *Proc. Natl. Acad. Sci.* U. S. A. 106, 8877–8882. doi: 10.1073/pnas.0900646106

Fukada-Tanaka, S., Inagaki, Y., Yamaguchi, T., Saito, N., and Iida, S. (2000). Colour-enhancing protein in blue petals. *Nature* 407:581. doi: 10.1038/35036683

Gaskett, A. C. (2011). Orchid pollination by sexual deception: pollinator perspectives. *Biol. Rev.* 86, 33–75. doi: 10.1111/j.1469-185X.2010.00134.x

Gaskett, A. C., Endler, J. A., and Phillips, R. D. (2017). Convergent evolution of sexual deception via chromatic and achromatic contrast rather than colour mimicry. *Evol. Ecol.* 31, 205–227. doi: 10.1007/s10682-016-9863-2

George, A., Gonzales, C., Strauss, M. S., and Arditti, J. (1973). Chemotaxonomic and ecological implications of anthocyanins in *Elythranthera*. *Biochem. Syst. Ecol.* 1, 45–49. doi: 10.1016/0305-1978(73)90034-3

Glover, B. J., Walker, R. H., Moyroud, E., and Brockington, S. F. (2013). How to spot a flower. *New Phytol.* 197, 687–689. doi: 10.1111/nph.12112

Gomez, C., Terrier, N., Torregrosa, L., Vialet, S., Fournier-Level, A., Verries, C., et al. (2009). Grapevine MATE-type proteins act as vacuolar H+-dependent acylated anthocyanin transporters. *Plant Physiol.* 150, 402–415. doi: 10.1104/pp.109.135624

Grotewold, E. (2006). The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* 57, 761–780. doi: 10.1146/annurev.arplant.57.032905.105248

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

Hatier, J.-H. B., and Gould, K. S. (2007). Black coloration in leaves of *Ophiopogon planiscapus* "Nigrescens". Leaf optics, chromaticity, and internal light gradients. *Funct. Plant Biol.* 34:130. doi: 10.1071/FP06220

Hayashi, T., Bohman, B., Scaffidi, A., Peakall, R., and Flematti, G. R. (2021). An unusual tricosatriene is crucial for male fungus gnat attraction and exploitation by sexually deceptive *Pterostylis* orchids. *Curr. Biol.* 31, 1954.e–1961.e. doi: 10.1016/j.cub.2021.01.095

Hörtensteiner, S. (2006). Chlorophyll degradation during senescence. *Annu. Rev. Plant Biol.* 57, 55–77. doi: 10.1146/annurev.arplant.57.032905.105212

Hsu, C.-C., Chen, Y.-Y., Tsai, W.-C., Chen, W.-H., and Chen, H.-H. (2015). Three R2R3-MYB transcription factors regulate distinct floral pigmentation patterning in *Phalaenopsis* spp. *Plant Physiol.* 168, 175–191. doi: 10.1104/pp.114.254599

Hsu, H. F., Chen, W. H., Shen, Y. H., Hsu, W. H., Mao, W. T., and Yang, C. H. (2021). Multifunctional evolution of B and AGL6 MADS box genes in orchids. *Nat. Commun.* 12, 1–12. doi: 10.1038/s41467-021-21229-w

Hugueney, P., Provenzano, S., Verries, C., Ferrandino, A., Meudec, E., Batelli, G., et al. (2009). A novel cation-dependent O-methyltransferase involved in anthocyanin methylation in grapevine. *Plant Physiol.* 150, 2057–2070. doi: 10.1104/pp.109.140376

Iwashina, T. (2015). Contribution to flower colors of flavonoids including anthocyanins: a review. *Nat. Prod. Commun.* 10, 529–544. doi: 10.1177/1934578x1501000335

Johnson, S. D., and Schiestl, F. P. (2016). *Floral Mimicry*, 1st Edn. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198732693.001.0001

Jones, D. L. (2021). *A Complete Guide to Native Orchids of Australia*, 3rd Edn. Wahroonga, NSW: New Holland Publishers.

Kellenberger, R. T., Byers, K. J. R. P., De Brito Francisco, R. M., Staedler, Y. M., LaFountain, A. M., Schönenberger, J., et al. (2019). Emergence of a floral colour polymorphism by pollinator-mediated overdominance. *Nat. Commun.* 10, 1–11. doi: 10.1038/s41467-018-07936-x

Khunmuang, S., Kanlayanarat, S., Wongs-Aree, C., Meir, S., Philosoph-Hadas, S., Oren-Shamir, M., et al. (2019). Ethylene induces a rapid degradation of petal anthocyanins in cut *Vanda* 'Sansai Blue' orchid flowers. *Front. Plant Sci.* 10:1004. doi: 10.3389/fpls.2019.01004

Kitamura, S., Akita, Y., Ishizaka, H., Narumi, I., and Tanaka, A. (2012). Molecular characterization of an anthocyanin-related glutathione S-transferase gene in cyclamen. *J. Plant Physiol.* 169, 636–642. doi: 10.1016/j.jplph.2011.12.011

LaFountain, A. M., and Yuan, Y. W. (2021). Repressors of anthocyanin biosynthesis. *New Phytol.* 231, 933–949. doi: 10.1111/nph.17397

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, B. J., Zheng, B. Q., Wang, J. Y., Tsai, W. C., Lu, H. C., Zou, L. H., et al. (2020). New insight into the molecular mechanism of colour differentiation among floral segments in orchids. *Commun. Biol.* 3, 1–13. doi: 10.1038/s42003-020-0821-8

Li, Q., Wang, J., Sun, H. Y., and Shang, X. (2014). Flower color patterning in pansy (*Viola*×*wittrockiana* Gams.) is caused by the differential expression of three genes from the anthocyanin pathway in acyanic and cyanic flower areas. *Plant Physiol. Biochem.* 84, 134–141. doi: 10.1016/j.plaphy.2014.09.012

Liang, C. Y., Rengasamy, K. P., Huang, L. M., Hsu, C. C., Jeng, M. F., Chen, W. H., et al. (2020). Assessment of violet-blue color formation in *Phalaenopsis* orchids. *BMC Plant Biol.* 20:212. doi: 10.1186/s12870-020-02402-7

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

Lim, E. K., Ashford, D. A., Hou, B., Jackson, R. G., and Bowles, D. J. (2004). *Arabidopsis* glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnol. Bioeng.* 87, 623–631. doi: 10.1002/bit.20154

Liu, Y., Tikunov, Y., Schouten, R. E., Marcelis, L. F. M., Visser, R. G. F., and Bovy, A. (2018). Anthocyanin biosynthesis and degradation mechanisms in solanaceous vegetables: a review. *Front. Chem.* 6:52. doi: 10.3389/fchem.2018.00052

Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., et al. (2014). Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant Cell Environ.* 37, 1250–1258. doi: 10.1111/pce.12231

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

Lu, Z., Cao, H., Pan, L., Niu, L., Wei, B., Cui, G. C., et al. (2021). Two loss-of-function alleles of the glutathione S-transferase (GST) gene cause anthocyanin deficiency in flower and fruit skin of peach (*Prunus persica*). *Plant J.* 107, 1320–1331. doi: 10.1111/tpj.15312

Lücker, J., Martens, S., and Lund, S. T. (2010). Characterization of a *Vitis vinifera* cv. *Cabernet Sauvignon 3',5'-O*-methyltransferase showing strong preference for anthocyanins and glycosylated flavonols. *Phytochemistry* 71, 1474–1484. doi: 10.1016/j.phytochem.2010.05.027

Luo, J., Nishiyama, Y., Fuell, C., Taguchi, G., Elliott, K., Hill, L., et al. (2007). Convergent evolution in the BAHD family of acyl transferases: identification and characterization of anthocyanin acyl transferases from arabidopsis thaliana. *Plant J.* 50, 678–695. doi: 10.1111/j.1365-313X.2007.03079.x

Luo, P., Ning, G., Wang, Z., Shen, Y., Jin, H., Li, P., et al. (2016). Disequilibrium of flavonol synthase and dihydroflavonol-4-reductase expression associated tightly to white vs. red color flower formation in plants. *Front. Plant Sci.* 6, 1–12. doi: 10.3389/fpls.2015.01257

Markham, K. R., Bloor, S. J., Nicholson, R., Rivera, R., Shemluck, M., Kevan, P. G., et al. (2004). Black flower coloration in wild *Lisianthius nigrescens*: its chemistry and ecological consequences. zeitschrift fur naturforsch. - Sect. C *J. Biosci.* 59, 625–630. doi: 10.1515/znc-2004-9-1003

Martel, C., Cairampoma, L., Stauffer, F. W., and Ayasse, M. (2016). *Telipogon peruvianus* (Orchidaceae) flowers elicit pre-mating behaviour in *Eudejeania* (Tachinidae) males for pollination. *PLoS One* 11:e0165896. doi: 10.1371/journal.pone.0165896

Martel, C., Francke, W., and Ayasse, M. (2019). The chemical and visual bases of the pollination of the Neotropical sexually deceptive orchid *Telipogon peruvianus*. *New Phytol.* 223, 1989–2001. doi: 10.1111/nph.15902

Martins, T. R., Berg, J. J., Blinka, S., Rausher, M. D., and Baum, D. A. (2013). Precise spatio-temporal regulation of the anthocyanin biosynthetic pathway leads to petal spot formation in *Clarkia gracilis* (Onagraceae). *New Phytol.* 197, 958–969. doi: 10.1111/nph.12062

Martins, T. R., Jiang, P., and Rausher, M. D. (2017). How petals change their spots: cis-regulatory re-wiring in *Clarkia* (Onagraceae). *New Phytol.* 216, 510–518. doi: 10.1111/nph.14163

Mehrtens, F., Kranz, H., Bednarek, P., and Weisshaar, B. (2005). The arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 138, 1083–1096. doi: 10.1104/pp.104.058032

Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121. doi: 10.1093/nar/gkt263

Morita, Y., Ishiguro, K., Tanaka, Y., Iida, S., and Hoshino, A. (2015). Spontaneous mutations of the UDP-glucose:flavonoid 3-O-glucosyltransferase gene confers pale- and dull-colored flowers in the Japanese and common morning glories. *Planta* 242, 575–587. doi: 10.1007/s00425-015-2321-5

Morita, Y., Takagi, K., Fukuchi-Mizutani, M., Ishiguro, K., Tanaka, Y., Nitasaka, E., et al. (2014). A chalcone isomerase-like protein enhances flavonoid production and flower pigmentation. *Plant J.* 78, 294–304. doi: 10.1111/tpj.12469

Nadot, S., and Carrive, L. (2021). The colourful life of flowers. *Bot. Lett.* 168, 120–130. doi: 10.1080/23818107.2020.1839789

Ohmiya, A., Hirashima, M., Yagi, M., Tanase, K., and Yamamizo, C. (2014). Identification of genes associated with chlorophyll accumulation in flower petals. *PLoS One* 9:e113738. doi: 10.1371/journal.pone.0113738

Ohmiya, A., Sasaki, K., Nashima, K., Oda-Yamamizo, C., Hirashima, M., and Sumitomo, K. (2017). Transcriptome analysis in petals and leaves of chrysanthemums with different chlorophyll levels. *BMC Plant Biol.* 17:202. doi: 10.1186/s12870-017-1156-6

Ollerton, J., Winfree, R., and Tarrant, S. (2011). How many flowering plants are pollinated by animals? *Oikos* 120, 321–326. doi: 10.1111/j.1600-0706.2010.18644.x

Pan, Z. J., Chen, Y. Y., Du, J. S., Chen, Y. Y., Chung, M. C., Tsai, W. C., et al. (2014). Flower development of *Phalaenopsis* orchid involves functionally divergent SEPALLATA-like genes. *New Phytol.* 202, 1024–1042. doi: 10.1111/nph.12723

Peakall, R., Ebert, D., Poldy, J., Barrow, R. A., Francke, W., Bower, C. C., et al. (2010). Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually deceptive *Chiloglottis* orchids: implications for pollinator-driven speciation. *New Phytol.* 188, 437–450. doi: 10.1111/j.1469-8137.2010.03308.x

Peakall, R., Wong, D. C. J., Bohman, B., Flematti, G. R., and Pichersky, E. (2020). "Floral volatiles for pollinator attraction and speciation in sexually deceptive orchids," in *Biology of Plant*, eds E. Pichersky and N. Dudareva (Boca Raton, FL: CRC Press), 271–295. doi: 10.1201/9780429455612-18

Peakall, R., Wong, D. C. J., Phillips, R. D., Ruibal, M., Eyles, R., Rodriguez-Delgado, C., et al. (2021). A multitiered sequence capture strategy spanning broad evolutionary scales: application for phylogenetic and phylogeographic studies of orchids. *Mol. Ecol. Resour.* 21, 1118–1140. doi: 10.1111/1755-0998.13327

Pellegrino, G., Bellusci, F., and Palermo, A. M. (2017). Functional differentiation in pollination processes among floral traits in *Serapias* species (Orchidaceae). *Ecol. Evol.* 7, 7171–7177. doi: 10.1002/ece3.3264

Piñeiro Fernández, L., Byers, K. J. R. P., Cai, J., Sedeek, K. E. M., Kellenberger, R. T., Russo, A., et al. (2019). A phylogenomic analysis of the floral transcriptomes of sexually deceptive and rewarding European orchids. *Ophrys* and *Gymnadenia*. *Front. Plant Sci.* 10, 1–13. doi: 10.3389/fpls.2019.01553

Provenzano, S., Spelt, C., Hosokawa, S., Nakamura, N., Brugliera, F., Demelis, L., et al. (2014). Genetic control and evolution of anthocyanin methylation. *Plant Physiol.* 165, 962–977. doi: 10.1104/pp.113.234526

Raguso, R. A. (2004). Flowers as sensory billboards: progress towards an integrated understanding of floral advertisement. *Curr. Opin. Plant Biol.* 7, 434–440. doi: 10.1016/j.pbi.2004.05.010

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rodrigues, J. A., Espley, R. V., and Allan, A. C. (2021). Genomic analysis uncovers functional variation in the C-terminus of anthocyanin-activating MYB transcription factors. *Hortic. Res.* 8, 1–14. doi: 10.1038/s41438-021-00514-1

Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., et al. (2012). RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry* 82, 38–45. doi: 10.1016/j.phytochem.2012.07.007

Schiestl, F. P., Ayasse, M., Paulus, H. F., Löfstedt, C., Hansson, B. S., Ibarra, F., et al. (1999). Orchid pollination by sexual swindle. *Nature* 399, 421–421. doi: 10.1038/20829

Schiestl, F. P., Peakall, R., Mant, J. G., Ibarra, F., Schulz, C., Franke, S., et al. (2003). The chemistry of sexual deception in an orchid-wasp pollination system. *Science* 302, 437–438. doi: 10.1126/science.1087835

Sedeek, K. E. M., Qi, W., Schauer, M. A., Gupta, A. K., Poveda, L., Xu, S., et al. (2013). Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. *PLoS One* 8:e64621. doi: 10.1371/journal.pone.0064621

Shan, X., Li, Y., Yang, S., Yang, Z., Qiu, M., Gao, R., et al. (2020). The spatio-temporal biosynthesis of floral flavonols is controlled by differential phylogenetic MYB regulators in Freesia hybrida. *New Phytol.* 228, 1864–1879. doi: 10.1111/nph.16818

Sheehan, H., Moser, M., Klahre, U., Esfeld, K., Dell'olivo, A., Mandel, T., et al. (2016). MYB-FL controls gain and loss of floral UV absorbance, a key trait affecting pollinator preference and reproductive isolation. *Nat. Genet.* 48, 159–166. doi: 10.1038/ng.3462

Singer, R. B., Flach, A., Koehler, S., Marsaioli, A. J., and Amaral, M. D. C. E. (2004). Sexual mimicry in *Mormolyca ringens* (Lindl.) Schltr. (Orchidaceae: Maxillariinae). *Ann. Bot.* 93, 755–762. doi: 10.1093/aob/mch091

Strack, D., Busch, E., and Klein, E. (1989). Anthocyanin patterns in european orchids and their taxonomic and phylogenetic relevance. *Phytochemistry* 28, 2127–2139. doi: 10.1016/S0031-9422(00)97931-7

Stracke, R., Ishihara, H., Huep, G., Barsch, A., Mehrtens, F., Niehaus, K., et al. (2007). Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* 50, 660–677. doi: 10.1111/j.1365-313X.2007.03078.x

Streisfeld, M. A., and Rausher, M. D. (2011). Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. *Evolution* 65, 629–642. doi: 10.1111/j.1558-5646.2010.01165.x

Sun, J., Lin, L. Z., and Chen, P. (2012a). Study of the mass spectrometric behaviors of anthocyanins in negative ionization mode and its applications for characterization of anthocyanins and non-anthocyanin polyphenols. *Rapid Commun. Mass Spectrom.* 26, 1123–1133. doi: 10.1002/rcm.6209

Sun, Y., Li, H., and Huang, J. R. (2012b). Arabidopsis TT19 functions as a carrier to transport anthocyanin from the cytosol to tonoplasts. *Mol. Plant* 5, 387–400. doi: 10.1093/mp/ssr110

Suzuki, H., Nakayama, T., Yonekura-Sakakibara, K., Fukui, Y., Nakamura, N., Yamaguchi, M. A., et al. (2002). cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme a:anthocyanidin 3-O-glucoside-6′-O-malonyltransferase from dahlia flowers. *Plant Physiol.* 130, 2142–2151. doi: 10.1104/pp.010447

Tanaka, A., and Tanaka, R. (2006). Chlorophyll metabolism. *Curr. Opin. Plant Biol.* 9, 248–255. doi: 10.1016/j.pbi.2006.03.011

Tanaka, Y., and Brugliera, F. (2013). Flower colour and cytochromes P450. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368:20120432. doi: 10.1098/rstb.2012.0432

Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-313X.2008.03447.x

Thill, J., Miosic, S., Ahmed, R., Schlangen, K., Muster, G., Stich, K., et al. (2012). Le rouge et le noir": a decline in flavone formation correlates with the rare color of black dahlia (Dahlia variabilis hort.) flowers. *BMC Plant Biol.* 12:225. doi: 10.1186/1471-2229-12-225

Thomas, M. M., Rudall, P. J., Ellis, A. G., Savolainen, V., and Glover, B. J. (2009). Development of a complex floral trait: the pollinator-attracting petal spots of the beetle daisy. *Gorteria diffusa* (asteraceae). *Am. J. Bot.* 96, 2184–2196. doi: 10.3732/ajb.0900079

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

Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J. I., Awazuhara, M., et al. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of arabidopsis plants over-expressing an MYB transcription factor. *Plant J.* 42, 218–235. doi: 10.1111/j.1365-313X.2005.02371.x

Vereecken, N. J., Wilson, C. A., Hotling, S., Schulz, S., Banketov, S. A., and Mardulyn, P. (2012). Pre-adaptations and the evolution of pollination by sexual deception: cope's rule of specialization revisited. *Proc. R. Soc. B Biol. Sci* 279, 4786–4794. doi: 10.1098/rspb.2012.1804

Verweij, W., Spelt, C., Di Sansebastiano, G. P., Pietro, Vermeer, J., Reale, L., et al. (2008). An H+ P-ATPase on the tonoplast determines vacuolar pH and flower colour. *Nat. Cell Biol.* 10, 1456–1462. doi: 10.1038/ncb1805

Vignolini, S., Davey, M. P., Bateman, R. M., Rudall, P. J., Moyroud, E., Tratt, J., et al. (2012). The mirror crack'd: both pigment and structure contribute to the glossy blue appearance of the mirror orchid. *Ophrys speculum*. *New Phytol.* 196, 1038–1047. doi: 10.1111/j.1469-8137.2012.04356.x

Waki, T., Mameda, R., Nakano, T., Yamada, S., Terashita, M., Ito, K., et al. (2020). A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nat. Commun.* 11, 1–14. doi: 10.1038/s41467-020-14558-9

Wang, L., Albert, N. W., Zhang, H., Arathoon, S., Boase, M. R., Ngo, H., et al. (2014). Temporal and spatial regulation of anthocyanin biosynthesis provide diverse flower colour intensities and patterning in *Cymbidium orchid*. *Planta* 240, 983–1002. doi: 10.1007/s00425-014-2152-9

Wessinger, C. A., and Rausher, M. D. (2012). Lessons from flower colour evolution on targets of selection. *J. Exp. Bot.* 63, 5741–5749. doi: 10.1093/jxb/ers267

Weston, P. H., Perkins, A. J., Indsto, J. O., and Clements, M. A. (2014). "Phylogeny of Orchidaceae tribe Diurideae and its implications for the evolution of pollination systems," in *Darwin's Orchids: Then and Now*, eds P. Bernhardt and R. Edens-Meier (Chicago, IL: University of Chicago Press), 91–154.

Wong, D. C. J., Amarasinghe, R., Falara, V., Pichersky, E., and Peakall, R. (2019). Duplication and selection in β-ketoacyl-ACP synthase gene lineages in the sexually deceptive *Chiloglottis* (Orchidaceae). *Ann. Bot.* 123, 1053–1066. doi: 10.1093/aob/mcz013

Wong, D. C. J., Amarasinghe, R., Pichersky, E., and Peakall, R. (2018). Evidence for the involvement of fatty acid biosynthesis and degradation in the formation of insect sex pheromone-mimicking chiloglottones in sexually deceptive *Chiloglottis* Orchids. *Front. Plant Sci.* 9:389. doi: 10.3389/fpls.2018.00839

Wong, D. C. J., Amarasinghe, R., Rodriguez-Delgado, C., Eyles, R., Pichersky, E., and Peakall, R. (2017a). Tissue-specific floral transcriptome analysis of the sexually deceptive orchid *Chiloglottis trapeziformis* provides insights into the biosynthesis and regulation of its unique UV-B dependent floral volatile, chiloglottone 1. *Front. Plant Sci.* 8:1260. doi: 10.3389/fpls.2017.01260

Wong, D. C. J., Pichersky, E., and Peakall, R. (2017b). The biosynthesis of unusual floral volatiles and blends involved in orchid pollination by deception: current progress and future prospects. *Front. Plant Sci.* 8:1955. doi: 10.3389/fpls.2017.01955

Wong, D. C. J., Schlechter, R., Vannozzi, A., Holl, J., Hmmam, I., Bogs, J., et al. (2016). A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. *DNA Res.* 23, 451–466. doi: 10.1093/dnares/dsw028

Yang, X., Xia, X., Zhang, Z., Nong, B., Zeng, Y., Wu, Y., et al. (2019). Identification of anthocyanin biosynthesis genes in rice pericarp using PCAMP. *Plant Biotechnol. J.* 17, 1700–1702. doi: 10.1111/pbi.13133

Yoshida, K., Toyama-Kato, Y., Kameda, K., and Kondo, T. (2003). Sepal color variation of *Hydrangea macrophylla* and vacuolar pH measured with a proton-selective microelectrode. *Plant Cell Physiol.* 44, 262–268. doi: 10.1093/pcp/pcg033

Yuan, Y. W., Rebocho, A. B., Sagawa, J. M., Stanley, L. E., and Bradshaw, H. D. (2016). Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2448–2453. doi: 10.1073/pnas.1515294113

Zhang, Y., Cheng, Y., Ya, H., Xu, S., and Han, J. (2015). Transcriptome sequencing of purple petal spot region in tree peony reveals differentially expressed anthocyanin structural genes. *Front. Plant Sci.* 6:964. doi: 10.3389/fpls.2015.00964

Zhang, Y., Zhou, T., Dai, Z., Dai, X., Li, W., Cao, M., et al. (2020). Comparative transcriptomics provides insight into floral color polymorphism in a *Pleione limprichtii* orchid population. *Int. J. Mol. Sci* 21, 247. doi: 10.3390/ijms21010247

Zhao, D., and Tao, J. (2015). Recent advances on the development and regulation of flower color in ornamental plants. *Front. Plant Sci.* 6:261. doi: 10.3389/fpls.2015.00261

Zhao, Y., Dong, W., Zhu, Y., Allan, A. C., Lin-Wang, K., and Xu, C. (2020). PpGST1, an anthocyanin-related glutathione S-transferase gene, is essential for fruit coloration in peach. *Plant Biotechnol. J.* 18, 1284–1295. doi: 10.1111/pbi.13291

Zheng, J., Wu, H., Zhu, H., Huang, C., Liu, C., Chang, Y., et al. (2019). Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. *New Phytol.* 223, 705–721. doi: 10.1111/nph.15807

Zhong, C., Tang, Y., Pang, B., Li, X., Yang, Y., Deng, J., et al. (2020). The R2R3-MYB transcription factor GhMYB1a regulates flavonol and anthocyanin accumulation in *Gerbera* hybrida. *Hortic. Res.* 7:78. doi: 10.1038/s41438-020-0296-2

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fpls.2022.916081

Characterization of Three *SEPALLATA*-Like MADS-Box Genes Associated With Floral Development in *Paphiopedilum henryanum* (Orchidaceae)

Hao Cheng[1,2†], Xiulan Xie[2†], Maozhi Ren[2], Shuhua Yang[1], Xin Zhao[1], Nasser Mahna[3], Yi Liu[2], Yufeng Xu[1], Yukai Xiang[4], Hua Chai[4], Liang Zheng[4], Hong Ge[1] and Ruidong Jia[1*]*

[1]Key Laboratory of Biology and Genetic Improvement of Flower Crops (North China), Ministry of Agriculture and Rural Affairs, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, [2]National Agricultural Science & Technology Center, Institute of Urban Agriculture, Chinese Academy of Agricultural Sciences, Chengdu, China, [3]Department of Horticultural Sciences, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, [4]Department of High-Performance Computing, National Supercomputing Center in Chengdu, Chengdu, China

OPEN ACCESS

Edited by:
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

Reviewed by:
Pablo Bolaños-Villegas,
University of Costa Rica, Costa Rica
Sagheer Ahmad,
Guangdong Academy of Agricultural Sciences, China
Pandiyan Muthuramalingam,
Gyeongsang National University, South Korea

***Correspondence:**
Hong Ge
gehong@caas.cn
Ruidong Jia
jiaruidong@caas.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: 08 April 2022
Accepted: 03 May 2022
Published: 26 May 2022

Citation:
Cheng H, Xie X, Ren M, Yang S, Zhao X, Mahna N, Liu Y, Xu Y, Xiang Y, Chai H, Zheng L, Ge H and Jia R (2022) Characterization of Three SEPALLATA-Like MADS-Box Genes Associated With Floral Development in Paphiopedilum henryanum (Orchidaceae). Front. Plant Sci. 13:916081. doi: 10.3389/fpls.2022.916081

Paphiopedilum (Orchidaceae) is one of the world's most popular orchids that is found in tropical and subtropical forests and has an enormous ornamental value. *SEPALLATA*-like (*SEP*-like) MADS-box genes are responsible for floral organ specification. In this study, three *SEP*-like MADS-box genes, *PhSEP1*, *PhSEP2*, and *PhSEP3*, were identified in *Paphiopedilum henryanum*. These genes were 732–916 bp, with conserved SEPI and SEPII motifs. Phylogenetic analysis revealed that *PhSEP* genes were evolutionarily closer to the core eudicot *SEP3* lineage, whereas none of them belonged to core eudicot *SEP1/2/4* clades. *PhSEP* genes displayed non-ubiquitous expression, which was detectable across all floral organs at all developmental stages of the flower buds. Furthermore, subcellular localization experiments revealed the localization of PhSEP proteins in the nucleus. Yeast two-hybrid assays revealed no self-activation of PhSEPs. The protein–protein interactions revealed that PhSEPs possibly interact with B-class DEFICIENS-like and E-class MADS-box proteins. Our study suggests that the three *SEP*-like genes may play key roles in flower development in *P. henryanum*, which will improve our understanding of the roles of the *SEP*-like MADS-box gene family and provide crucial insights into the mechanisms underlying floral development in orchids.

Keywords: expression analysis, flower development, gene cloning, *Paphiopedilum*, *SEPALLATA*-like MADS-box genes

INTRODUCTION

Paphiopedilum Pfitzer (Orchidaceae), commonly known as "slipper orchid," is one of the world's most popular orchids in the Orchidaceae family, owing to its remarkable diversity in terms of the shape, size, and color of flowers (Ng and Mohd Saleh, 2011; Zeng et al., 2013; Guo et al., 2021). This orchid can be mainly found in tropical and subtropical forests extending from Asia to the Pacific Islands. More than 18 species are widely distributed across Southwest

China (Guan et al., 2011). *Paphiopedilum henryanum*, a species threatened with extinction, mainly occurs in the crevices of shady cliffs or rocks and well-drained habitats of the mountains along the Sino-Vietnamese border (Xu et al., 2018). The perianth of the *Paphiopedilum* flower consists of two petal-like sepals (whorl I), two lateral petals, and a highly diversified lip (whorl II). The inner fertile organ is adapted to represent gynostemium (whorl III; Pi et al., 2009). The reproductive organ of this ornamental plant is highly diversified and thus may serve as models for studying the molecular development of flowers in monocots.

Flower formation is known to be controlled by different regulatory genes, including several MADS-box family members (Ma et al., 2019). These MADS-box genes can be divided into two lineages, Type I and Type II, originating from a single-gene duplication that occurred before the divergence of plants and animals (Alvarez-Buylla et al., 2000). MADS-box proteins contain a highly conserved motif of 55–60 amino acids known as the MADS domain, which is essential for DNA-binding activity (De Bodt et al., 2003). As important transcriptional factors, MADS-box genes participate in various plant developmental processes, including the regulation of floral organ identity, inflorescence meristem identity, fruit ripening, and several other processes (Goto and Meyerowitz, 1994; Mandel and Yanofsky, 1995; Liljegren et al., 2000; Guo et al., 2017).

The developmental pathways for determining floral organ identity have been well-studied in several eudicot model species, such as *Arabidopsis thaliana* and *Antirrhinum majus* (Schwarz-Sommer et al., 1990; Coen and Meyerowitz, 1991). The ABCDE model of floral development was established as unifying paradigm and underlying principle of flower development and evolution. This model comprises five major classes of homeotic genes: A, B, C, D, and E. Except *APETALA2* (*AP2*), all of these genes belong to MADS-box genes (Theißen, 2001; Mondragõn-Palomino and Theißen, 2011). According to this model, the expression of A- and E-class genes leads to the development of sepals; the expression of A-, B-, and E-class genes give rise to petals; the expression of B-, C-, and E-class genes in the meristematic regions allows the development of stamens; carpels are formed when the C- and E-class genes are expressed; and ovules develop when the D- and E-class genes are expressed (Ditta et al., 2004; Theissen and Melzer, 2007; Pu and Xu, 2021).

SEPALLATA (*SEP*) are E-class MADS-box genes that act as important mediators of the higher-order complex and participate in various aspects of plant development together with B-, C-, and D-class MADS-box genes (Becker and Theißen, 2003; Immink et al., 2009; Pu et al., 2020). *SEP* genes have undergone two gene duplications during their evolution; the first duplication preceded the origin of the extant angiosperms, resulting in two clades, *AGL2/3/4* (*SEP1/2*) and *AGL9* (*SEP3*). Subsequent duplications have occurred independently within these clades after the divergence of eudicots and monocots (Shan et al., 2009). As for eudicots, *SEP* genes have been reported in tomatoes (*Solanum lycopersicum*), petunias (*Petunia hybrida*), and orchids (Ferrario et al., 2003; Uimari et al., 2004). Moreover, members of the SEP family have been identified in monocots, such as maize and rice (Becker and Theißen, 2003; Cui et al., 2010). In *Arabidopsis*, four *SEP* genes (*AtSEP1*, *AtSEP2*, *AtSEP3*, and *AtSEP4*) play a role in the development of all floral whorl and meristem determinacy (Ditta et al., 2004).

In orchids, the function of some MADS-box genes has been reported, and a specific model was established (Mondragón-Palomino, 2013). According to the "Homeotic Orchid Tepal" (HOT) model, the B-class genes in combination with genes of other classes, such as the E-class genes, regulate the complexity of sepal, petal, and lip identity (Pan et al., 2011). However, few *SEP*-like genes have been identified in orchid species, such as *AdOM1* in Aranda, *DcSEP1* in *Dendrobium crumenatum*, *DOMADS1* and *DOMADS3* in *Dendrobium grex Madame* Thong-IN, and *PeSEP1/2/3/4* in *Phalaenopsis equestris* (Lu et al., 1993; Xu et al., 2006; Pan et al., 2014). In Dendrobium, DcOSEP1/DcOPI/DcOAP3A or DcOAP3B (SEP-like/PI-like/AP3-like) could form multimeric proteins (Hsu et al., 2015). Functional analysis showed that virus-induced silencing of *PeSEP3* in *P. equestris* could alter the epidermal identity of tepals and the contents of anthocyanin and chlorophyll, causing tepals to become leaf-like organs (Pan et al., 2014). Agreeing well with research from Phalaenopsis, defects of *CeSEP1/3*-clade genes of the Chinese orchid *Cymbidium ensifolium* contributed to the leaf-like flower phenotype in the mutant, indicating that SEP paralogs differed in their ability to regulate floral organ specificity (Wei et al., 2020). Interestingly, *CeSEP-2* is important for the development of a specialized lip in Cymbidium orchids, while its downregulation resulted in the formation of a peloric flower shape in *C. ensifolium* (Ai et al., 2021). Further study revealed that the E-class MADS-box protein PeMADS8 in *P. equestris* could also interact with the B_{sister} protein PeMADS28, and a higher-order protein complex formed by C-E-D-Bsister genes (PeMADS1-PeMADS8-PeMADS-PeMADS28) was likely to be associated with regulation of orchid ovule development (Shen et al., 2021). In *Habenaria radiata*, the *SEP*-like gene *HrSEP1* plays an important role in column, lip, and petal development. The mutation in this gene can cause the greenish flower phenotype of *H. radiata* (Mitoma and Kanno, 2018). In contrast with the four genes found in other orchids, only two SEP transcripts were expressed in the inflorescence of *Orchis italica*, and both genes were detectable in all floral organs, which was consistent with the expression pattern in all the floral whorls of class E genes involved in the formation of all the organs of the flower (Valoroso et al., 2019). Moreover, *SEP*-like genes were also involved in orchid fruit development. In *Erycina pusilla*, the *SEP-like* genes *EpMADS8* and− *9* were expressed throughout fruit development, and protein–protein interaction studies revealed that MADS domain complexes comprised of SEP, FRUITFULL (FUL), and AGAMOUS (AG)/SHATTERPROOF (SHP) orthologs can also be formed in *E. pusilla* (Lin et al., 2016; Dirks-Mulder et al., 2019). To date, no *SEP* genes have been identified in *Paphiopedilum*. In addition, the flowers of *Paphiopedilum* contain a pocket-like lip and synsepal, distinguishing them from those of other orchids. Therefore, it is necessary to isolate and characterize the E-class genes of *Paphiopedilum* and address their developmental role in perianth identity.

In the present study, three *SEP*-like MADS-box genes were isolated from *P. henryanum*. The sequences of these genes and their encoded proteins were analyzed. Moreover, the expression

patterns of *PhSEP* genes were explored using quantitative real-time PCR (qRT-PCR) performed on different tissues and organs and across various floral bud developmental stages. The subcellular localization and self-activation of the corresponding proteins were also investigated using the gene gun-mediated transformation and yeast two-hybrid system, respectively. In addition, the identification of PhSEPs' binding sites was predicted by the DeepMind's AlphaFold2 program. Thus, the present study aimed to establish a foundation for further studies by elaborating the molecular mechanisms of floral organ determination in slipper orchids.

MATERIALS AND METHODS

Plant Material and Bacterial Strains

The *P. henryanum* plants used in this study were grown in a greenhouse at the Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences. Floral buds representing developmental stages B1–B4 (stage B1: 2.0–3.0 mm in length; stage B2: 3.0–4.0 mm in length; stage B3: 4.0–5.0 mm in length; stage B4: 5.0–6.0 mm in length), various floral organs of mature flowers (sepal, petal, lip, ovary, gynostemium, and bract), scape, roots, and leaves were collected, immediately frozen in liquid nitrogen, and then stored at −80°C. Plasmid pEASY®-T3 (Takara, Japan) was used to clone the cDNA sequences, whereas the pBI221-EGFP and pGBKT7 (Clontech, United States) vectors were modified to clone overexpression constructs. *Escherichia coli* DH5α and *Saccharomyces cerevisiae* AH109 were used for transformation and self-activation, respectively.

Cloning and Characterization of *PhSEP* Genes From *Paphiopedilum henryanum*

To identify, clone, and characterize the *SEP*-like genes of *P. henryanum*, total RNA was extracted from the harvested tissues using the RNAprep Pure Plant Kit (TIANGEN Biotech Co., Ltd., Beijing, China), according to the manufacturer's instructions. Reverse transcription was carried out with 1.0 μg of each RNA sample using the FastQuant RT Kit (with gDNase; TIANGEN Biotech Co., Ltd., Beijing, China). Three *SEP*-like genes were identified from the transcriptome of *P. henryanum* (Accession nos. SRP131426 and PRJNA431671, available at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) database). Based on the sequences retrieved, specific primer pairs (PhSEP1-F/PhSEP1-R, PhSEP2-F/PhSEP2-R, and PhSEP3-F/PhSEP3-R; **Supplementary Table S1**) were designed for cloning the coding sequences (CDS) of *PhSEP* genes. The amount of 1–2 μl of the synthesized cDNA (100 ng/μL) was used for PCR with primers and the high-fidelity Taq DNA polymerase (Ex Taq, TaKaRa Bio, Japan). The amplified products were evaluated by agarose gel electrophoresis and then cloned into the pEASY®-T3 vector. The recombinant clones were selected for amplification and identification. The nucleic acid sequences obtained were then compared with the homologous gene sequences retrieved from GenBank using Blastn. Open reading frame (ORF) search was performed using the online server getorf;[1] the molecular weights and isoelectric points of the predicted proteins were analyzed using ProtParam,[2] whereas their hydrophilicity was assessed using ProtScale;[3] and the amino acid signal peptides and subcellular localization were predicted using SignalP3.0[4] and PSORT,[5] respectively.

Multiple Sequence Alignment and Phylogenetic Analysis

SEP-like genes and *AP1/SQUA*-like genes were retrieved from previously published studies and other publicly available databases using BLAST searches (Pan et al., 2014). During the BLAST searches, multiple genes of the each subfamily from different lineages were used as queries. The following databases were used in the search: NCBI. Each of the databases was searched using TBLASTN. We obtained the sequences whose E-values were below le–5 and redundant sequences with identity of at least 95% were removed from our data set (**Supplementary Table S2**). Protein sequences were first aligned with MEGA11.[6] Sequences of the alignment were ordered according to their phylogenetic placements in the preliminary tree, then, they were aligned manually using MEGA11and DNAMAN version 4.0 (Lynnon Biosoft Company; Kumar et al., 2016).

Phylogenetic analyses about *SEP3* were conducted using DNA alignments that included the conserved M-, I-, and K-domain regions and the C-terminal residues with higher than 12 quality scores. The quality score for each residue was estimated in CLUSTALX 2.1(Thompson et al., 1997). The PhyML software was used to construct ML tree with the most appropriate model, GTR + I + C, which was estimated by running MODELTEST version 3.06 and 1,000 bootstrap replicates (Posada and Crandall, 1998; Guindon and Gascuel, 2003). Bootstrapping was performed by resampling the data 1,000 times. Tree files were viewed using iTOL (Letunic and Bork, 2021).

Gene Expression Analysis *via* qRT-PCR

To investigate the spatio-temporal expression patterns of the *SEP* genes, a quantitative reverse transcription PCR (qRT-PCR) was performed as described previously (Pan et al., 2014; Omondi et al., 2015; Shen et al., 2021) using tissues of roots, stems, leaves, floral organs of mature flowers, and developing floral buds at different stages. Gene-specific primers (qSEP1-F/qSEP1-R, qSEP2-F/qSEP2-R, and qSEP3-F/qSEP3-R) were designed within the non-conserved C-terminal region for each gene (**Supplementary Table S1**) using the Primer 5 software. The expected size of amplification products was 100–150 bp. TB Green® Premix Ex Taq™ II (Tli RNaseH Plus; TaKaRa, Japan) was used for transcript quantification. The cycling program was as follows: an initial denaturation step at 95°C for 30 s, followed by 40 cycles of denaturation at 95°C (5 s), annealing at 60°C (30 s), and extension at 72°C (30 s). The relative mRNA

[1]http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf
[2]https://web.expasy.org/protparam/
[3]https://web.expasy.org/protscale/
[4]http://www.cbs.dtu.dk/services/SignalP-3.0/
[5]https://www.genscript.com/psort.html
[6]https://www.megasoftware.net/

abundance of the *SEP* genes and the reference gene, *Actin*, was analyzed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Three independent biological replicates and three technical replicates were used for each experimental or control sample.

Subcellular Localization of *PhSEP* Genes

To examine the subcellular localization of SEP genes, a gene gun was used for introducing DNA into onion (*Allium cepa*) inner epidermal cells as described previously by Wang et al. (2016) and modified by Li et al. (2020). To do this, the coding regions of *PhSEP* genes were amplified using KAPA HiFi™ HotStart DNA polymerase (KAPA Biosystems, United States), with the primers (DWSEP1-F/DWSEP1-R, DWSEP2-F/DWSEP2-R, DWSEP3-F/DWSEP3-R) listed in **Supplementary Table S1**. After purification, the PCR products were cloned downstream of the synthetic green fluorescent protein (EGFP) reporter gene in the pBI221-EGFP binary vector using the SE Seamless Cloning and Assembly Kit (Zomanbio, Beijing, China). The recombinant vector harboring the *SEP* fusion and the negative control (empty pBI221-EGFP vector) were used to transform living onion epidermal cells by biolistic bombardment using a Biolistc® PDS-1000/He Particle Delivery System (Bio-Rad Laboratories, CA, United States) according to the manufacturer's instructions (helium pressure, 9 MPa; Yu and Goh, 2000). Fluorescence was observed using a fluorescence microscope (BX53 Upright Microscope, Olympus, Tokyo, Japan).

Yeast Assay and Protein–Protein Interactions Prediction

The yeast two-hybrid assay is a powerful and classic method of screening protein–protein interactions (Hu et al., 2021). To screen protein–protein interactions, a yeast two-hybrid assay was performed as described by Dirks-Mulder et al. (2019). The CDS of *SEP* genes were cloned in-frame downstream of the GAL4-binding domain of the pGBKT7 vector (Clontech, USA) after amplification with the primers (BDSEP1-F/BDSEP1-R, BDSEP2-F/BDSEP2-R, BDSEP3-F/BDSEP3-R) listed in **Supplementary Table S1**. The constructs were prepared using the SE Seamless Cloning and Assembly. The recombinant plasmids and the negative control (empty pGBKT7 vector) were used to transform *S. cerevisiae* AH109 competent cells according to the Yeast Protocols Handbook (Clontech, United States; De Folter and Immink, 2011). The cultures were serially diluted at a ratio of 1:10. Thereafter, 2 µl aliquots of the undiluted, 1:10, and 1:100 diluted cell cultures were spotted onto a non-selective medium, that is, the synthetic dropout medium without leucine and tryptophan (SD-LW), and selective media, including SD-LWH +3-AT (SD–leucine–tryptophan–histidine+5 mM 3-AT) and SD-LWHA (SD–leucine–tryptophan–histidine–adenine). The respective plates were incubated at 30°C for 7 days before being photographed. The self-activation of each protein was evaluated for its host status (Yeast Protocols Handbook; Clontech). AlphaFold2 (AF2) is a protein structure prediction model developed by DeepMind, which can predict the protein–protein complex structures and interaction accurately (Pozzati et al., 2021; Bryant et al., 2022). To predict the interaction ability of SEP proteins, a poly-glycine linker was added between each chain before running it as a single chain through the AlphaFold model (Humphreys et al., 2021; Tsaban et al., 2022). Molecular modeling was performed using the PyMOL molecular viewer for visualizing hydrogen bond interactions.

RESULTS

Identification of *SEP* Genes From *Paphiopedilum henryanum* and Sequence Analysis

Three *SEP*-like genes were isolated from *P. henryanum* and named *PhSEP1*, *PhSEP2*, and *PhSEP3* (GenBank accession nos. MN274961, MN274962, and MN809620, respectively; **Figure 1**). *PhSEP1* was 916 bp in length and contained an ORF of 732 bp. *PhSEP2* was 839 bp in length and contained an ORF of 726 bp, whereas *PhSEP3* was 889 bp in length with a 732 bp ORF. *PhSEP1* shared 86% identity with its *P. equestris* homolog, *PeSEP1*. *PhSEP2* and *PhSEP3* independently showed 81% identity with *PeSEP2* and *PeSEP1*, respectively. The predicted proteins showed length of 241 (PhSEP2) and 243 (PhSEP1/3) amino acids and pI ranging between 8.71 (PhSEP2) and 8.94 (PhSEP1). Further bioinformatics analysis showed that these PhSEP proteins displayed a theoretical molecular mass of 28 kDa and harbored a nuclear localization signal. In addition, these proteins lacked transmembrane domains. Multiple sequence alignment with homologous SEP proteins from orchids (**Figure 2**) indicated that PhSEPs harbored a conserved MIK domain and divergent C-terminal domain with conserved SEP I and SEP II motifs, which are characteristic of E-class MADS-box proteins.

Phylogenetic Analysis of *SEP*-Like Genes

To determine the evolutionary relationships of the SEP subfamily within orchids and with other angiosperms, we constructed a phylogenetic tree using nucleic acid sequences. Totally, 35 *SEP* genes from orchids, other monocots, asterids, and rosids were obtained (**Supplementary Table S2**). We then performed phylogenetic analyses on the nucleotide sequences of these genes using Maximum Likelihood (ML), with 3 *AP1/SQUA*-like genes as the outgroups. The phylogenetic tree showed that *SEP* genes from monocots formed a well-supported single clade (**Figure 3**). They were divided into two major clades, *M1* and M2, with strong supporting values. The clade *M1* (including *PhSEP1* and *PhSEP3*) was clustered together with the *SEP3* genes of eudicots, leaving the clade *M2* (including *PhSEP2*) alone (well-supported), suggesting a duplicate event before the separation of monocots and eudicots. Monocots *SEPs* selected were not included in core eudicot *SEP1/2/4*, differently from previous study (Pan et al., 2014). A neighbor-joining based phylogenetic tree of 50 *SEP*-like genes showed that *PhSEP1* and *PhSEP3* were included in the *SEP3* clade while *PhSEP2* was grouped into the *SEP1/2/4* clade (**Supplementary Figure S2**).

Expression Analysis of *PhSEP* Genes

qRT-PCR was performed to determine the spatio-temporal expression pattern of *PhSEP* genes across different tissues and

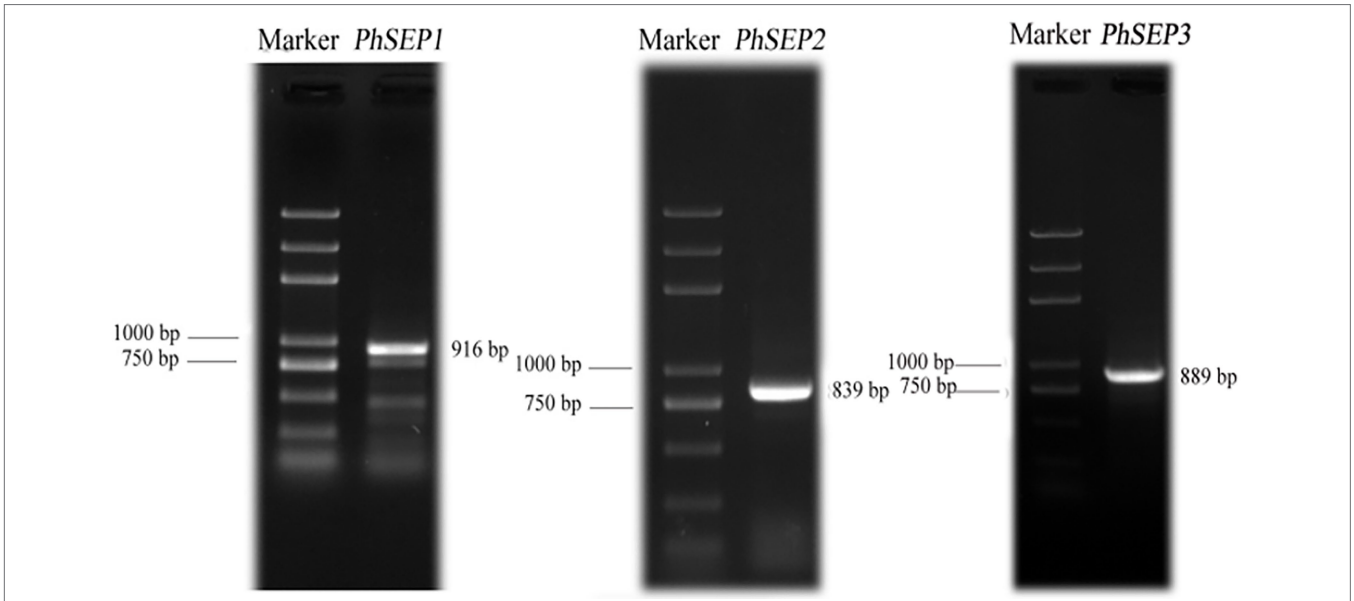


FIGURE 1 | Amplification of three *PhSEP* genes from *Paphiopedilum henryanum*.

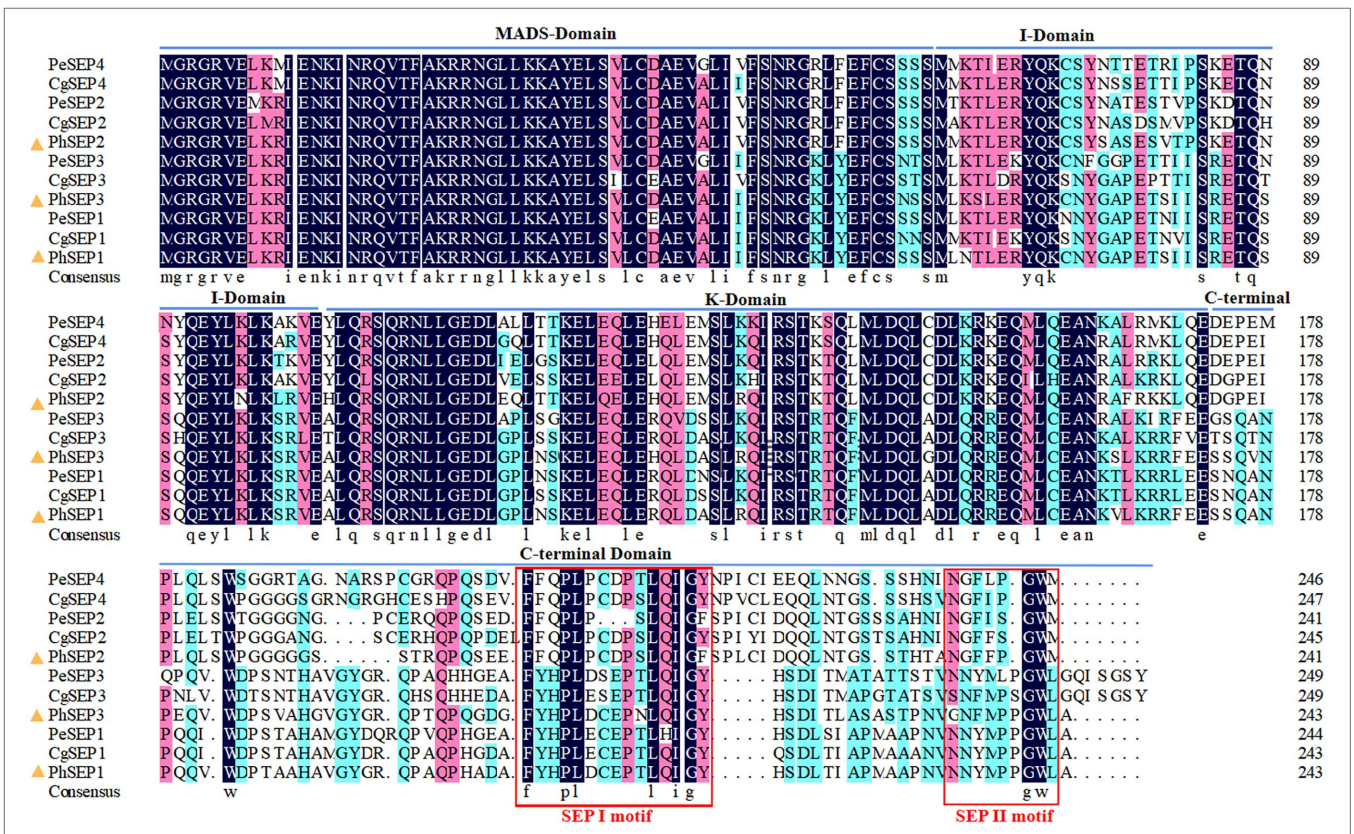


FIGURE 2 | Amino acid sequence alignment of PhSEPs and closely related homologs in orchids using the MEGA11 and DNAMAN version 4.0. The PhSEP1/2/3 proteins of *P. henryanum* are highlighted by yellow triangles.

organs of *P. henryanum* (**Figures 4A–D**). As shown in **Figure 4E**, the expression of *PhSEP1* was specific to reproductive tissue and was especially high in the gynostemium and synsepal. In contrast, its expression was negligible in vegetative tissues, including the scape, roots, and leaves. *PhSEP2* was expressed in all reproductive tissues and the scape. However, it displayed negligible expression

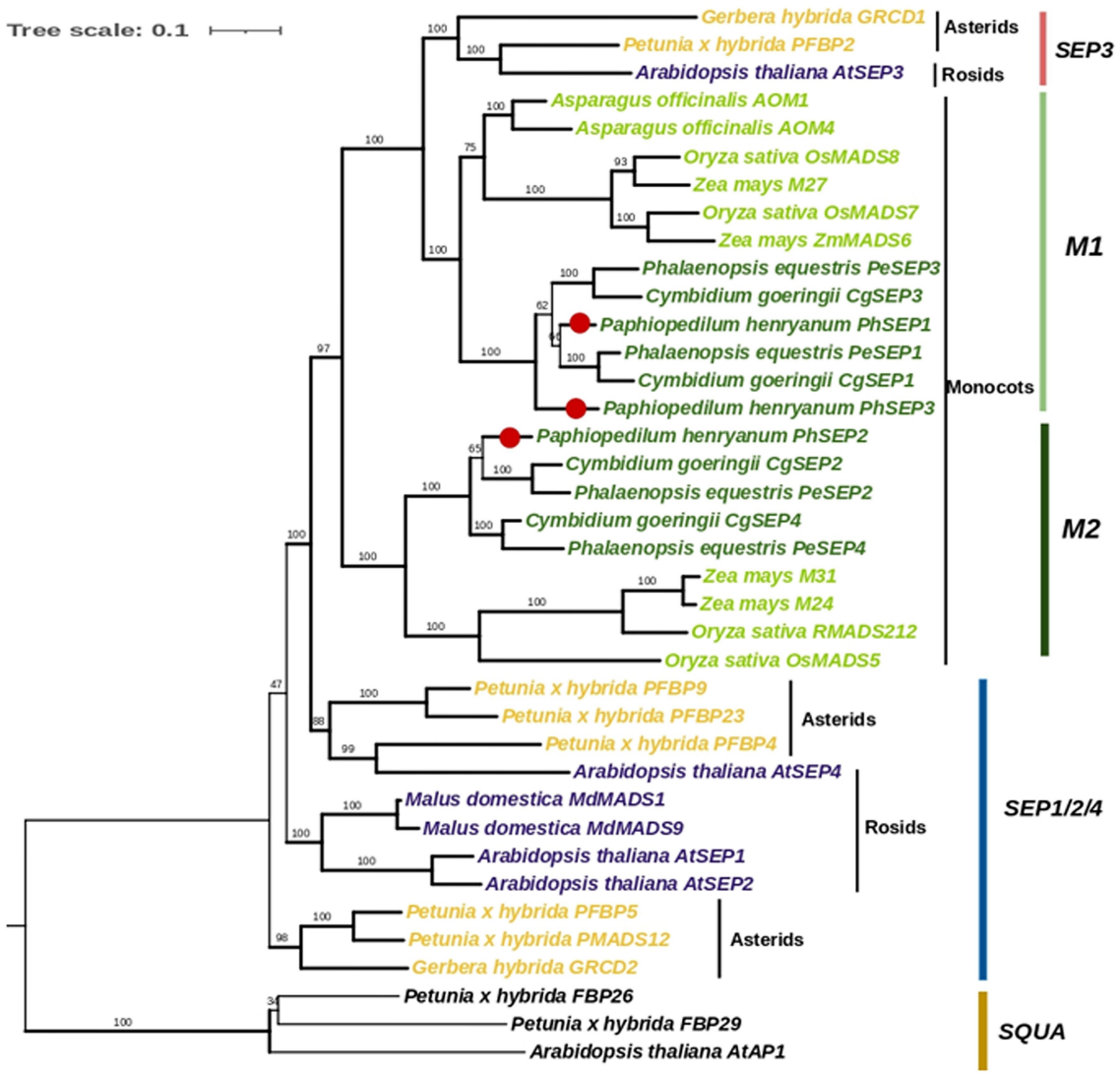


FIGURE 3 | Phylogenetic analysis of *SEP*-like genes with *SQUA* genes as an outgroup. The topology of this tree was generated using PhyML. ML bootstrap support (MLBS) values are indicated on each branch. Thick branches indicate high support values with MLBS ≥ 70. PhSEPs are marked with red dots.

in the roots and leaves. *PhSEP3* was predominantly expressed in the petals, dorsal sepals, ovaries, and especially lips. We examined the temporal expression pattern of *PhSEP* genes in floral buds at four developmental stages (**Figures 4D,F**). Abundant *PhSEP* transcripts were found throughout floral development, whereas stage B2 showed the highest transcript accumulation. Overall, the expression patterns of *PhSEP* genes indicated that *PhSEP* genes play multiple roles in the flower development of *P. henryanum*.

Subcellular Localization of PhSEP Proteins

To investigate the subcellular localization of PhSEP proteins, the EGFP–PhSEP fusion constructs and EGFP control were cloned in pBI221 under the regulatory control of the CaMV35S promoter. These constructs were transiently expressed in onion epidermal cells and analyzed by fluorescence microscopy. It was found that the PhSEP1–GFP, PhSEP2–GFP, and PhSEP1–GFP fusion proteins harboring nuclear localization signals were targeted to the nucleus, whereas the control GFP protein was localized in the cytosol and nucleus (**Figure 5**). These results indicated that PhSEPs are in fact nuclear proteins, in agreement with their role as transcription factors (TFs).

Self-Activation Detection and Protein–Protein Interaction Prediction

To investigate whether PhSEPs could be self-activated, we analyzed the ability of PhSEPs to activate the reporter genes *LacZ*, *TRP1*, *LEU1*, and *ADE2* in budding yeast. To this end, the CDS of the *PhSEP* genes were fused to the *GAL4* DNA-binding domain, and their ability to activate transcription from the *GAL4* upstream activation sequence (UAS) was assessed

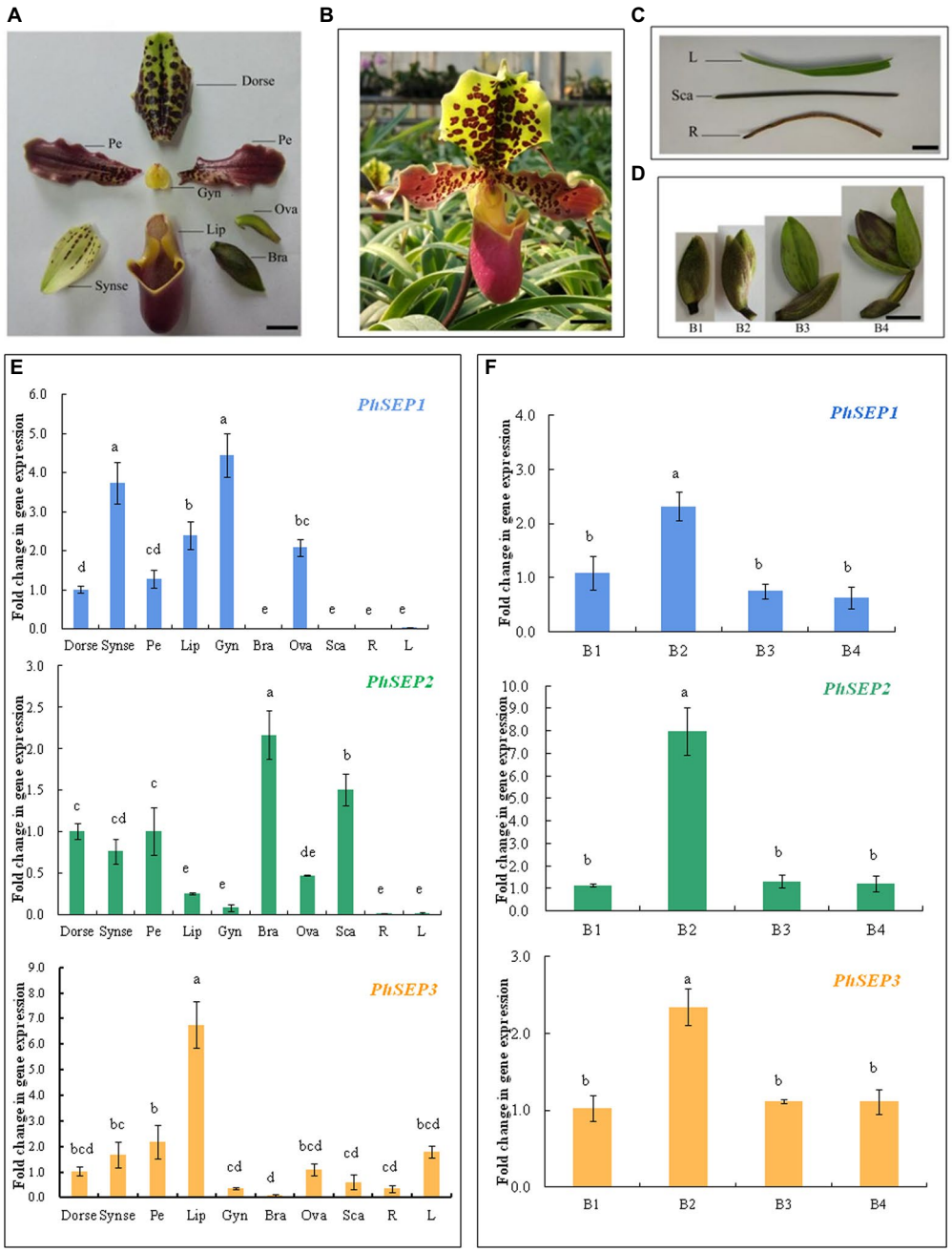


FIGURE 4 | Transcript levels of PhSEP genes in different tissues and organs and at different developmental stages of floral buds of *P. henryanum*. **(A)** A flower dissected as follows: Dorse, dorsal; sepal; Synse, synsepal; Pe, petal; Lip; Gyn, gynostemium; Bra, bract; Ova, ovary. **(B)** Mature flowers. **(C)** Vegetative tissues dissected as follows: R, root; Sca, scape; L, leaf. **(D)** Floral buds at different developmental stages. **(E)** Relative expression patterns of PhSEP1, PhSEP2, and PhSEP3 in different tissues and organs. **(F)** Relative expression patterns of PhSEP1, PhSEP2, and PhSEP3 at four stages of floral development. Scale bars: 10 mm; The values are means of three replicates ± SE. Statistical analysis was performed using one-way ANOVA test ($p < 0.05$). Different letters represent significant difference. The expression level of gene in Dorse or at stage B1 was set to 1, and those of others were normalized to it.

in terms of yeast growth. The yeast cells contained individual *PhSEP* plasmids and the control plasmid were sustained well on the non-selective SD-LW medium, whereas growth was absent on the selective SD-LWH + 3-AT or SD-LWHA medium, suggesting that no self-activation activity of PhSEPs was confirmed in yeast cells (**Figure 6**). In addition, the change in the color of the colonies from pink to red further indicated that the *ADE2* reporter gene was not expressed.

The neural network AlphaFold2 developed by the artificial intelligence company DeepMind was trained using multiple sequence alignments (MSA) and experimental protein structures deposited before April 30, 2018. It could be used to predict

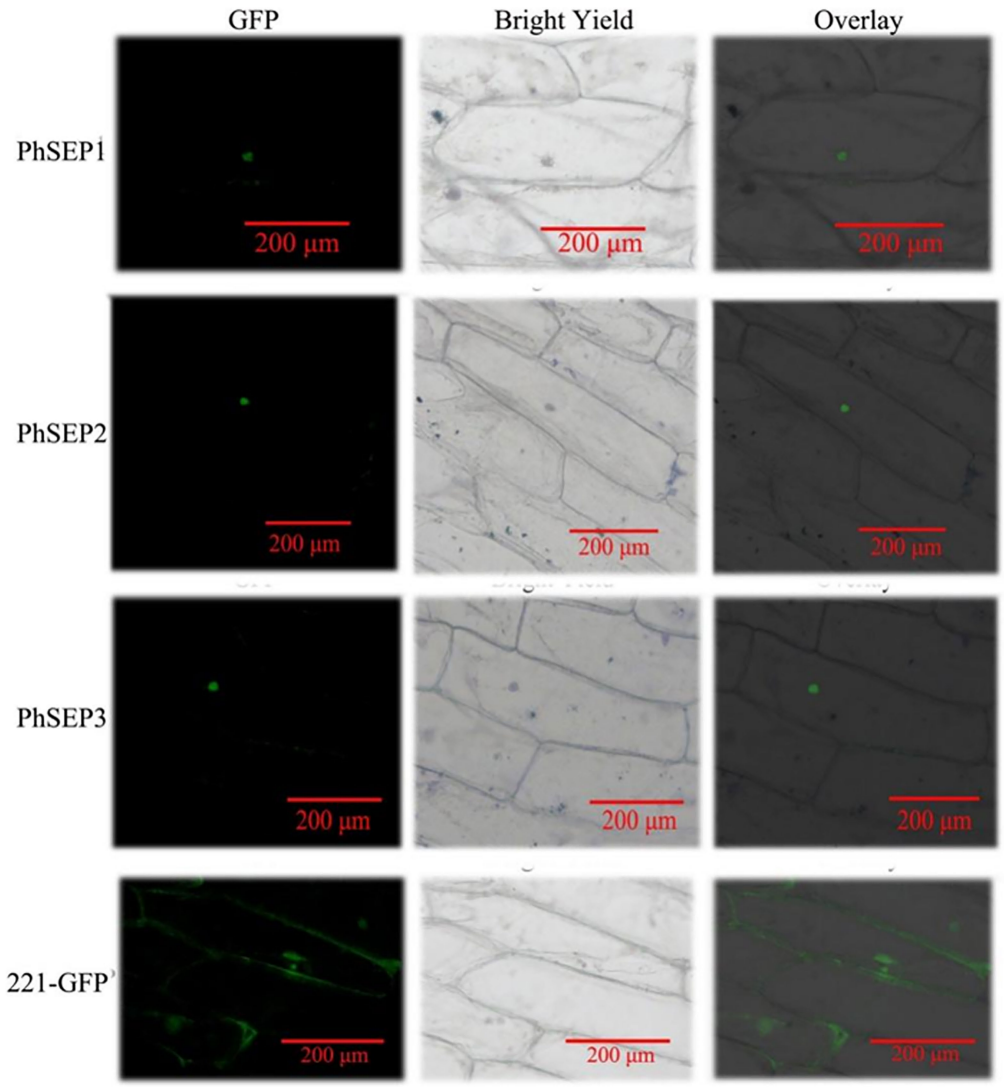


FIGURE 5 | Subcellular localization of the putative PhSEPs in onion (*Allium cepa*) inner epidermal cells. The onion epidermal cells transiently expressing PhSEP1–GFP, PhSEP2–GFP, and PhSEP1–GFP fusion proteins and the EGFP control were visualized through fluorescence microscopy. In each horizontal panel, the extreme left represents GFP fluorescence, the middle image represents bright field, and the right image represents an overlay of the other two images.

the protein structure at the atomic level with high accuracy. Besides, the protein–protein interactions could be also predicted by AlphaFold2 (Tunyasuvunakool et al., 2021; Hegeds et al., 2022). In this study, three SEP proteins, DEFICIENS-like protein (AKC93996.1) and flower meristem identity protein LEAFY (AKC94104.1) in *P. henryanum* were selected for the prediction of the protein–protein interactions (**Figure 7**). The structural models of the protein complexes revealed that PhSEP1-PhSEP2, PhSEP1-PhSEP3, and PhSEP2-PhSEP3 interaction were likely to have existed (**Figure 7A**). Besides, we also demonstrated that PhSEPs might interact with DEFICIENS-like proteins (**Figure 7B**). In **Figure 7C**, the structure of PhSEPs-LEAFY complexes was dispersed, and no hydrogen bonds were observed between the proteins, which indicates a potential lack of protein–protein interactions between PhSEPs and the central floral development protein LEAFY.

DISCUSSION

With more than 25,000 species, orchids are the second-largest plant family (Stokstad, 2015; Andriamihaja et al., 2021). They have a unique zygomorphic floral structure, including three sepals, two petals, and a highly diversified lip (Rudall and Bateman, 2002). The highly specialized and diverse morphology of flowers in orchids makes them excellent models for examining the complex network of regulatory genes involved in floral morphogenesis (Pan et al., 2011). Recently, *SEP*-like genes have been identified and characterized in a wide range of eudicots and monocots, including Arabidopsis, rice, and orchid (Pu et al., 2020; Adal et al., 2021; Zhu et al., 2022). Numerous reports have shown that these genes are instrumental in the floral evolution of diverse plants and play fundamental roles in floral organ fate determination during development

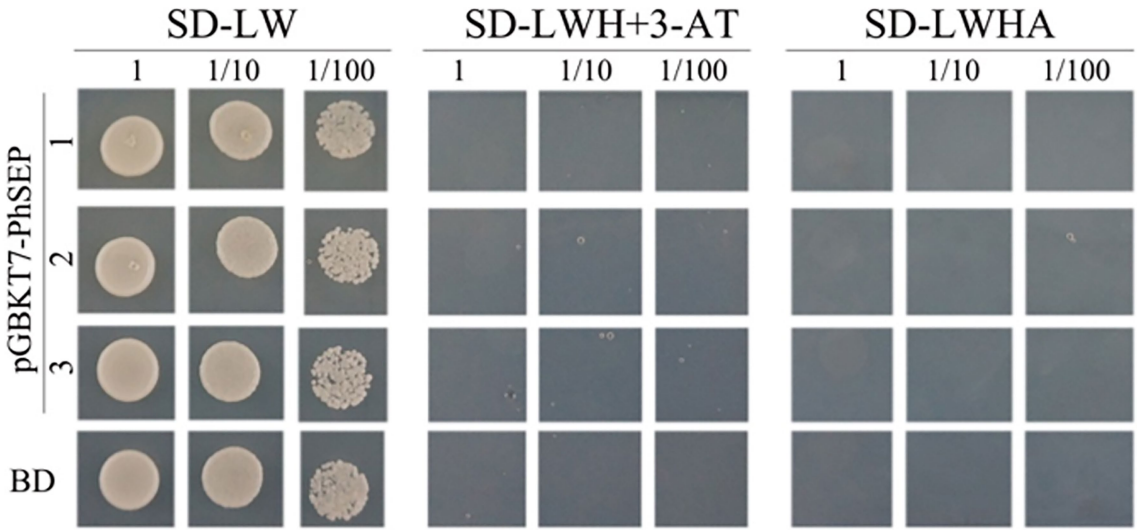


FIGURE 6 | Evaluation of the self-activation ability of PhSEPs in budding yeast. The yeast cells of strain AH109 harboring the indicated plasmids were grown on non-selective (SD-LW) or selective (SD-LWH + 3-AT and SD-LWHA) media. Decreasing cell densities represent the 10-fold dilution series. BD in the last row represents an empty GAL4 DNA-binding domain containing a vector.

by interacting with other MADS-box gene products, such as those from A-, B-, and C-class genes (Malcomber and Kellogg, 2005; Mitoma and Kanno, 2018; Qi et al., 2020; Pu and Xu, 2021).

In the present study, three *SEP*-like genes, *PhSEP1*, *PhSEP2*, and *PhSEP3*, were cloned from *Paphiopedilum* orchid. Sequence and phylogenetic analysis revealed that *PhSEP* genes from *Paphiopedilum* were highly conserved. In addition, the predicted amino acid sequences of PhSEPs showed a high degree of identity with homologous proteins from *P. equestris*, *Cymbidium goeringii*, and *A. thaliana* (Ditta et al., 2004; Pan et al., 2014; Yang et al., 2021). In general, a similar primary structure of proteins represents a relatively close evolutionary relationship, analogous structure, and identical functions (Ma et al., 2019). Furthermore, the conservation of SEP I and SEP II motifs in the highly variable C-terminus of PhSEPs supported their characterization as E-class floral meristem identity genes and suggested a similar functionality to their orthologs in other plants. Duplication events are common in MIKC-type MADS-box TFs and many MIKC-type homoeologs are functionally important and not redundant (Shan et al., 2009; Schilling et al., 2020). Consistent with this fact, several duplication events have been reported in the evolutionary lineages of *SEP* genes in both eudicots and monocots, resulting in four *SEP* members in *Arabidopsis*, *Cymbidium*, and *Phalaenopsis* (Ditta et al., 2004; Chang et al., 2009; Mondragón-Palomino, 2013; Pan et al., 2014). Extensive duplication of MADS-box genes and the resulting subfunctional and expressional differentiation were associated with regulation of species-specific flower traits, such as floral patterning, seasonal flowering, and ecological adaption (Yang et al., 2021). Phylogenetic analysis showed that the three *PhSEP* genes were sorted into two diversified clades (*M1* and the *M2* clade in monocots), in consistent with the findings of previous studies reporting the phylogeny of *SEP*-like genes from *P. equestris* and *C. goeringii*. According to the phylogenetic tree, three *PhSEP* genes from *Paphiopedilum* were clustered together with *SEP1/2/3* genes from other *P. equestris* and *C. goeringii*, respectively (Pan et al., 2014; Yang et al., 2021), so it is quite possible to clone *PhSEP4*, the orthologous orchid *SEP4* gene. Hence, according to our analysis result, the monocots *SEPs* did not belong to *SEP1/2/4* from eudicots, differently from previous study (Pan et al., 2014). We speculated that there might be more unidentified *SEP* genes in orchids if they were not lost in plants evolution. The frequent duplication of *SEP* genes might be one of the main cause of diversity of flower structure in angiosperms.

SEP-like genes encode MADS transcription factors required for the formation of all the organs of the flower and for the determinacy of the floral meristems (Valoroso et al., 2019; Pu and Xu, 2021). In this research, the *PhSEP* genes displayed differential spatial expression patterns in vegetative and reproductive tissues of *P. henryanum*. *PhSEP* genes were collectively expressed in all flower organs, as observed earlier in other plants (Xu et al., 2006; Pan et al., 2014; Adal et al., 2021). The expression levels of *PhSEP* genes in roots and leaves were negligible. These findings suggest that the *SEP*-like genes in orchids are involved in the specification of floral organ identity. Interestingly, *PhSEP1* displayed expression patterns complementary to those of *PhSEP2*. While high expression levels of *PhSEP1* were noted in the synsepal, lip, gynostemium, and ovary, *SEP2* showed high expression levels in bracts and scapes. This result agreed well from research from *C. ensifolium*. The expression of the *CeSEP1/3*-clade genes TDN29274 and TDN28990, the orthologs of *PeSEP1* and *PeSEP3*, respectively, was obviously reduced. However, the other two *CeSEP* genes showed equal or slightly higher expression in the leaf-like flower mutant of *C. ensifolium*

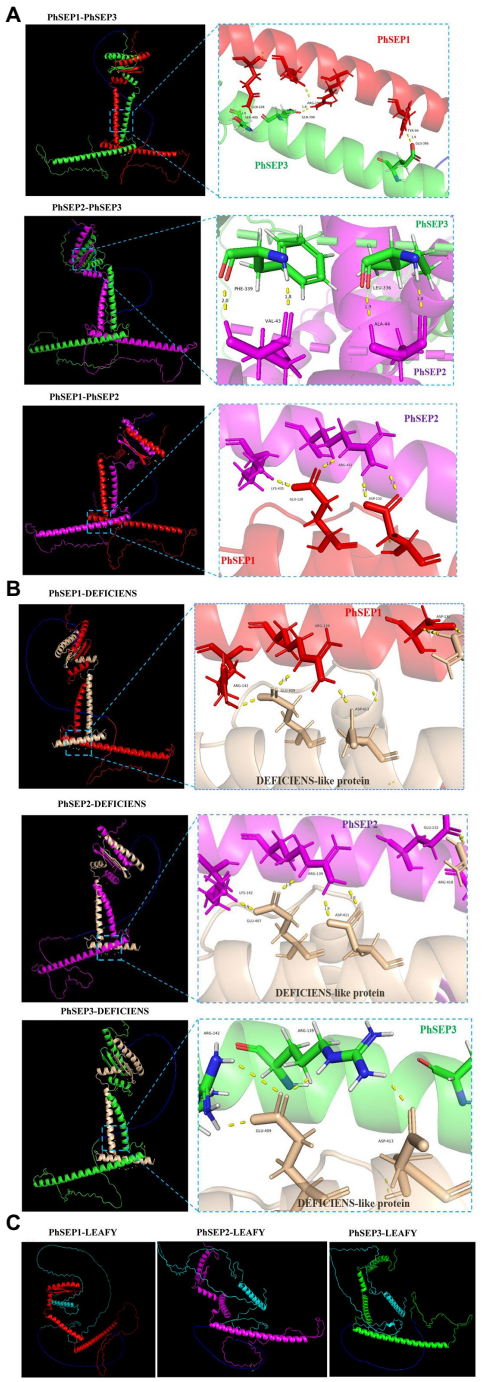


FIGURE 7 | The prediction of protein–protein interactions by DeepMind's AlphaFold2 program. **(A)** The prediction of PhSEPs-PhSEPs interactions. **(B)** The prediction of PhSEPs-DEFICIENS interactions. **(C)** The prediction of PhSEPs-LEAFY interactions. The red chain represents the structure of the PhSEP1 protein. The magenta chain represents the structure of the PhSEP2 protein. The green chain represents the structure of the PhSEP3 protein. The wheat chain represents the structure of the DEFICIENS-like protein. The cyan chain represents the structure of the LEAFY protein. The blue chain represents the linker. The yellow stick represents the hydrogen bonds between any of the two proteins. In **(A,B)** each horizontal panel, the left image represents the predicted structure of the protein complex, and the right image represents the predicted protein interaction sites.

(Wei et al., 2020). Similarly, the expression levels of *BroaSEP1/2/3* genes from *Brassica oleracea* were very different at different developmental stages, also in the wild type, mutant flower with increased petals, and mutant flower with decreased petals, which indicated that different patterns of gene expression may cause the flowers to increase or decrease the petal number (Xiang et al., 2020). Moreover, the expression patterns of *PhSEP2* were reminiscent of those of its orthologs *CgSEP2* and *PeSEP2*, which also display substantial expression in sepals and petals and are minimally expressed in the lip (Pan et al., 2014; Yang et al., 2021). The non-overlapping expression profiles of the *PhSEP* genes indicate possible functional divergence. One possible reason for this divergence may be problems caused by changes in the exon–intron structure of the SEP subfamily (Yu et al., 2016; Schilling et al., 2020). *SEP3* and its orthologs, such as *FBP2* (petunia), *TM5* (tomato), *WSEP* (wheat), and *EScaAGL9* from the basal eudicot California poppy (*Eschscholzia californica*), are only expressed in the inner three whorls of the flower (Angenent et al., 1992, 1994; Pelaz et al., 2000). In contrast, the *PhSEP3* transcripts were detected in all floral organ whorls, especially in lip, indicating that *PhSEP3* might be the key gene associated with the lip; thus, the mRNA expression pattern of *PhSEP3* was slightly different from that of its aforementioned orthologs. This finding may be attributed to the remarkable similarity between the sepals and petals of the flowers of *P. henryanum*, implying that the genes that control petal formation in slipper orchids might be similarly expressed in sepals (Chang et al., 2009).

TFs, a major driver in evolution and in domestication, can facilitate or obstruct the access of RNA polymerases to the DNA template in association with other transcriptional regulators, including chromatin-remodeling/modifying proteins (Udvardi et al., 2007; Martínez-Ainsworth and Tenaillon, 2016). MADS-box genes constitute one of the largest families of plant TFs (Riechmann et al., 2000). The *SEP* genes, which are E-class MADS-box TFs, play vital roles in various aspects of plant growth and development (Qi et al., 2020). In this study, three PhSEPs were located in the nucleus, indicating the possible involvement of these TFs proteins in the regulation of the expression of downstream genes associated with floral development. Recent yeast two-hybrid experiments demonstrated that SEP proteins have conserved interactions with other MADS-box proteins of the SQUA, DEF/GLO, and AG subfamilies (Zahn et al., 2005). Moreover, SEP-like proteins can interact with FUL-like proteins during fruit patterning of *E. pusilla* (Dirks-Mulder et al., 2019). Consistent with this role, PhSEPs were incapable of self-activation and the prediction of protein–protein interactions by AlphaFold2 showed that PhSEPs might interact with PhSEPs and B-class DEFICIENS-like proteins. Furthermore, the central floral development protein LEAFY is necessary in triggering flower formation on inflorescences, while the SEP family can reprogram cauline leaves into the floral organs. LEAFY promotes floral fate through upregulation of the floral commitment factor A-class APETALA1 (AP1; Jin et al., 2021). We found that LEAFY might not interact with E-class SEP proteins in *P. henryanum*.

Previous studies reported that the greenish flower phenotype of *Habenaria radiata* (Orchidaceae) is caused by a mutation in the *SEP*-like MADS-box gene *HrSEP-1* (Mitoma and Kanno, 2018). In *Lavandula angustifolia*, the expression of lavender *SEP*-like genes promote early flowering and alter leaf morphology in *A. thaliana* (Adal et al., 2021). In *Apostasia shenzhenica*, the adaxial petal does not differentiate into a specialized lip due to the loss of class B-AP3 and E-class genes (Zhang et al., 2017). Although there was no loss of B and E clade genes in *C. ensifolium*, transcriptomic analysis showed that the upregulation of *CeSEP-2* is necessary for the development of a specialized lip in *Cymbidium* orchids, while its downregulation results in the formation of a peloric flower shape (Ai et al., 2021). In addition, *CsSEP4* was originally found to positively regulate gynostemium development in *Cymbidium sinense*. The gene was ectopically expressed in the gynostemium of the wild-type flower and expended to all floral organs of a gynostemium-like perianth variant in *C. sinense*, and the *35S:CsSEP4 Arabidopsis* showed a severe flower phenotype whereas the *35S:CsSEP3* had an abnormal stamen and ovule (Yang et al., 2021). These results reveals that *SEP*-like genes are associated with the development of flower organs, especially for the lip. Thus, *PhSEP1/2/3* genes might have a similar function in floral organ identity. As in planta transformation of *Paphiopedilum maudiae* by agrobacterium-mediated ovary-injection was established, *FT* (*Flowering Locus T*) functional genes of *P. maudiae* were transformed into Paphiopedilum to elucidate their role during floral bud development and shorten the juvenile phase (Luo et al., 2020). This may be a useful way for the transformation of *PhSEP* genes in *P. henryanum*. There are differences in expression patterns and functional differentiation among paralogous genes or even among orthologous genes in closely related species (Morel et al., 2019; Yang et al., 2021). Functional differentiation might exist within the three *SEP*-like genes of *P. henryanum*.

The molecular basis of orchid flower development is accomplished through a specific regulatory program, and *SEP*-like genes enrich the molecular program underpinning the orchid perianth development, resulting in the expansion of the original "orchid code" in an even more complex gene regulatory network (Lucibelli et al., 2021). According to the quartet model, E-class genes are essential for the formation of quaternary complexes (Mitoma and Kanno, 2018). We suspected that *PhSEPs* genes may act as the "glue" for MADS-box transcription factor complex formation to regulate perianth formation in *P. henryanum*. In recent years, the genomes of some orchids including *Apostasia shenzhenica*, *Phalaenopsis aphrodite*, *Vanilla planifolia*, *C. goeringii*, *Dendrobium chrysotoxum*, *Platanthera zijinensis*, and *Platanthera guangdongensis* were published (Zhang et al., 2017, 2021; Chao et al., 2018; Hasing et al., 2020; Sun et al., 2021; Li et al., 2022). By contrast, the occurrence of whole-genome duplication in *Paphiopedilum* results in a genome that is very large and complex Besides, A-, B-, C-, D-class MADS-box genes in *P. henryanum* have not been systematically identified, which restricts the analysis of the protein–protein interaction network. Moreover, the tissue culture and genetic transformation systems are not applied widely in *Paphiopedilum* industry, and the mutant of this flower is difficult to be obtained and preserved. More functional data are required to validate our orchid flower regulatory model, such as breeding *PhSEP* genes-overexpressing and gene-silenced mutants by virus-induced gene silencing and transgenic technology, protein–protein interaction validation by yeast two-hybrid system and bimolecular fluorescence complementation, and downstream target genes detection by chromatin immunoprecipitation, electrophoretic mobility shift assay, and dual-luciferase. The biological function of *PhSEP* genes still needs to be evaluated in the *Paphiopedilum* or model plant *Arabidopsis*. Further research is required to explore the mechanisms underlying floral development.

CONCLUSION

In this study, three *SEP*-like MADS-box genes in slipper orchids were identified for the first time, and the characteristics and expression patterns of the gene and protein sequences were systematically analyzed. All three homologs were structurally conserved and were characterized as E-class MADS-box transcription factors. Phylogenetic analysis revealed that *PhSEP1*, *PhSEP2*, and *PhSEP3* were evolutionarily closer to the core eudicot *SEP3* lineage, whereas none of them belonged to core eudicot *SEP1/2/4* clade. *PhSEP* genes were expressed during flower development and exhibited non-ubiquitous expression patterns. All SEP proteins were localized to the nucleus. Furthermore, we observed no self-activation of SEP proteins and the prediction of protein–protein interactions by AlphaFold2 revealed that SEP proteins might interact with SEP and DEFICIENS-like proteins. Consequently, these results illustrate three *SEP*-like MADS-box genes *PhSEP1*, *PhSEP2*, and *PhSEP3* might play a vital role in flower development in *P. henryanum*. Future research needs to be conducted to further elucidate the regulatory networks underlying the floral development and organ identity in the slipper orchid.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

RJ, HG, and MR: conceptualization and supervision. XX, SY, XZ, and YL: methodology. RJ, YXi, HC, and LZ: software. YXu and NM: validation. HC and XX: formal analysis and investigation. RJ, HG, SY, and XZ: resources. HC and YL: data curation. HC: writing—original draft preparation and visualization. HC, XX, NM, and YL: writing—review and editing. HG and RJ: project

administration and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (grant number 31301810), the Special Fund for Agro-scientific Research in the Public Interest (grant number 201203071), and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (grant number 34-IUA-02).

ACKNOWLEDGMENTS

We gratefully thank Jiang Y.-Q. and Rehmani M. S. for providing some help in terms of research methodologies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.916081/full#supplementary-material

REFERENCES

Adal, A. M., Binson, E., Remedios, L., and Mahmoud, S. S. (2021). Expression of lavender *AGAMOUS*-like and *SEPALLATA3*-like genes promote early flowering and alter leaf morphology in *Arabidopsis thaliana*. *Planta* 254, 54–12. doi: 10.1007/s00425-021-03703-3

Ai, Y., Li, Z., Sun, W. H., Chen, J., Zhang, D., Ma, L., et al. (2021). The *Cymbidium* genome reveals the evolution of unique morphological traits. *Hortic. Res.* 8:255. doi: 10.1038/s41438-021-00683-z

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., et al. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5328–5333. doi: 10.1073/pnas.97.10.5328

Andriamihaja, C. F., Ramarosandratana, A. V., Grisoni, M., Vololoniaina, G., and Besse, J. P. (2021). Drivers of population divergence and species differentiation in a recent group of indigenous orchids (*Vanilla* spp.) in Madagascar. *Ecol. Evol.* 11, 2681–2700. doi: 10.1002/ece3.7224

Angenent, G. C., Busscher, M., Franken, J., Mol, J. N. M., and Van Tunen, A. J. (1992). Differential expression of two MADS box genes in wild-type and mutant petunia flowers. *Plant Cell* 4, 983–993. doi: 10.1105/tpc.4.8.983

Angenent, G. C., Franken, J., Busscher, M., Weiss, D., and Van Tunen, A. J. (1994). Co-suppression of the petunia homeotic gene *FBP2* affects the identity of the generative meristem. *Plant J.* 5, 33–44. doi: 10.1046/j.1365-313X.1994.5010033.x

Becker, A., and Theißen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol. Phylogenet. Evol.* 29, 464–489. doi: 10.1016/S1055-7903(03)00207-0

Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using alphafold2. *Nat. Commun.* 13:1694. doi: 10.1038/s41467-022-29480-5

Chang, Y. Y., Chiu, Y. F., Wu, J. W., and Yang, C. H. (2009). Four Orchid (*Oncidium Gower* Ramsey) *AP1/AGL9*-like MADS box genes show novel expression patterns and cause different effects on floral transition and formation in *Arabidopsis thaliana*. *Plant Cell Physiol.* 50, 1425–1438. doi: 10.1093/pcp/pcp087

Chao, Y. T., Chen, W. C., Chen, C. Y., Ho, H. Y., Yeh, C. H., Kuo, Y. T., et al. (2018). Chromosome-level assembly, genetic and physical mapping of *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnol. J.* 16, 2027–2041. doi: 10.1111/pbi.12936

Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0

Cui, R., Han, J., Zhao, S., Su, K., Wu, F., Du, X., et al. (2010). Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). *Plant J.* 61, 767–781. doi: 10.1111/j.1365-313X.2009.04101.x

De Bodt, S., Raes, J., Florquin, K., Rombauts, S., Rouzé, P., Theißen, G., et al. (2003). Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. *J. Mol. Evol.* 56, 573–586. doi: 10.1007/s00239-002-2426-x

De Folter, S., and Immink, R. G. (2011). "Yeast protein-protein interaction assays and screens," in *Plant Transcription Factors. Methods and Protocols (Methods in Molecular Biology)*. eds. L. Yuan and S. E. Perry (New York, NY: Springer), 145–165.

Dirks-Mulder, A., Ahmed, I., Uit Het Broek, M., Krol, L., Menger, N., and Snier, J. (2019). Morphological and molecular characterization of orchid fruit development. *Front. Plant Sci.* 10:137. doi: 10.3389/fpls.2019.00137

Ditta, G., Pinyopich, A., Robles, P., Pelaz, S., and Yanofsky, M. F. (2004). The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr. Biol.* 14, 1935–1940. doi: 10.1016/j.cub.2004.10.028

Ferrario, S., Immink, R. G. H., Shchennikova, A., Busscher-Lange, J., and Angenent, G. C. (2003). The MADS box gene *FBP2* is required for *SEPALLATA* function in petunia. *Plant Cell* 15, 914–925. doi: 10.1105/tpc.010280

Goto, K., and Meyerowitz, E. M. (1994). Function and regulation of the Arabidopsis floral homeotic gene *PISTILLATA*. *Genes Dev.* 8, 1548–1560. doi: 10.1101/gad.8.13.1548

Guan, Z. J., Zhang, S. B., Guan, K. Y., Li, S. Y., and Hu, H. (2011). Leaf anatomical structures of *Paphiopedilum* and *Cypripedium* and their adaptive significance. *J. Plant Res.* 124, 289–298. doi: 10.1007/s10265-010-0372-z

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Guo, X., Chen, G., Naeem, M., Yu, X., Tang, B., Li, A., et al. (2017). The MADS-box gene *SlMBP11* regulates plant architecture and affects reproductive development in tomato plants. *Plant Sci.* 258, 90–101. doi: 10.1016/j.plantsci.2017.02.005

Guo, B., Zeng, S., Yin, Y., Li, L., Ma, G., Wu, K., et al. (2021). Characterization of phytohormone and transcriptome profiles during protocorm-like bodies development of Paphiopedilum. *BMC Genomics* 22:806. doi: 10.1186/s12864-021-08087-y

Hasing, T., Tang, H., Brym, M., Khazi, F., Chambers, A., and H., (2020). A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nat. Food* 1, 811–819. doi: 10.1038/s43016-020-00197-2

Hegeds, T., Geisler, M., Lukács, G. L., and Farkas, B. (2022). Ins and outs of AlphaFold2 transmembrane protein structure predictions. *Cell. Mol. Life Sci.* 79, 73–12. doi: 10.1007/s00018-021-04112-1

Hsu, H., Hsu, W., Chen, Y. I., Mao, W., Yang, J., Li, J., et al. (2015). Model for perianth formation in orchids. *Nat. Plants* 1:15046. doi: 10.1038/nplants.2015.46

Hu, Y., Yu, Q., Chen, Y., Zhang, L., Fu, X. R., Ding, L., et al. (2021). Labour-saving construction of a target protein interaction network by selective culture and high-throughput sequencing. *Biotechnol. J.* 16:e2100204. doi: 10.1002/biot.202100204

Humphreys, I. R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., et al. (2021). Computed structures of core eukaryotic protein complexes. *Science* 3745:eabm4805. doi: 10.1126/science.abm4805

Immink, R. G. H., Tonaco, I. A. N., de Folter, S., Shchennikova, A., van Dijk, A. D. J., Busscher-Lange, J., et al. (2009). SEPALLATA3: The 'glue' for MADS box transcription factor complex formation. *Genome Biol.* 10:R24. doi: 10.1186/gb-2009-10-2-r24

Jin, R., Klasfeld, S., Zhu, Y., Garcia, M. F., Xiao, J., Han, S. K., et al. (2021). LEAFY is a pioneer transcription factor and licenses cell reprogramming to floral fate. *Nat. Commun.* 12:626. doi: 10.1038/s41467-020-20883-w

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

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

Li, W., Dang, C., Ye, Y., Wang, Z., and Wang, F. (2020). Overexpression of grapevine *vviaa18* gene enhanced salt tolerance in tobacco. *Int. J. Mol. Sci.* 21:1323. doi: 10.3390/ijms21041323

Li, M. H., Liu, K. W., Li, Z., Lu, H. C., Ye, Q. L., and Zhang, (2022). Genomes of leafy and leafless platanthera orchids illuminate the evolution of mycoheterotrophy. *Nat. Plants* 8, 373–388. doi: 10.1038/s41477-022-01127-9

Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowmant, J. L., and Yanofsky, M. F. (2000). *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770. doi: 10.1038/35008089

Lin, C. S., Hsu, C. T., Liao, D. C., Chang, W. J., Chou, M. L., Huang, Y. T., et al. (2016). Transcriptome-wide analysis of the MADS-box gene family in the orchid *Erycina pusilla*. *Plant Biotechnol. J.* 14, 284–298. doi: 10.1111/pbi.12383

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2 [-Delta Delta C(T)] method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lu, Z. X., Wu, M., Loh, C. S., Yeong, C. Y., and Goh, C. J. (1993). Nucleotide sequence of a flower-specific MADS box cDNA clone from orchid. *Plant Mol. Biol.* 23, 901–904. doi: 10.1007/BF00021545

Lucibelli, F., Valoroso, M. C., Theißen, G., Nolden, S., Mondragon-Palomino, M., and Aceto, S. (2021). Extending the toolkit for beauty: differential co-expression of drooping *leaf*-like and class B MADS-box genes during *Phalaenopsis* flower development. *Int. J. Mol. Sci.* 22:7025. doi: 10.3390/ijms22137025

Luo, B. X., Zhang, L., Zheng, F., Wu, K. L., Li, L., Zhang, X. H., et al. (2020). Ovule development and in planta transformation of *Paphiopedilum maudiae* by agrobacterium-mediated ovary-injection. *Int. J. Mol. Sci.* 22:84. doi: 10.3390/ijms22010084

Ma, Y. Q., Pu, Z. Q., Zhang, L., Lu, M. X., Zhu, Y., Hao, C. Y., et al. (2019). A *SEPALLATA1-like* gene of *Isatis indigotica* fort. Regulates flowering time and specifies floral organs. *Gene* 713:143974. doi: 10.1016/j.gene.2019.143974

Malcomber, S. T., and Kellogg, E. A. (2005). *SEPALLATA* gene diversification: brave new whorls. *Trends Plant Sci.* 10, 427–435. doi: 10.1016/j.tplants.2005.07.008

Mandel, M. A., and Yanofsky, M. F. (1995). A gene triggering flower formation in Arabidopsis. *Nature* 377, 522–524. doi: 10.1038/377522a0

Martínez-Ainsworth, N. E., and Tenaillon, M. I. (2016). Superheroes and masterminds of plant domestication. *C. R. Biol.* 339, 268–273. doi: 10.1016/j.crvi.2016.05.005

Mitoma, M., and Kanno, A. (2018). The greenish flower phenotype of habenaria radiata (Orchidaceae) is caused by a mutation in the *SEPALLATA-like* MADS-box gene *HrSEP-1*. *Front. Plant Sci.* 9:831. doi: 10.3389/fpls.2018.00831

Mondragón-Palomino, M. (2013). Perspectives on MADS-box expression during orchid flower evolution and development. *Front. Plant Sci.* 4:377. doi: 10.3389/fpls.2013.00377

Mondragõn-Palomino, M., and Theißen, G. (2011). Conserved differential expression of paralogous *DEFICIENS*- and *GLOBOSA*-like MADS-box genes in the flowers of Orchidaceae: refining the 'orchid code'. *Plant J.* 66, 1008–1019. doi: 10.1111/j.1365-313X.2011.04560.x

Morel, P., Chambrier, P., Boltz, V., Chamot, S., Rozier, F., Rodrigues Bento, S., et al. (2019). Divergent functional diversification patterns in the SEP/AGL6/AP1 MADS-box transcription factor superclade. *Plant Cell* 31, 3033–3056. doi: 10.1105/tpc.19.00162

Ng, C. Y., and Mohd Saleh, N. (2011). In vitro propagation of Paphiopedilum orchid through formation of protocorm-like bodies. *Plant Cell Tissue Organ Cult.* 105, 193–202. doi: 10.1007/s11240-010-9851-0

Omondi, B. A., Latorre-Estivalis, J. M., Oliveira, I. H. R., Ignell, R., and Lorenzo, M. G. (2015). Evaluation of reference genes for insect olfaction studies. *Parasit. Vectors* 8:243. doi: 10.1186/s13071-015-0862-x

Pan, Z. J., Chen, Y. Y., Du, J. S., Chen, Y. Y., Chung, M. C., Tsai, W. C., et al. (2014). Flower development of Phalaenopsis orchid involves functionally divergent *SEPALLATA-like* genes. *New Phytol.* 202, 1024–1042. doi: 10.1111/nph.12723

Pan, Z. J., Cheng, C. C., Tsai, W. C., Chung, M. C., Chen, W. H., Hu, J. M., et al. (2011). The duplicated B-class MADS-box genes display dualistic characters in orchid floral organ identity and growth. *Plant Cell Physiol.* 52, 1515–1531. doi: 10.1093/pcp/pcr092

Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require *SEPALLATTA* MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103

Pi, Q. X., Yang, N., Hu, H., and Li, S. Y. (2009). Flower development and cultivation of *Paphiopedilum armeniacum* (orchidaceae). *Acta Bot. Yunnanica* 31, 296–302. doi: 10.3724/SP.J.1143.2009.09048

Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817

Pozzati, G., Zhu, W., Lamb, J., Bassot, C., and Elofsson, A. (2021). Limits and potential of combined folding and docking. *Bioinformatics* 38, 954–961. doi: 10.1101/2021.06.04.446442

Pu, Z. Q., Ma, Y. Y., Lu, M. X., Ma, Y. Q., and Xu, Z. Q. (2020). Cloning of a *SEPALLATA4*-like gene (*Iisep4*) in *Isatis indigotica* fortune and characterization of its function in *Arabidopsis thaliana*. *Plant Physiol Bioch.* 154, 229–237. doi: 10.1016/j.plaphy.2020.05.031

Pu, Z. Q., and Xu, Z. Q. (2021). Functions of the E-class floral homeotic genes in several common dicotyledons. *J. Plant Growth Regul.* 41, 524–534. doi: 10.1007/s00344-021-10318-1

Qi, X., Liu, C., Song, L., and Li, M. (2020). PaMADS7, a MADS-box transcription factor, regulates sweet cherry fruit ripening and softening. *Plant Sci.* 301:110634. doi: 10.1016/j.plantsci.2020.110634

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

Rudall, P. J., and Bateman, R. M. (2002). Roles of synorganisation, zygomorphy and heterotopy in floral evolution: the gynostemium and labellum of orchids and other lilioid monocots. *Biol. Rev. Camb. Philos. Soc.* 77, 403–441. doi: 10.1017/S1464793102005936

Schilling, S., Kennedy, A., Pan, S., Jermiin, L. S., and Melzer, R. (2020). Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol.* 225, 511–529. doi: 10.1111/nph.16122

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 250, 931–936. doi: 10.1126/science.250.4983.931

Shan, H., Zahn, L., Guindon, S., Wall, P. K., Kong, H., Ma, H., et al. (2009). Evolution of plant MADS box transcription factors: evidence for shifts in selection associated with early angiosperm diversification and concerted gene duplications. *Mol. Biol. Evol.* 26, 2229–2244. doi: 10.1093/molbev/msp129

Shen, C. Y., Chen, Y. Y., Liu, K. W., Lu, H. C., Chang, S. B., and Hsiao, Y. Y. (2021). Orchid Bsister gene *PeMADS28* displays conserved function in ovule integument development. *Sci. Rep.* 11:1205. doi: 10.1038/s41598-020-79877-9

Stokstad, E. (2015). Orchids' dazzling diversity explained. *Science* 349, 914. doi: 10.1126/science.aad1667

Sun, Y., Chen, G. Z., Huang, J., Liu, D. K., Xue, F., Chen, X. L., et al. (2021). The *Cymbidium goeringii* genome provides insight into organ development and adaptive evolution in orchids. *Ornam. Plant Res* 1:10, 10.48130/OPR-2021-0010

Theißen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85. doi: 10.1016/S1369-5266(00)00139-4

Theissen, G., and Melzer, R. (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Ann. Bot.* 100, 603–619. doi: 10.1093/aob/mcm143

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882. doi: 10.1093/nar/25.24.4876

Tsaban, T., Varga, J. K., Avraham, O., Ben-Aharon, Z., Khramushin, A., and Schueler-Furman, O. (2022). Harnessing protein folding neural networks for peptide-protein docking. *Nat. Commun.* 13:176. doi: 10.1038/s41467-021-27838-9

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., and Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature* 596, 590–596. doi: 10.1038/s41586-021-03828-1

Udvardi, M. K., Kakar, K., Wandrey, M., Montanari, O., Murray, J., Andriankaja, A., et al. (2007). Legume transcription factors: global regulators of plant development and response to the environment. *Plant Physiol.* 144, 538–549. doi: 10.1104/pp.107.098061

Uimari, A., Kotilainen, M., Elomaa, P., Yu, D., Albert, V. A., and Teeri, T. H. (2004). Integration of reproductive meristem fates by a *SEPALLATA-like*

MADS-box gene. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15817–15822. doi: 10.1073/pnas.0406844101

Valoroso, M. C., Censullo, M. C., and Aceto, S. (2019). The MADS-box genes expressed in the inflorescence of *Orchis italica* (Orchidaceae). *PLoS One* 14:e0213185. doi: 10.1371/journal.pone.0213185

Wang, F., Tong, W., Zhu, H., Kong, W., Peng, R., Liu, Q., et al. (2016). A novel *Cys2/His2* zinc finger protein gene from sweetpotato, *IbZFP1*, is involved in salt and drought tolerance in transgenic *Arabidopsis*. *Planta* 243, 783–797. doi: 10.1007/s00425-015-2443-9

Wei, Y., Jin, J., Yao, X., Lu, C., Zhu, G., and Yang, F. (2020). Transcriptome analysis reveals clues into leaf-like flower mutant in Chinese orchid *Cymbidium ensifolium*. *Plant Divers.* 42, 92–101. doi: 10.1016/j.pld.2019.12.001

Xiang, Y., Huang, Y., He, H., and Xu, Q. (2020). Phylogenetic and expression analysis of *SEPALLATA*-like gene in *Brassica oleracea* L. var. acephala. *Sheng Wu Gong Cheng Xue Bao* 36, 2398–2412. doi: 10.13345/j.cjb.200422

Xu, Y., Jia, R., Zhou, Y., Cheng, H., Zhao, X., and Ge, H. (2018). Development and characterization of polymorphic EST-SSR markers for *Paphiopedilum henryanum* (Orchidaceae). *Appl. Plant Sci.* 6:e01152. doi: 10.1002/aps3.1152

Xu, Y., Teo, L. L., Zhou, J., Kumar, P. P., and Yu, H. (2006). Floral organ identity genes in the orchid *Dendrobium crumenatum*. *Plant J.* 46, 54–68. doi: 10.1111/j.1365-313X.2006.02669.x

Yang, F. X., Guo, J., Wei, Y. L., Ren, R., Zhang, G. Q., Lu, C. Q., et al. (2021). The genome of *Cymbidium sinense* revealed the evolution of orchid traits. *Plant Biotechnol. J.* 19, 2501–2516. doi: 10.1111/pbi.13676

Yu, H., and Goh, C. J. (2000). Identification and characterization of three orchid MADS-box genes of the AP1/AGL9 subfamily during floral transition. *Plant Physiol.* 123, 1325–1336. doi: 10.1104/pp.123.4.1325

Yu, X., Duan, X., Zhang, R., Fu, X., Ye, L., Kong, H., et al. (2016). Prevalent exon-intron structural changes in the *APETALA1/FRUITFULL*, *SEPALLATA*, *AGAMOUS*-like 6, and *FLOWERING LOCUS C* MADS-box gene subfamilies provide new insights into their evolution. *Front. Plant Sci.* 7:598. doi: 10.3389/fpls.2016.00598

Zahn, L. M., Kong, H., Leebens-Mack, J. H., Kim, S., Soltis, P. S., Landherr, L. L., et al. (2005). The evolution of the *SEPALLATA* subfamily of MADS-box genes: a preangiosperm origin with multiple duplications throughout angiosperm history. *Genetics* 169, 2209–2223. doi: 10.1534/genetics.104.037770

Zeng, S., Wang, J., Wu, K., Teixeira da Silva, J. A., Zhang, J., and Duan, J. (2013). In vitro propagation of *Paphiopedilum hangianum* Perner & Gruss. *Sci. Hortic.* 151, 147–156. doi: 10.1016/j.scienta.2012.10.032

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

Zhang, Y. X., Zhang, G. Q., Zhang, D. Y., Liu, X. D., Xu, X. Y., Sun, W. H., et al. (2021). Chromosome-scale assembly of the *Dendrobium chrysotoxum* genome enhances the understanding of orchid evolution. *Hortic. Res.* 8:183. doi: 10.1038/s41438-021-00621-z

Zhu, W., Yang, L., Wu, D., Meng, Q., Deng, X., Huang, G., et al. (2022). Rice *SEPALLATA* genes *OsMADS5* and *OsMADS34* cooperate to limit inflorescence branching by repressing the TERMINAL FLOWER1-like gene RCN4. *New Phytol.* 233, 1682–1700. doi: 10.1111/nph.17855

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.

frontiers | Frontiers in Plant Science

PERSPECTIVE
published: 27 May 2022
doi: 10.3389/fpls.2022.910362

Orchid Phylotranscriptomics: The Prospects of Repurposing Multi-Tissue Transcriptomes for Phylogenetic Analysis and Beyond

Darren C. J. Wong and Rod Peakall*

Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia

The Orchidaceae is rivaled only by the Asteraceae as the largest plant family, with the estimated number of species exceeding 25,000 and encompassing more than 700 genera. To gain insights into the mechanisms driving species diversity across both global and local scales, well-supported phylogenies targeting different taxonomic groups and/or geographical regions will be crucial. High-throughput sequencing technologies have revolutionized the field of molecular phylogenetics by simplifying the process of obtaining genome-scale sequence data. Consequently, there has been an explosive growth of such data in public repositories. Here we took advantage of this unprecedented access to transcriptome data from predominantly non-phylogenetic studies to assess if it can be repurposed to gain rapid and accurate phylogenetic insights across the orchids. Exhaustive searches revealed transcriptomic data for more than 100 orchid species spanning 5 subfamilies, 13 tribes, 21 subtribes, and 50 genera that were amendable for exploratory phylotranscriptomic analysis. Next, we performed re-assembly of the transcriptomes before strategic selection of the final samples based on a gene completeness evaluation. Drawing on these data, we report phylogenetic analyses at both deep and shallow evolutionary scales *via* maximum likelihood and shortcut coalescent species tree methods. In this perspective, we discuss some key outcomes of this study and conclude by highlighting other complementary, albeit rarely explored, insights beyond phylogenetic analysis that repurposed multi-tissue transcriptome can offer.

Keywords: phylogeny, orchids, transcriptome, next-generation sequencing–NGS, plastome, target sequence capture sequencing, phylogenomics, phylotranscriptomics

OPEN ACCESS

Edited by:
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

Reviewed by:
Tatiana Arias,
Marie Selby Botanical Gardens, United States
Ali Raza,
Fujian Agriculture and Forestry University, China

****Correspondence:***
Darren C. J. Wong
darren.wong@anu.edu.au;
wongdcj@gmail.com

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: *01 April 2022*
Accepted: *21 April 2022*
Published: *27 May 2022*

Citation:
Wong DCJ and Peakall R (2022) Orchid Phylotranscriptomics: The Prospects of Repurposing Multi-Tissue Transcriptomes for Phylogenetic Analysis and Beyond. Front. Plant Sci. 13:910362. doi: 10.3389/fpls.2022.910362

INTRODUCTION

The Orchidaceae is rivaled only by the Asteraceae as the largest plant family, with an estimate of some 25,000 species (∼ 8% of plants) shared across at least 700 genera (Chase et al., 2015). Orchids can be found in a wide range of habitats across tropical, subtropical, and temperate regions (Swarts and Dixon, 2009), and hold great interest from an ecological and evolutionary standpoint. They are particularly well known for their diverse morphological adaptations such as epiphytism, with at least 18,000 species adopting this lifestyle (Gravendeel et al., 2004), and their

dependence on mycorrhizal fungi interactions for germination (Phillips et al., 2020). Additionally, many orchids employ highly specialized or unusual pollination strategies (Jersáková et al., 2006; Schiestl and Schlüter, 2009; Bohman et al., 2016; Wong et al., 2017b; Trunschke et al., 2021). These characteristics make orchids ideal targets for exploring diverse ecological and evolutionary questions (Peakall, 2007).

Plastome and Mitochondrial Genome Phylogenomics in Orchids

To gain insights into the mechanisms driving the species diversity of the Orchidaceae across both global and local scales, investigations targeting different taxonomic and geographic scales (e.g., across specific tribe to genus) and a well-supported phylogeny are crucial. Rapid advancements in high-throughput sequencing and bioinformatics methods have enabled the use of genome (nuclear, plastid, and mitochondrial) and transcriptome sequences (including coding and non-coding regions) amendable for phylogenetic inference (McKain et al., 2018). Currently, the newfound ease of sequencing whole plastomes makes cpDNA the most widely used target for both small- and broad-scale orchid phylogenetic studies in terms of size (i.e., from tens to hundreds of species) and taxonomic breadth (i.e., within a genus or across five subfamilies) (Barrett et al., 2014; Givnish et al., 2015; Li Y. X. et al., 2019; Kim et al., 2020; Liu et al., 2020; Smidt et al., 2020; Serna-Sánchez et al., 2021). Notably, the first broad-scale plastome phylogeny for the Orchidaceae was based on maximum-likelihood (ML) analysis of 75 cpDNA coding-genes across 39 genera representing all five subfamilies and 16 (of 17) tribes (Givnish et al., 2015). Many previously underrepresented genera are now included in a cpDNA study of 78 genes spanning 264 species, 117 genera, 18 tribes, and 28 subtribes (Serna-Sánchez et al., 2021).

Phylotranscriptomics in Orchids

Notwithstanding the extraordinary insight from plastome studies gained in the last decade, phylogenetic studies leveraging transcriptome datasets to identify putative single-copy genes (or orthogroups) are now emerging as a new tool to aid plant molecular systematics. Compared to the linked genes in cpDNA which collectively represent just one evolutionary history, nuclear genes offer several advantages including their biparental inheritance, higher substitution rates, and access to thousands of unlinked genes each with the potential to represent independent evolutionary histories, among others (Zhang et al., 2012; Smith, 2015).

To date, several broad-scale transcriptome-based phylogenetic analyses of the Orchidaceae have been performed, but with limited taxon sampling compared to the available plastome-based phylogenies (Deng et al., 2015; Zhang et al., 2017; Guo et al., 2018; Unruh et al., 2018; Piñeiro Fernández et al., 2019; Wong et al., 2019). Nonetheless, one study incorporated the transcriptomes of 13 orchid species representing five subfamilies and 10 phylogenetic informative non-orchid outgroups to build a high-confidence phylogenetic tree of the Orchidaceae from >700 single-copy orthogroups using both concatenation and coalescence-based summary methods (Unruh et al., 2018). This analysis firmly placed the genus *Cypripedium* sister to the rest of the slipper orchid genera (e.g., *Phragmipedium*, *Mexipedium*, *Paphiopedilum*, *Selenipedium*), a previously uncertain relationship. The effectiveness of using transcriptomes to resolve relationships at shallower scales have also been demonstrated. For example, previously uncertain phylogenetic relationships among closely related orchid species (e.g., 8 *Cypripedium*, 4 *Ophrys*, and 5 *Gymnadenia* spp.) have also been clarified using hundreds of putative single-copy genes (Guo et al., 2018; Piñeiro Fernández et al., 2019).

Target Sequence Capture Phylogenomics in Orchids

Putative single-copy genes are already being used to design target enrichment probe sets capable of spanning broad evolutionary scales. Two universal probe sets that provide orchid coverage are: Angiosperms353 (Johnson et al., 2019) and Orchidaceae963 (Eserman et al., 2021). For example, a recent phylogenomic analysis of 75 species representing 69 genera, 16 tribes, and 24 subtribes used the Angiosperms353 universal probe set to target 294 low-copy nuclear genes (Pérez-Escobar et al., 2021). This study revealed that higher-level phylogenetic relationships were robust and largely congruent with earlier plastome- and mitochondrial genome-based phylogenies (Li Y. X. et al., 2019; Serna-Sánchez et al., 2021). However, several instances of strongly supported discordances in both shallow and deep time were also revealed. The adaptability of universal probe sets, such as Angiosperms353, for resolving within-genus relationships of some *Epidendrum* (Granados Mendoza et al., 2020) and *Lepanthes* (Bogarín et al., 2018) orchids, has also been demonstrated. Nonetheless, the use of more targeted probe sets may be necessary to provide greater phylogenetic resolution at the subtribe to subspecific levels, as exemplified by a study of 30 genera spanning the tribe Diurideae and 24 closely related *Caladenia* species (Peakall et al., 2021).

Opportunities for Repurposing Transcriptomes for Phylogenetic Analysis in Orchids

In this perspective, we demonstrate that rapid and accurate phylogenetic insights across the Orchidaceae can be gained from repurposed multi-tissue transcriptomes that were sequenced primarily for non-phylogenetic studies. We conclude by discussing some feasible avenues of research beyond phylogenetic analysis that repurposed multi-tissue transcriptome can offer. In 2021, we conducted an exhaustive search for orchid transcriptome datasets in public repositories and the literature. Our search revealed transciptomes were available for more than 100 orchid species spanning 5 subfamilies, 13 tribes, and 20 subtribes, and drawn from over 50 published studies (**Supplementary Table 1**).

To use this large volume of data effectively, it was necessary to first re-assemble the transcriptomes from the publically accessible raw data. In total, the transcriptomes of 133 target orchid species were reassembled from a total of ∼ 8.5 billion high-quality reads

using *Trinity* (Haas et al., 2013). Next, the outcomes of a gene completeness evaluation using BUSCOs (Waterhouse et al., 2018) was used to gauge assembly quality and to cull samples with poor scores (i.e., < 60%). Finally, generic level representation was restricted to a maximum of two (when available) species per genus (**Supplementary Figure 1 and Text**). Thus, our final high-quality orchid-wide phylotranscriptome data set contained 69 orchids (48 genera) spanning 5 subfamilies, 13 tribes, and 21 subtribes. Four non-orchid Asparagales species (i.e., *Molineria capitulata*, *Hypoxis hemerocallidea*, *Lanaria larata*, *Borya sphaerocephala*) were included as outgroups. Single-copy (strict and relaxed) phylogenetic-informed orthogroups were then identified using OrthoFinder (Emms and Kelly, 2019) and concatenated alignments of protein sequences were evaluated by maximum-likelihood and shortcut-coalescent phylogenetic analysis using IQTREE (Minh et al., 2020) and ASTRAL (Zhang et al., 2018), respectively (see **Supplementary Text and Methods** for additional details).

Taxonomic Representation Across the Phylotranscriptome Data Set

Despite originating from public databases which constrains the choice of samples available, the final phylotranscriptome data set of 69 species achieved wide taxonomic coverage across the Orchidaceae and outgroups (**Figure 1**). This included all known genera of the Apostasioideae and Cypripedioideae subfamilies, while only the tribe Pogonieae lacked representation within the Vanilloideae. Tribes spanning both the lower and higher Epidendroideae were also well covered, with a total of 16 genera, 9 tribes, and 10 subtribes represented. Finally, despite a trend of the Orchidoideae being previously under-represented, this gap has been filled in this study. In particular, genome-scale phylogenetic studies (based on whole plastome, transcriptome, or mitochondria genome) have had poor representation of the diverse Australasian tribe the Diurideae (Givnish et al., 2015; Li Y. X. et al., 2019; Kim et al., 2020; Smidt et al., 2020). Here our phylogenetic analysis included 18 Diurideae species from 10 genera and 8 subtribes. Consequently, this phylotranscriptome dataset provides one of the most comprehensive genera and subtribe representations for the Orchidoideae in phylogenomic studies to date, with a total number of genera, tribes, and subtribes of 22, 3, and 11, respectively (**Figure 1**).

Deep and Shallow Phylotranscriptome Relationships Are Robust and Consistent With Other Phylogenomics Datasets

The outcomes of our phylotranscriptomic analysis (**Figure 1**) revealed robust and largely consistent findings with other phylogenomics datasets (Givnish et al., 2015; Kim et al., 2020; Liu et al., 2020; Eserman et al., 2021; Pérez-Escobar et al., 2021; Serna-Sánchez et al., 2021). As expected, the backbone phylogeny of the *Orchidaceae* revealed five distinct clades: Apostasioideae (4 species) was placed sister to the other four subfamilies, followed by Vanilloideae (4 species), and by Cypripedioideae (7 species) which is placed sister to Orchidoideae (30 species) and Epidendroideae (24 species). Within the Vanilloideae subfamily, *Lecanorchis* was placed sister to a clade formed by *Galeola* and *Vanilla*. Within the Cypripedioideae, *Cypripedium* is basal, followed by *Selenipedium* successively sister to a clade formed by *Phragmipedium*, *Mexipedium,* and *Paphiopedilum* species.

Within the Epidendroideae, for almost all branches leading to the relevant tribes except *Vandeae*, *Epidendreae*, and *Cymbidieae*, the topology was also consistent with other orchid-wide species trees inferred from genome-scale datasets. For example, clear separation between lower (i.e., tribes Neottieae, Gastrodieae, Nervilieae) and higher (i.e., in the order of tribes Arethuseae, Malaxideae, Collabieae, Epidendreae, Vandeae, and Cymbidieae) Epidendroid species were evident. The only case of poor branch support, low gCF and sCF (i.e., proportion of decisive gene trees and alignment sites supporting a given branch supporting a given branch, respectively), and local PP scores pertains to the placement of the tribe Vandeae (i.e., *Phalaenopsis* spp.) in the ML and ASTRAL species tree (**Supplementary Figure 2** and Text for further information). This point of uncertainty is consistent with findings in other recent broad-scale nuclear orchid phylogenomic studies (Eserman et al., 2021; Pérez-Escobar et al., 2021). This now well characterized disparity is postulated to reflect widespread incomplete lineage sorting of nuclear genes during the rapid radiations at the early stages of the evolution of the Vandeae, Epidendreae, and Cymbideae tribes (Eserman et al., 2021).

Within the Orchidoideae, the tribe Orchideae (11 species) is placed sister to a clade formed by Diurideae (13 species) and Cranichideae (6 species) with robust support. Subtribe relationships within Cranichideae were resolved with the subtribe Goodyerinae forming a well-supported clade, sister to *Pterostylis* (subtribe Pterostylidinae). Relatively short branch lengths and low gCF/sCF scores along the backbone and some terminal taxa of the tribes Diurideae and Orchideae were also observed. Within Orchideae, the genus *Habenaria* was the outermost group followed by the genus *Hemipilia* and *Orchis*. Next, the genus *Serapias* forms a clade with *Ophrys* and is placed sister to a clade containing *Platanthera*, *Dactylorhiza*, and *Gymnadenia* species.

Concerning the tribe Diurideae, most subtribe and genus level relationships were congruent with those obtained from a recent phylogenomic study of this tribe using a multi-tiered sequence capture strategy (Peakall et al., 2021). Like the former study, poor support and very low gene concordance factors were observed on the very short but deep branch that separates the sole outermost genus, *Microtis* (Prasophyllinae) from the rest of the tribe. As in the Epidendroideae, such uncertainly likely reflects the incomplete lineage sorting during the rapid radiations underpinning the formation of the major subtribes (Peakall et al., 2021).

To ascertain if robust phylogenetic relationships can be attained at shallower evolutionary depths (i.e., within specific subfamily and genus) from repurposed transcriptomes, samples corresponding to the Cypripedioideae subfamily (19 species) and the genus *Phalaenopsis* (11 species) were targeted. Almost all nodes showed robust support and the topologies closely mirror those inferred using a few chloroplast and low-copy nuclear genes but with broader taxon sampling (Tsai et al., 2010; Guo et al., 2012). As already indicated by other phylotranscriptomic

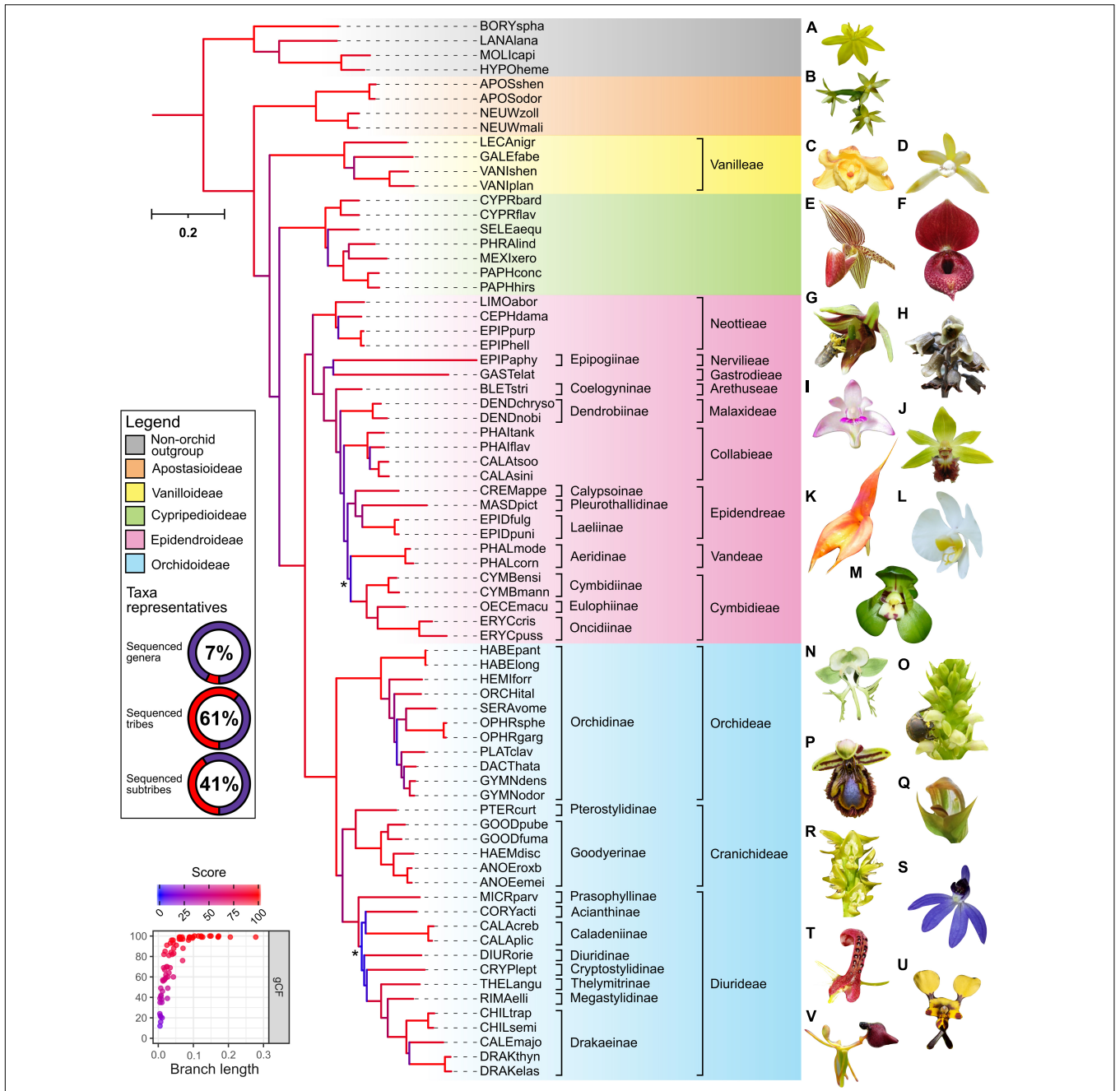


FIGURE 1 | Outcomes of phylogenetic analysis across the Orchidaceae from repurposed multi-tissue transcriptomes. The maximum-likelihood IQ-TREE phylogeny of 69 orchids (spanning 5 subfamilies, 13 tribes, 21 subtribes, and 48 genera) and four non-orchid species as outgroup was based on partition analysis of 633 amino-acid alignments with a total of 317,221 (139,913 parsimony-informative) sites (see **Supplementary Methods** for details). The abbreviated species name is shown (see **Supplementary Table 1** for details). Background color indicates subfamily grouping. The donut chart depict the taxonomic representativeness of included orchid genera, subtribe, and tribe (see relevant placeholder in the phylogeny for details). Asterisk indicate ultrafast bootstrap (UFboot) or SH-aLRT branch support <95 and <80%, respectively. Branch length indicates the total number of substitutions per site and branch colors correspond to gene concordance factors (gCF)–the proportion of gene trees that decisively support the given bifurcation. Inset (bottom left) shows a strong positive relationship of gCF scores with branch length. See **Supplementary Figure 2** for additional details on the corresponding shortcut coalescent ASTRAL phylogeny. Representative species of the outgroup **(A)** *Hypoxis hirsuta* and subfamilies across the Orchidaceae are indicated **(B)** *Apostasia nipponica*–Apostasioideae; **(C)** *Galeola lindleyana* and **(D)** *Lecanorchis suginoana*–Vanilloideae; **(E)** *Paphiopedilum rothschildianum* and **(F)** *Cypripedium lichiangense*–Cypripedioideae; **(G)** *Epipactis veratrifolia*, **(H)** *Gastrodia fontinalis*, **(I)** *Dendrobium kingianum*, **(J)** *Calanthe tricarinata*, **(K)** *Masdevallia veitchiana*, **(L)** *Phalaenopsis amabilis*, and **(M)** *Cymbidium goeringii*–Epidendroideae; **(N)** *Habenaria limprichtii*, **(O)** *Satyrium microrrhynchum*, **(P)** *Ophrys speculum*, **(Q)** *Pterostylis curta*, **(R)** *Prasophyllum elatum*, **(S)** *Cyanicula caerulea*, **(T)** *Cryptostylis leptochila*, **(U)** *Diuris pardina*, and **(V)** *Drakaea glyptodon*–Orchidoideae. All images have been reproduced with permission from the respective copyright holders. Please refer to the section "Acknowledgments" for image credits.

studies (Guo et al., 2018; Unruh et al., 2018; Piñeiro Fernández et al., 2019), our analyses re-confirm the potential for the phylogenetic analysis of transcriptomes to resolve relationships at relatively shallow scales.

The Prospects of Using Transcriptomic Datasets for Phylogenetic Inference and Beyond

While our study has convincingly demonstrated the potential of phylotranscriptomic studies across the Orchidaceae, there are still several major challenges posed across multiple levels. For example, despite our recent exhaustive search of orchid transcriptomes in public repositories, comprehensive coverage spanning: i. many genera, tribes, and subtribes (Li Y. X. et al., 2019; Kim et al., 2020; Serna-Sánchez et al., 2021) or ii. expanded coverage within a particular genus (Guo et al., 2021) is lacking when compared to those from recent plastome phylogenies. Thus, given the size of the Orchidaceae, we recommend targeted sequencing to help fill obvious gaps not yet available in the public datasets. Nonetheless, some phylotranscriptome studies demonstrate that past transcriptomic datasets can be a key resource for increasing taxonomic breadth, which when lacking is known to affect overall phylogenetic accuracy (Zwickl and Hillis, 2002). One example is the phylotranscriptome of Asterids where approximately 40% of the 365 species represented were largely repurposed (Zhang et al., 2020).

In the same way that transcriptomes can be utilized for phylogenetic studies, assembled whole genome sequences may be repurposed for combined phylogenomic/transcriptomic studies of the Orchidaceae. For example, high-depth whole-genome sequencing of 689 vascular plant species from a botanical garden (Liu et al., 2019) contained at least 10 orchid species spanning the genera *Dendrobium*, *Renanthera*, *Vanda*, *Tropidia*, and *Hylophila*, among others, may be used to complement the phylotranscriptomic analysis presented here.

In the Orchidaceae, mycoheterotrophy has evolved sporadically with estimates of some 200 almost fully mycoheterotrophic species, spanning 43 genera (Merckx, 2013). Partial mycoheterotrophy is also likely to be much more widespread across the orchids, than currently documented (Gebauer et al., 2016). Fully mycoheterotrophic orchids often exhibit severely reduced plastid genome, thus, they are often devoid of adequate plastid gene sequences for phylogenetic analysis (Feng et al., 2016). Partially mycoheterotrophic orchid species may also exhibit varying degrees of plastome degradation. In such cases, nuclear DNA sequences may provide the only reliable data for robust phylogenetic analysis (Cameron, 2004; Górniak et al., 2010; Freudenstein and Chase, 2015). Our findings from transcriptomic data show that the placement of *Gastrodia elata* is consistent with earlier orchid-wide phylotranscriptome studies (Deng et al., 2015) while the relationship of *Epipogium aphyllum* (tribe Nervilieae) being sister to *G. elata* (tribe Gastrodieae) appears to be a new finding from this present study. Interestingly, the exceptionally long branch lengths in the clades formed by the mycoheterotrophic *G. elata* and *E. aphyllum* also points to accelerated substitution rates of nuclear genes at a genome-wide scale, a general feature of many parasitic plants (Bromham et al., 2013).

The difficulty in scaling up standard phylogenetic procedures to hundreds, if not thousands of species, will likely persist in future orchid phylotranscriptome studies. Therefore, we take the opportunity to highlight one promising approach using homologous *k*-mer (substring of sequence) blocks that bypass several arduous steps including annotation, ortholog detection, and alignment (Sanderson et al., 2017). Interestingly, Sanderson et al. (2017) recovered trees using homologous *k*-mer blocks and published ones were nearly identical in all seven genome-scale datasets tested containing tens to hundreds of species highlighting its robustness and scalability. Preliminary evaluation of this approach on target sequence capture data encompassing hundreds of loci and samples–67 samples spanning the tribe Diurideae and 72 samples involving 24 closely related *Caladenia* species–revealed its potential to recover species trees matching conventional workflows but also better group morphologically similar members of species complexes (Peakall et al., 2021).

Past and emerging transcriptome datasets can also be mined for the Angiosperms353 (Johnson et al., 2019) and Orchidaceae963 single-copy genes (Eserman et al., 2021) for downstream phylotranscriptomic analysis. This approach bypasses the need for the arduous ortholog inference that often scales poorly with large datasets spanning hundreds of species. However, it remains to be seen if these predefined single-copy genes will be effective in resolving shallow and/or complex evolutionary relationships (e.g., in recent and/or rapid radiating lineages). For the latter purpose, the use of thousands of single-copy genes unique to the study system enabled by targeted studies may be more effective as exemplified in this present study (**Supplementary Figures 3**, **4**) and several others (Guo et al., 2018; Wu et al., 2018; Piñeiro Fernández et al., 2019; Züst et al., 2020; Peakall et al., 2021).

Benefits of Phylotranscriptomics for Gene Duplication and Selection

The prospects of using transcriptomic datasets to gain insights into whole-genome (WGD) or lineage-specific gene duplication events (Zhang et al., 2017; Unruh et al., 2018) and genome-wide detection of positively selected genes of specific evolutionary branches (Balao et al., 2017; Guo et al., 2018; Wong et al., 2019) are also promising avenues of research. For example, two studies have identified a single WGD event shared by all (extant) orchids using transcriptomes from a largely independent list of orchid and non-orchid species (Zhang et al., 2017; Unruh et al., 2018). Lineage-specific gene duplication events leading to the diversification of Cypripedioideae have also been inferred and mapped onto the phylogeny (Unruh et al., 2018).

Signatures of genes under positive selection that may have contributed to species diversification or adaptive evolution in specific evolutionary branches of orchids have also been revealed. For example, elevated evolutionary rates at several candidate volatile biosynthesis-related genes (e.g., fatty acid metabolism) have been independently observed in the unrelated genera of *Chiloglottis* and *Cypripedium* (Guo et al., 2018; Wong et al., 2019).

In *Chiloglottis*, the genes encoding the β-ketoacyl-ACP synthase homologs implicated in the biosynthesis of the 2,5-dialkyl cyclohexane-1,3-dione volatiles (chiloglottones) that attract the specific male wasp pollinators (Peakall et al., 2010), show strong signatures of selection (Wong et al., 2019). In two *Dactylorhiza* orchids, positively selected genes were notably enriched with biotic defense-related pathways related to physical and chemical adaptations (Balao et al., 2017).

CONCLUSION

In this study, we showcase for the first time that a large set of publically available transcriptomes that were sequenced primarily for non-phylogenetic studies can be effectively repurposed to gain phylogenetic insights across broad evolutionary scales of the Orchidaceae. This strategy greatly increased the genus, tribe and subtribe representation of both small (e.g., Cypripedioideae) and larger subfamilies (e.g., Orchidoideae and Epidendroideae) and provided access to thousands of informative sequences (e.g., putative single-copy genes or orthogroups) amendable for phylogenetic inference. We predict that phylotranscriptomic studies will provide an additional platform for future molecular systematic studies and other investigations into the mechanisms driving the extraordinary species diversity of the Orchidaceae.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories *via* NCBI Sequence Read Archive (SRA) accession from respective publications listed in **Supplementary Table 1**. When records of the original publication(s) are not available, the designated Bioproject accession (PRJNA#) is shown.

AUTHOR CONTRIBUTIONS

DW designed the study and analyzed the data. RP and DW secured funding. DW wrote the manuscript with assistance from RP. Both authors have read and approved the manuscript.

FUNDING

This work was supported by the Australian Research Council projects DE190100249 to DW and DP150102762 to RP.

ACKNOWLEDGMENTS

We thank the plant research community for making a wide variety of high-throughput sequencing (especially RNA-seq) datasets publically available. Photographs in **Figure 1** were taken by Zong-Xin Ren (a,c,f,g,j,n), Kenji Suetsugu (b,d,h), Jorun Tharaldsen (p), RP (e,k,l,m,o,r,t,u,v), and DW (i,q,s).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.910362/full#supplementary-material

REFERENCES

Balao, F., Trucchi, E., Wolfe, T. M., Hao, B. H., Lorenzo, M. T., Baar, J., et al. (2017). Adaptive sequence evolution is driven by biotic stress in a pair of orchid species (*Dactylorhiza*) with distinct ecological optima. *Mol. Ecol.* 26, 3649–3662. doi: 10.1111/mec.14123

Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., et al. (2014). Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol. Biol. Evol.* 31, 3095–3112. doi: 10.1093/molbev/msu252

Bogarín, D., Pérez-Escobar, O. A., Groenenberg, D., Holland, S. D., Karremans, A. P., Lemmon, E. M., et al. (2018). Anchored hybrid enrichment generated nuclear, plastid and mitochondrial markers resolve the *Lepanthes horrida* (Orchidaceae: Pleurothallidinae) species complex. *Mol. Phylogenet. Evol.* 129, 27–47. doi: 10.1016/j.ympev.2018.07.014

Bohman, B., Flematti, G. R., Barrow, R. A., Pichersky, E., and Peakall, R. (2016). Pollination by sexual deception - it takes chemistry to work. *Curr. Opin. Plant Biol.* 32, 37–46. doi: 10.1016/j.pbi.2016.06.004

Bromham, L., Cowman, P. F., and Lanfear, R. (2013). Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evol. Biol.* 13:126. doi: 10.1186/1471-2148-13-126

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W., et al. (2015). The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* 47, 65–72. doi: 10.1038/ng.3149

Cameron, K. M. (2004). Utility of plastid psaB gene sequences for investigating intrafamilial relationships within Orchidaceae. *Mol. Phylogenet. Evol.* 31, 1157–1180. doi: 10.1016/j.ympev.2003.10.010

Chao, Y. T., Chen, W. C., Chen, C. Y., Ho, H. Y., Yeh, C. H., Kuo, Y. T., et al. (2018). Chromosome-level assembly, genetic and physical mapping of *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnol. J.* 16, 2027–2041. doi: 10.1111/pbi.12936

Chao, Y.-T., Yen, S.-H., Yeh, J.-H., Chen, W.-C., and Shih, M.-C. (2017). Orchidstra 2.0-A transcriptomics resource for the orchid family. *Plant Cell Physiol.* 58:e9. doi: 10.1093/pcp/pcw220

Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G., van den Berg, C., et al. (2015). An updated classification of Orchidaceae. *Bot. J. Linn. Soc.* 177, 151–174. doi: 10.1111/boj.12234

Chuang, Y. C., Hung, Y. C., Tsai, W. C., Chen, W. H., and Chen, H. H. (2018). PbbHLH4 regulates floral monoterpene biosynthesis in *Phalaenopsis* orchids. *J. Exp. Bot.* 69, 4363–4377. doi: 10.1093/jxb/ery246

De Paolo, S., Salvemini, M., Gaudio, L., and Aceto, S. (2014). De novo transcriptome assembly from inflorescence of *Orchis italica*: analysis of coding and non-coding transcripts. *PLoS One* 9:e102155. doi: 10.1371/journal.pone.0102155

Deng, H., Zhang, G. Q., Lin, M., Wang, Y., and Liu, Z. J. (2015). Mining from transcriptomes: 315 single-copy orthologous genes concatenated for the phylogenetic analyses of Orchidaceae. *Ecol. Evol.* 5, 3800–3807. doi: 10.1002/ece3.1642

Dhiman, N., Sharma, N. K., Thapa, P., Sharma, I., Kumar Swarnkar, M., Chawla, A., et al. (2019). De novo transcriptome provides insights into the growth behaviour and resveratrol and trans-stilbenes biosynthesis in *Dactylorhiza hatagirea* - An endangered alpine terrestrial orchid of western Himalaya. *Sci. Rep.* 9:13133. doi: 10.1038/s41598-019-49446-w

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y

Eserman, L. A., Thomas, S. K., Coffey, E. E. D., and Leebens-Mack, J. H. (2021). Target sequence capture in orchids: developing a kit to sequence hundreds of single-copy loci. *Appl. Plant Sci.* 9:e11416. doi: 10.1002/aps3.11416

Fang, L., Kong, X., Wen, Y., Li, J., Yin, Y., Li, L., et al. (2021). Characterization of embryo and protocorm development of *Paphiopedilum spicerianum*. *Plant Physiol. Biochem.* 167, 1024–1034. doi: 10.1016/j.plaphy.2021.09.001

Feng, Y. L., Wicke, S., Li, J. W., Han, Y., Lin, C. S., Li, D. Z., et al. (2016). Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol. Evol.* 8, 2164–2175. doi: 10.1093/gbe/evw144

Fochi, V., Chitarra, W., Kohler, A., Voyron, S., Singan, V. R., Lindquist, E. A., et al. (2017). Fungal and plant gene expression in the *Tulasnella calospora–Serapias vomeracea* symbiosis provides clues about nitrogen pathways in orchid mycorrhizas. *New Phytol.* 213, 365–379. doi: 10.1111/nph.14279

Freudenstein, J. V., and Chase, M. W. (2015). Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Ann. Bot.* 115, 665–681. doi: 10.1093/aob/mcu253

Gebauer, G., Preiss, K., and Gebauer, A. C. (2016). Partial mycoheterotrophy is more widespread among orchids than previously assumed. *New Phytol.* 211, 11–15. doi: 10.1111/nph.13865

Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., et al. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. R. Soc. B Biol. Sci.* 282:20151553. doi: 10.1098/rspb.2015.1553

Górniak, M., Paun, O., and Chase, M. W. (2010). Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, Xdh: congruence with organellar and nuclear ribosomal DNA results. *Mol. Phylogenet. Evol.* 56, 784–795. doi: 10.1016/j.ympev.2010.03.003

Granados Mendoza, C., Jost, M., Hágsater, E., Magallón, S., van den Berg, C., Lemmon, E. M., et al. (2020). Target nuclear and off-target plastid hybrid enrichment data inform a range of evolutionary depths in the orchid genus *Epidendrum*. *Front. Plant Sci.* 10:1761. doi: 10.3389/fpls.2019.01761

Gravendeel, B., Smithson, A., Slik, F. J. W., and Schuiteman, A. (2004). Epiphytism and pollinator specialization: drivers for orchid diversity? *Philos. Trans. R. Soc. B Biol. Sci.* 359, 1523–1535. doi: 10.1098/rstb.2004.1529

Guo, Y. Y., Luo, Y. B., Liu, Z. J., and Wang, X. Q. (2012). Evolution and biogeography of the slipper orchids: Eocene vicariance of the conduplicate genera in the old and new world tropics. *PLoS One* 7:e38788. doi: 10.1371/journal.pone.0038788

Guo, Y. Y., Yang, J. X., Bai, M. Z., Zhang, G. Q., and Liu, Z. J. (2021). The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* 21:248. doi: 10.1186/s12870-021-03053-y

Guo, Y.-Y., Zhang, Y.-Q., Zhang, G.-Q., Huang, L.-Q., and Liu, Z.-J. (2018). Comparative transcriptomics provides insight into the molecular basis of species diversification of section Trigonopedia (*Cypripedium*) on the Qinghai-Tibetan Plateau. *Sci. Rep.* 8:11640. doi: 10.1038/s41598-018-30147-9

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

Hamabata, T., Kinoshita, G., Kurita, K., Cao, P. L., Ito, M., Murata, J., et al. (2019). Endangered island endemic plants have vulnerable genomes. *Commun. Biol.* 2:244. doi: 10.1038/s42003-019-0490-7

Heyduk, K., Hwang, M., Albert, V., Silvera, K., Lan, T., Farr, K., et al. (2019). Altered gene regulatory networks are associated with the transition from C3 to crassulacean acid metabolism in *Erycina* (Oncidiinae: Orchidaceae). *Front. Plant Sci.* 9:2000. doi: 10.3389/fpls.2018.02000

Hu, C., Yang, H., Jiang, K., Wang, L., Yang, B., Hsieh, T., et al. (2018). Development of polymorphic microsatellite markers by using de novo transcriptome assembly of *Calanthe masuca* and *C. sinica* (Orchidaceae). *BMC Genomics* 19:800. doi: 10.1186/s12864-018-5161-4

Hu, Y., Resende, M. F. R., Bombarely, A., Brym, M., Bassil, E., and Chambers, A. H. (2019). Genomics-based diversity analysis of *Vanilla* species using a *Vanilla planifolia* draft genome and genotyping-by-sequencing. *Sci. Rep.* 9:3416. doi: 10.1038/s41598-019-40144-1

Huang, H., Kuo, Y. W., Chuang, Y. C., Yang, Y. P., Huang, L. M., Jeng, M. F., et al. (2021). Terpene synthase-b and Terpene synthase-e/f genes produce monoterpenes for *Phalaenopsis bellina* floral scent. *Front. Plant Sci.* 12:700958. doi: 10.3389/fpls.2021.700958

Huolin, L., Yuan, T., Wenjing, Y., Liping, L., and Boyun, Y. (2019). De novo transcriptome assembly and analysis of the codon usage bias of the MADS-box gene family in *Cymbidium kanran*. *Indian J. Genet. Plant Breed.* 79, 485–493.

Jąkalski, M., Minasiewicz, J., Caius, J., May, M., Selosse, M. A., and Delannoy, E. (2021). The genomic impact of mycoheterotrophy in orchids. *Front. Plant Sci.* 12:632033. doi: 10.3389/fpls.2021.632033

Jersáková, J., Johnson, S. D., and Kindlmann, P. (2006). Mechanisms and evolution of deceptive pollination in orchids. *Biol. Rev.* 81, 219–235. doi: 10.1017/S1464793105006986

Jiang, Y., Tian, M., Wang, C., and Zhang, Y. (2021). Transcriptome sequencing and differential gene expression analysis reveal the mechanisms involved in seed germination and protocorm development of *Calanthe tsoongiana*. *Gene* 772:145355. doi: 10.1016/j.gene.2020.145355

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., et al. (2019). A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using *k*-medoids clustering. *Syst. Biol.* 68, 594–606. doi: 10.1093/sysbio/syy086

Kim, Y. K., Jo, S., Cheon, S. H., Joo, M. J., Hong, J. R., Kwak, M., et al. (2020). Plastome evolution and phylogeny of Orchidaceae, with 24 new sequences. *Front. Plant Sci.* 11:22. doi: 10.3389/fpls.2020.00022

Lallemand, F., Martin-Magniette, M. L., Gilard, F., Gakière, B., Launay-Avon, A., Delannoy, É, et al. (2019). *In situ* transcriptomic and metabolomic study of the loss of photosynthesis in the leaves of mixotrophic plants exploiting fungi. *Plant J.* 98, 826–841. doi: 10.1111/tpj.14276

Leal, B. S. S., Brandão, M. M., Palma-Silva, C., and Pinheiro, F. (2020). Differential gene expression reveals mechanisms related to habitat divergence between hybridizing orchids from the Neotropical coastal plains. *BMC Plant Biol.* 20:554. doi: 10.1186/s12870-020-02757-x

Lee, Y. I., Chen, M. C., Lin, L., Chung, M. C., and Leu, W. M. (2018). Increased expression of 9-cis-epoxycarotenoid dioxygenase, *PtNCED1*, associated with inhibited seed germination in a terrestrial orchid, *Phaius tankervilliae*. *Front. Plant Sci.* 9:1043. doi: 10.3389/fpls.2018.01043

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., et al. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2

Li, D. M., Wu, W., Zhang, D., Liu, X. R., Liu, X. F., and Lin, Y. J. (2015). Floral transcriptome analyses of four *Paphiopedilum* orchids with distinct flowering behaviors and development of simple sequence repeat markers. *Plant Mol. Biol. Report.* 33, 1928–1952. doi: 10.1007/s11105-015-0886-6

Li, J. W., Zhang, S. B., Xi, H. P., Bradshaw, C. J. A., and Zhang, J. L. (2020). Processes controlling programmed cell death of root velamen radicum in an epiphytic orchid. *Ann. Bot.* 126, 261–275. doi: 10.1093/aob/mcaa077

Li, J., Xu, Y., and Wang, Z. (2019). Construction of a high-density genetic map by RNA sequencing and eQTL analysis for stem length and diameter in *Dendrobium* (*Dendrobium nobile* × *Dendrobium wardianum*). *Ind. Crops Prod.* 128, 48–54. doi: 10.1016/j.indcrop.2018.10.073

Li, X., Luo, J., Yan, T., Xiang, L., Jin, F., Qin, D., et al. (2013). Deep sequencing-based analysis of the *Cymbidium ensifolium* floral transcriptome. *PLoS One* 8:e85480. doi: 10.1371/journal.pone.0085480

Li, Y. X., Li, Z. H., Schuitman, A., Chase, M. W., Li, J. W., Huang, W. C., et al. (2019). Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Mol. Phylogenet. Evol.* 139:106540. doi: 10.1016/j.ympev.2019.106540

Lin, C. S., Chen, J. J. W., Chiu, C. C., Hsiao, H. C. W., Yang, C. J., Jin, X. H., et al. (2017). Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* 90, 994–1006. doi: 10.1111/tpj.13525

Lin, C. S., Hsu, C. T., Liao, D. C., Chang, W. J., Chou, M. L., Huang, Y. T., et al. (2016). Transcriptome-wide analysis of the MADS-box gene family in the orchid *Erycina pusilla*. *Plant Biotechnol. J.* 14, 284–298. doi: 10.1111/pbi.12383

Liu, D. K., Tu, X., Zhao, Z., Zeng, M. Y., Zhang, S., Ma, L., et al. (2020). Plastid phylogenomic data yield new and robust insights into the phylogeny

of *Cleisostoma–Gastrochilus* clades (Orchidaceae, Aeridinae). *Mol. Phylogenet. Evol.* 145:106729. doi: 10.1016/j.ympev.2019.106729

Liu, H., Wei, J., Yang, T., Mu, W., Song, B., Yang, T., et al. (2019). Molecular digitization of a botanical garden: high-depth whole-genome sequencing of 689 vascular plant species from the Ruili botanical garden. *Gigascience* 8:giz007. doi: 10.1093/gigascience/giz007

Liu, S. S., Chen, J., Li, S. C., Zeng, X., Meng, Z. X., and Guo, S. X. (2015). Comparative transcriptome analysis of genes involved in GA-GID1-DELLA regulatory module in symbiotic and asymbiotic seed germination of *Anoectochilus roxburghii* (Wall.) Lindl. (Orchidaceae). *Int. J. Mol. Sci.* 16, 30190–30203. doi: 10.3390/ijms161226224

Lv, X., Zhang, M., Li, X., Ye, R., and Wang, X. (2018). Transcriptome profiles reveal the crucial roles of auxin and cytokinin in the "shoot branching" of *Cremastra appendiculata*. *Int. J. Mol. Sci.* 19:3354. doi: 10.3390/ijms19113354

Ma, X., Yin, X., Tang, Z., Ito, H., Shao, C., Meng, Y., et al. (2020). The RNA degradome: a precious resource for deciphering RNA processing and regulation codes in plants. *RNA Biol.* 17, 1223–1227. doi: 10.1080/15476286.2020.1757898

McKain, M. R., Johnson, M. G., Uribe-Convers, S., Eaton, D., and Yang, Y. (2018). Practical considerations for plant phylogenomics. *Appl. Plant Sci.* 6:e1038. doi: 10.1002/aps3.1038

Meng, X., Li, G., Gu, L., Sun, Y., Li, Z., Liu, J., et al. (2020). Comparative metabolomic and transcriptome analysis reveal distinct flavonoid biosynthesis regulation between petals of white and purple *Phalaenopsis amabilis*. *J. Plant Growth Regul.* 39, 823–840. doi: 10.1007/s00344-019-10025-y

Merckx, V. (ed.) (2013). *Mycoheterotrophy*. New York, NY: Springer. doi: 10.1007/978-1-4614-5209-6

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015

Niu, S. C., Huang, J., Zhang, Y. Q., Li, P. X., Zhang, G. Q., Xu, Q., et al. (2017). Lack of S-RNase-based gametophytic self-incompatibility in orchids suggests that this system evolved after the monocot-eudicot split. *Front. Plant Sci.* 8:1106. doi: 10.3389/fpls.2017.01106

Peakall, R. (2007). Speciation in the Orchidaceae: confronting the challenges. *Mol. Ecol.* 16, 2834–2837. doi: 10.1111/j.1365-294X.2007.03311.x

Peakall, R., Ebert, D., Poldy, J., Barrow, R. A., Francke, W., Bower, C. C., et al. (2010). Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually deceptive *Chiloglottis* orchids: implications for pollinator-driven speciation. *New Phytol.* 188, 437–450. doi: 10.1111/j.1469-8137.2010.03308.x

Peakall, R., Wong, D. C. J., Phillips, R. D., Ruibal, M., Eyles, R., Rodriguez-Delgado, C., et al. (2021). A multitiered sequence capture strategy spanning broad evolutionary scales: application for phylogenetic and phylogeographic studies of orchids. *Mol. Ecol. Resour.* 21, 1118–1140. doi: 10.1111/1755-0998.13327

Pérez-Escobar, O. A., Dodsworth, S., Bogarín, D., Bellot, S., Balbuena, J. A., Schley, R. J., et al. (2021). Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *Am. J. Bot.* 108, 1166–1180. doi: 10.1002/ajb2.1702

Phillips, R. D., Reiter, N., and Peakall, R. (2020). Orchid conservation: from theory to practice. *Ann. Bot.* 126, 345–362. doi: 10.1093/aob/mcaa093

Piñeiro Fernández, L., Byers, K., Cai, J., Sedeek, K. E. M., Kellenberger, R. T., Russo, A., et al. (2019). A phylogenomic analysis of the floral transcriptomes of sexually deceptive and rewarding European orchids, *Ophrys* and *Gymnadenia*. *Front. Plant Sci.* 10:1553. doi: 10.3389/fpls.2019.01553

Ramya, M., Park, P. H., Chuang, Y. C., Kwon, O. K., An, H. R., Park, P. M., et al. (2019). RNA sequencing analysis of *Cymbidium goeringii* identifies floral scent biosynthesis related genes. *BMC Plant Biol.* 19:337. doi: 10.1186/s12870-019-1940-6

Sanderson, M. J., Nicolae, M., and McMahon, M. M. (2017). Homology-aware phylogenomics at gigabase scales. *Syst. Biol.* 66, 590–603. doi: 10.1093/sysbio/syw104

Schelkunov, M. I., Penin, A. A., and Logacheva, M. D. (2018). RNA-seq highlights parallel and contrasting patterns in the evolution of the nuclear genome of fully mycoheterotrophic plants. *BMC Genomics* 19:602. doi: 10.1186/s12864-018-4968-3

Schiestl, F. P., and Schlüter, P. M. (2009). Floral isolation, specialized pollination, and pollinator behavior in orchids. *Annu. Rev. Entomol.* 54, 425–446. doi: 10.1146/annurev.ento.54.110807.090603

Sedeek, K. E., Qi, W., Schauer, M. A., Gupta, A. K., Poveda, L., Xu, S., et al. (2013). Transcriptome and proteome data reveal candidate genes for pollinator attraction in sexually deceptive orchids. *PLoS One* 8:e64621. doi: 10.1371/journal.pone.0064621

Serna-Sánchez, M. A., Pérez-Escobar, O. A., Bogarín, D., Torres-Jimenez, M. F., Alvarez-Yela, A. C., Arcila-Galvis, J. E., et al. (2021). Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Sci. Rep.* 11:6858. doi: 10.1038/s41598-021-83664-5

Smidt, E. D. C., Páez, M. Z., Vieira, L. D. N., Viruel, J., De Baura, V. A., Balsanelli, E., et al. (2020). Characterization of sequence variability hotspots in Cranichideae plastomes (Orchidaceae, Orchidoideae). *PLoS One* 15:e0227991. doi: 10.1371/journal.pone.0227991

Smith, D. R. (2015). Mutation rates in plastid genomes: they are lower than you might think. *Genome Biol. Evol.* 7, 1227–1234. doi: 10.1093/gbe/evv069

Suetsugu, K., Yamato, M., Miura, C., Yamaguchi, K., Takahashi, K., Ida, Y., et al. (2017). Comparison of green and albino individuals of the partially mycoheterotrophic orchid *Epipactis helleborine* on molecular identities of mycorrhizal fungi, nutritional modes and gene expression in mycorrhizal roots. *Mol. Ecol.* 26, 1652–1669. doi: 10.1111/mec.14021

Sun, Y., Wang, G., Li, Y., Jiang, L., Yang, Y., and Guan, S. (2016). De novo transcriptome sequencing and comparative analysis to discover genes related to floral development in *Cymbidium faberi* Rolfe. *Springerplus* 5:1458. doi: 10.1186/s40064-016-3089-1

Swarts, N. D., and Dixon, K. W. (2009). Terrestrial orchid conservation in the age of extinction. *Ann. Bot.* 104, 543–556. doi: 10.1093/aob/mcp025

Trunschke, J., Lunau, K., Pyke, G. H., Ren, Z. X., and Wang, H. (2021). Flower color evolution and the evidence of pollinator-mediated selection. *Front. Plant Sci.* 12:617851. doi: 10.3389/fpls.2021.617851

Tsai, C. C., Chiang, Y. C., Huang, S. C., Chen, C. H., and Chou, C. H. (2010). Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) on the basis of plastid and nuclear DNA. *Plant Syst. Evol.* 288, 77–98. doi: 10.1007/s00606-010-0314-1

Unruh, S. A., McKain, M. R., Lee, Y. I., Yukawa, T., McCormick, M. K., Shefferson, R. P., et al. (2018). Phylotranscriptomic analysis and genome evolution of the Cypripedioideae (Orchidaceae). *Am. J. Bot.* 105, 631–640. doi: 10.1002/ajb2.1047

Valadares, R. B. S., Marroni, F., Sillo, F., Oliveira, R. R. M., Balestrini, R., and Perotto, S. (2021). A transcriptomic approach provides insights on the mycorrhizal symbiosis of the Mediterranean orchid *Limodorum abortivum* in nature. *Plants* 10:251. doi: 10.3390/plants10020251

Valadares, R. B. S., Perotto, S., Lucheta, A. R., Santos, E. C., Oliveira, R. M., and Lambais, M. R. (2020). Proteomic and transcriptomic analyses indicate metabolic changes and reduced defense responses in mycorrhizal roots of *Oeceoclades maculata* (Orchidaceae) collected in nature. *J. Fungi* 6:148. doi: 10.3390/jof6030148

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., et al. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548. doi: 10.1093/molbev/msx319

Wong, D. C. J., Amarasinghe, R., Falara, V., Pichersky, E., and Peakall, R. (2019). Duplication and selection in β-ketoacyl-ACP synthase gene lineages in the sexually deceptive *Chiloglottis* (Orchidaceae). *Ann. Bot.* 123, 1053–1066. doi: 10.1093/aob/mcz013

Wong, D. C. J., Amarasinghe, R., Pichersky, E., and Peakall, R. (2018). Evidence for the involvement of fatty acid biosynthesis and degradation in the formation of insect sex pheromone-mimicking chiloglottones in sexually deceptive *Chiloglottis* Orchids. *Front. Plant Sci.* 9:389. doi: 10.3389/fpls.2018.00839

Wong, D. C. J., Pichersky, E., and Peakall, R. (2017b). The biosynthesis of unusual floral volatiles and blends involved in orchid pollination by deception: current progress and future prospects. *Front. Plant Sci.* 8:1955. doi: 10.3389/fpls.2017.01955

Wong, D. C. J., Amarasinghe, R., Rodriguez-Delgado, C., Eyles, R., Pichersky, E., and Peakall, R. (2017a). Tissue-specific floral transcriptome analysis of the sexually deceptive orchid *Chiloglottis trapeziformis* provides insights into the biosynthesis and regulation of its unique UV-B dependent floral volatile, chiloglottone 1. *Front. Plant Sci.* 8:1260. doi: 10.3389/fpls.2017.01260

Wong, D. C. J., Perkins, J., and Peakall, R. (2022). Anthocyanin and flavonol glycoside metabolic pathways underpin floral color mimicry and contrast in a sexually deceptive orchid. *Front. Plant Sci.* 13:860997. doi: 10.3389/fpls.2022.860997

Wu, M., Kostyun, J. L., Hahn, M. W., and Moyle, L. C. (2018). Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. *Mol. Ecol.* 27, 3301–3316. doi: 10.1111/mec.14780

Xu, D., Chen, H., Aci, M., Pan, Y., Shangguan, Y., Ma, J., et al. (2018). De Novo assembly, characterization and development of EST-SSRs from *Bletilla striata* transcriptomes profiled throughout the whole growing period. *PLoS One* 13:e0205954. doi: 10.1371/journal.pone.0205954

Xu, H., Bohman, B., Wong, D. C. J., Rodriguez-Delgado, C., Scaffidi, A., Flematti, G. R., et al. (2017). Complex sexual deception in an orchid is achieved by co-opting two independent biosynthetic pathways for pollinator attraction. *Curr. Biol.* 27, 1867–1877. doi: 10.1016/j.cub.2017.05.065

Xu, Q., Niu, S.-C., Li, K.-L., Zheng, P.-J., Zhang, X.-J., Jia, Y., et al. (2022). Chromosome-scale assembly of the *Dendrobium nobile* genome provides insights into the molecular mechanism of the biosynthesis of the medicinal active ingredient of *Dendrobium*. *Front. Genet.* 13:844622. doi: 10.3389/fgene.2022.844622

Xu, Y., Jia, R., Zhou, Y., Cheng, H., Zhao, X., and Ge, H. (2018). Development and characterization of polymorphic EST-SSR markers for *Paphiopedilum henryanum* (Orchidaceae). *Appl. Plant Sci.* 6:e01152. doi: 10.1002/aps3.1152

Yu, J., Qiang, W., Qin-Qin, S., Bi-Ping, Z., and Jun-Rong, H. (2020). Transcriptome analysis reveals genes associated with leaf color mutants in *Cymbidium longibracteatum*. *Tree Genet. Genomes* 16:44. doi: 10.1007/s11295-020-01440-4

Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., et al. (2018). The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* 9:1615. doi: 10.1038/s41467-018-03423-5

Yuan, Y., Zhang, J., Kallman, J., Liu, X., Meng, M., and Lin, J. (2019). Polysaccharide biosynthetic pathway profiling and putative gene mining of *Dendrobium moniliforme* using RNA-Seq in different tissues. *BMC Plant Biol.* 19:521. doi: 10.1186/s12870-019-2138-7

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153. doi: 10.1186/s12859-018-2129-y

Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., et al. (2020). Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol. Biol. Evol.* 37, 3188–3210. doi: 10.1093/molbev/msaa160

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

Zhang, J., He, C., Wu, K., Teixeira da Silva, J. A., Zeng, S., Zhang, X., et al. (2016). Transcriptome analysis of *Dendrobium officinale* and its application to the identification of genes associated with polysaccharide synthesis. *Front. Plant Sci.* 7:5. doi: 10.3389/fpls.2016.00005

Zhang, J., Wu, K., Zeng, S., Teixeira da Silva, J. A., Zhao, X., Tian, C. E., et al. (2013). Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. *BMC Genomics* 14:279. doi: 10.1186/1471-2164-14-279

Zhang, L., Chen, F., Zhang, G. Q., Zhang, Y. Q., Niu, S., Xiong, J. S., et al. (2016). Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *Plant J.* 86, 175–185. doi: 10.1111/tpj.13159

Zhang, N., Zeng, L., Shan, H., and Ma, H. (2012). Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* 195, 923–937. doi: 10.1111/j.1469-8137.2012.04212.x

Zhang, Y. M., Su, Y., Dai, Z. W., Lu, M., Sun, W., Yang, W., et al. (2022). Integration of the metabolome and transcriptome reveals indigo biosynthesis in *Phaius flavus* flowers under freezing treatment. *PeerJ* 10:e13106. doi: 10.7717/peerj.13106

Zhang, Y., Zhang, G. Q., Zhang, D., Liu, X. D., Xu, X. Y., Sun, W. H., et al. (2021). Chromosome-scale assembly of the *Dendrobium chrysotoxum* genome enhances the understanding of orchid evolution. *Hortic. Res.* 8:183. doi: 10.1038/s41438-021-00621-z

Zhao, Y., Li, D., and Liu, T. (2019). Pollination-induced transcriptome and phylogenetic analysis in *Cymbidium tortisepalum* (Orchidaceae). *Russ. J. Plant Physiol.* 66, 618–627. doi: 10.1134/S1021443719040174

Zhou, Z., Ying, Z., Wu, Z., Yang, Y., Fu, S., Xu, W., et al. (2021). Anthocyanin genes involved in the flower coloration mechanisms of *Cymbidium kanran*. *Front. Plant Sci.* 12:737815. doi: 10.3389/fpls.2021.737815

Züst, T., Strickler, S. R., Powell, A. F., Mabry, M. E., An, H., Mirzaei, M., et al. (2020). Independent evolution of ancestral and novel defenses in a genus of toxic plants (*Erysimum*, Brassicaceae). *Elife* 9:e51712. doi: 10.7554/eLife.51712

Zwickl, D. J., and Hillis, D. M. (2002). Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51, 588–598. doi: 10.1080/10635150290102339

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fpls.2022.935540

Developmental Characteristics and Auxin Response of Epiphytic Root in *Dendrobium catenatum*

Jili Tian†, Weiwei Jiang†, Jinping Si, Zhigang Han, Cong Li and Donghong Chen**

State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, China

Dendrobium catenatum, a traditional precious Chinese herbal medicine, belongs to epiphytic orchids. Its special life mode leads to the specialization of roots, but there is a lack of systematic research. The aerial root in *D. catenatum* displays diverse unique biological characteristics, and it initially originates from the opposite pole of the shoot meristem within the protocorm. The root development of *D. catenatum* is not only regulated by internal cues but also adjusts accordingly with the change in growth environments. *D. catenatum* root is highly tolerant to auxin, which may be closely related to its epiphytic life. Exogenous auxin treatment has dual effects on *D. catenatum* roots: relatively low concentration promotes root elongation, which is related to the induced expression of cell wall synthesis genes; excessive concentration inhibits the differentiation of velamen and exodermis and promotes the overproliferation of cortical cells, which is related to the significant upregulation of WOX11-WOX5 regeneration pathway genes and cell division regulatory genes. Overexpression of *D. catenatum* WOX12 (DcWOX12) in Arabidopsis inhibits cell and organ differentiation, but induces cell dedifferentiation and callus production. Therefore, DcWOX12 not only retains the characteristics of ancestors as stem cell regulators, but also obtains stronger cell fate transformation ability than homologous genes of other species. These findings suggest that the aerial root of *D. catenatum* evolves special structure and developmental characteristics to adapt to epiphytic life, providing insight into ideal root structure breeding of simulated natural cultivation in *D. catenatum* and a novel target gene for improving the efficiency of monocot plant transformation.

Keywords: ***Dendrobium catenatum*, epiphytic life, aerial root, auxin response, DcWOX12**

OPEN ACCESS

Edited by:
Jen-Tsung Chen, National University of Kaohsiung, Taiwan

Reviewed by:
Randy Ortiz-Castro, National Council of Science and Technology (CONACYT), Mexico
Joseph George Ray, Mahatma Gandhi University, India
Pandiyan Muthuramalingam, Gyeongsang National University, South Korea

***Correspondence:**
Cong Li
congli@zafu.edu.cn
Donghong Chen
donghong.chen@zafu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: *04 May 2022*
Accepted: *30 May 2022*
Published: *23 June 2022*

Citation:
Tian J, Jiang W, Si J, Han Z, Li C and Chen D (2022) Developmental Characteristics and Auxin Response of Epiphytic Root in Dendrobium catenatum. Front. Plant Sci. 13:935540. doi: 10.3389/fpls.2022.935540

INTRODUCTION

Dendrobium catenatum is a traditional rare and precious Chinese medicinal material, which was first described in Shennong's Classic of Materia Medica (Si et al., 2017a,b). Modern pharmacology has confirmed that *D. catenatum* has a wide range of effects, such as enhancing immunity, anti-cancer, protecting liver and stomach, lowering blood sugar, lipid, and blood pressure, anti-fatigue, and anti-oxidation (Si et al., 2017a). Since the 1990s, with the breakthrough of key industrial technologies such as artificial pollination, tissue culture rapid propagation, and cultivation substrate selection, *D. catenatum* resources have changed from wild sporadic mining to large-scale facility cultivation, resulting in the transformation of *D. catenatum* from rare and endangered plants into bulk medicinal materials and formation of 10 billion level of industry (Si et al., 2017a). The yield of *D. catenatum* under facility cultivation is high, but the content of functional components is lower than that in the wild. Therefore, the simulated natural cultivation modes of *D. catenatum*,

such as "living tree epiphytic mode" and "lithophytic cultivation," which take both yield and quality into consideration, came into being (**Figure 1**). Root is closely associated with the epiphytic life of *D. catenatum* in simulated natural cultivation, and investigation on the development and regulation of root is helpful to the breeding of ideal root structure and the guidance of the production practice.

The roots of epiphytic orchids represented by *D. catenatum* are directly exposed to the air, the acquisition of water and nutrition is irregular, and they often face environmental stresses such as drought and heat. The unique lifestyle of epiphytic orchids led to modified aerial root, which is different from the common root characteristics. Epiphytic orchid roots mostly contain chlorophyll, and they are usually in a photosynthetic dormant state, but rapidly activated when wet and go dormant again when dry (Zotz and Winkler, 2013; Sma-Air and Ritchie, 2020). Though epiphytic orchid root still has the similar tissue pattern as the common root anatomically, the outermost tissue is not a single layer of epidermis, but a spongy multilayered velamen radicum, which is composed of dead epidermal cells at maturity, and is capable of rapidly absorbing water and binding cations with very slow water loss rate (Zotz and Winkler, 2013; Hauber et al., 2020). The outstanding ability of roots to absorb and store water is an important feature of epiphytic orchids to survive in water-limited environments. Moreover, the velamen radicum is also considered to be the adaptive characteristics with multiple roles, such as protection against UV-B radiation (Chomicki et al., 2015), and providing a habitat of microbial symbionts associated with nitrogen fixation and the synthesis of hormones and metabolites related to stress tolerance and growth promotion (Deepthi and Ray, 2018, 2020). In general, there is a lack of systematic understanding of the development and regulation mechanism of epiphytic orchid roots. The unique epiphytic characteristics and public reference genome resources of *D. catenatum* make it very suitable as a model system for the study on epiphytic aerial roots.

Plant root systems include embryonic roots, primary roots, lateral roots (LRs), and adventitious roots (ARs), playing critical functions in fixing, absorption, and storage. In model plant Arabidopsis, embryo roots originate from hypophysis as the uppermost cell of the suspensor at the globular stage during embryogenesis (Chen et al., 2019). During post-embryonic stage, primary roots derive from the continuous growth of embryonic roots and breakthrough of testis, and LRs are developed from LR primordia that are usually endogenously initiated from xylem pole pericycle cell in a developing root. ARs are regenerated from non-root organs (e.g., hypocotyl, stem, or leaf) or very old roots (Ge et al., 2019), which are typically induced and activated by diverse stresses and developmental cues, provide the adaption plasticity to ever-changed environments during the evolution of land plants, and are considered to be a strong natural selective trait (Lardon and Geelen, 2020). ARs mainly arise from pericycle-like cells neighboring vascular tissues and alternatively from phloem or xylem parenchyma cells, young secondary phloem cells, and procambium/cambium cells within vascular tissues (Lakehal and Bellini, 2019). Epiphytic orchids are adventitious root-based and do not make later roots (Shekhar et al., 2019).

Adventitious root is produced from non-root organs upon exogenous and endogenous cues and is a process of *de novo* root regeneration (DNRR) (Xu, 2018). Auxin is the major growth-promoting hormone for AR initiation. Auxin application by suitable concentration is widely used in asexual reproduction (e.g., cutting, air layering, and tissue culture) for DNRR in horticulture and forestry. The regulatory mechanisms of auxin are primarily orchestrated through biosynthesis, conjugation, polar transport, and signal transduction (Xie et al., 2021). For instance, overexpression of auxin receptor TIR1 and transcription factor AUX1 promotes AR development (Street et al., 2016). During DNRR, the regeneration-competent cells (e.g., pericycle-like) in explant experience fate transition guided by auxin (Yu et al., 2017; Xu, 2018). Cell fate transition from regeneration-competent cells to AR formation is an orderly and organized process, further divided into four steps: priming, initiation, patterning, and emergence (Xu, 2018). In addition, a high level of exogenous auxin can result in callus induction, which occurs from pericycle-like cells *via* a root developmental program (Sugimoto et al., 2010), similar to the DNRR (Liu et al., 2014). Both AR and callus regeneration events involve two cell fate transition steps: priming step from regeneration-competent cells to root founder cells (marked by WOX11/12), and initiation step from founder cells to root primordium or callus with remarkable root primordium-like feature (marked by WOX5), which is maintained by the constantly high endogenous auxin level (Hu and Xu, 2016; Sheng et al., 2017). Auxin distribution in callus mass is uneven, resulting in the entry of some cells with low auxin levels into patterning step in advance and the expression of RAM identity genes. Therefore, callus consists of a group of heterogeneous cells with root primordium features and partially differentiated RAM traits.

Here, we systematically revealed the developmental characteristics, origin, adaptive changes to the environment, and high tolerance and response to auxin of *D. catenatum* roots, providing guidance for production practice, and identified *D. catenatum* WOX12 high-efficient dedifferentiation-inducible property, providing novel candidate gene resource for regeneration and plant transformation.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The *D. catenatum* cultivars (JP966, JP7869, and D6) and *D. nobile* from Zhejiang A&F University and wildtype Arabidopsis Col-0 were used in this study. For Arabidopsis growth, seeds were surface sterilized (70 and 95% ethanol each for 10 min) and plated on 1/2 MS medium. After stratification at 4°C for 2 days in the dark, plates were incubated in a growth chamber at 22°C under a 16-h light/8-h dark photoperiod.

For *D. catenatum* tissue culture, seedlings were grown in greenhouse at a temperature of 25°C and 16-h light/8-h dark photoperiod. For pine bark pot cultivation and rock/trunk epiphytic cultivation of *D. catenatum*, all samples were grown in Hengtang Medicine-Expo Garden, Tianmu Mountain Eco-Agriculture Demonstration Zones, Hangzhou,

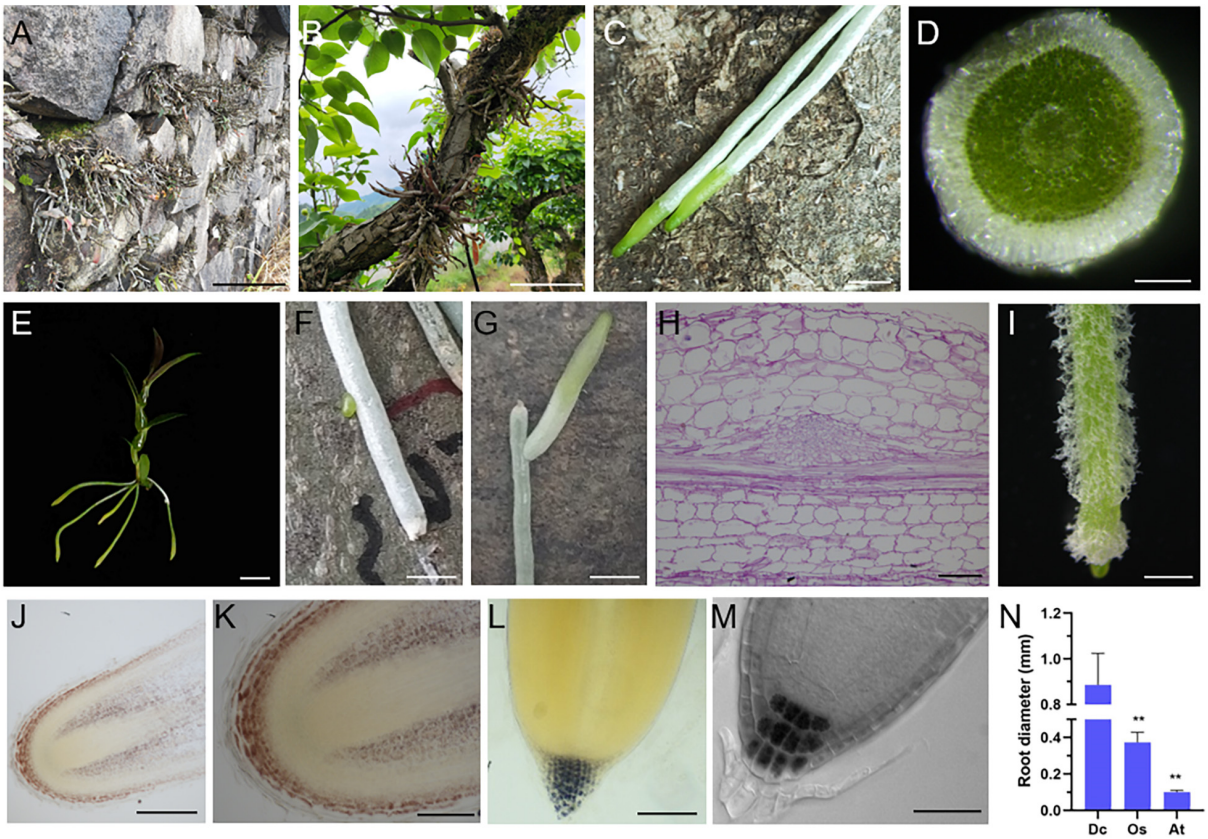


FIGURE 1 | Unique biological characteristics of *D. catenatum* root. **(A)** Rock epiphytic mode, **(B,C)** trunk epiphytic mode showing the aerial roots attached to a living tree **(C)**; **(D)** manual cross-section of an aerial root; **(E)** a plantlet of *D. catenatum* is absent of lateral roots. **(F,G)** lateral root production near the injury site; **(H)** paraffin section of lateral root primordia displayed by PAS staining; **(I)** root hairs from a tissue culture *D. catenatum* plantlet. **(J–L)** Iodine staining of roots in seedlings of *D. catenatum* **(J,K)**, rice **(L)**, and Arabidopsis **(M)**; **(N)** comparison of root diameters among *D. catenatum*, rice, and Arabidopsis. Dc, *D. catenatum*; Os, *Oryza sativa*; At, *Arabidopsis thaliana*. $^{**}p < 0.01$ (two-sample *t*-test, compared with Dc). Bars = 5 cm in **(A,B)**, 0.25 cm in **(C**,**F,G)**, 250 μm in **(D,J)**, 1 cm in **(E)**, 50 μm in **(H)**, 1 mm in **(I)**,100 μm in **(K,L)**, 25 μm in **(M)**.

Zhejiang Province, China (119°26′11″E, 30°20′30″N, 280 m above sea level).

For *D. catenatum* seed germination, a mature capsule from JP966 cultivar was subjected to surface sterilization including 75% ethanol for 1 min and 2% $HgCl_2$ for 10 min and washed by sterilized water for at least 5 times. Subsequently, the seeds from sterilized capsule were sowed on the germination medium (1/2 MS + 0.1 mg/L NAA + 0.5% active carbon + 3% sucrose + 0.8% agar, pH5.8).

For auxin treatment, 6-month-old *D. catenatum* JP966 tissue culture seedlings with the removal of the roots were inoculated to auxin-free banana medium (MS + 10% banana pulp + 3% sucrose + 0.8% agar, pH5.8) containing different concentrations of NAA (0, 2, 5, 10, and 20 $mg{\cdot}L^{-1}$).

For overexpressing *DcWOX12-GFP* in transgenic Arabidopsis, the full-length ORF of *DcWOX12* (LOC110108675) without stop codon was amplified from a cDNA pool of the *D. catenatum* seedlings and inserted into the modified pCAM1300-GFP vector, using the following primers: 5′-ggatccATGGAAGAGGAGAAGCCCCC-3′ and 5′-gtcgacGCTC CGGTTATCACCAAAGAAGT-3′. The construct was then introduced into Col-0 wildtype Arabidopsis plants *via Agrobacterium tumefaciens*-mediated transformation.

Histology and Microscopy

Vibratome sectioning was performed as described (Henry et al., 2015). Briefly, fresh root samples were directly embedded in 3% agarose blocks and then were glued on a vibratome plate to be sliced. Transverse sections were obtained using a Leica VT1200/VT1200S vibration microtome (Leica, Germany) with speed 0.01–1.5 mm/s, amplitude 0.8 mm, and thickness 50 μm. On the one hand, the obtained slices were used for direct observation or stained with 0.1% aniline blue solution. Lignin staining with phloroglucinol-HCl reagent was performed as previously described (Joca et al., 2020). Starch staining with Lugol's solution and fat red staining with Fat Red 7B were performed as previously described (Chen et al., 2010). On the other hand, the slices were transferred either onto chamber slides (Lab-teak 177402) for immunolocalization.

Immunolocalization assay was performed as previously described (Henry et al., 2015). Briefly, the sections in the chamber

slides experienced the rinse with 0.1 M glycine and phosphate-buffered saline (PBS) solution, blocking with 5% bovine fetal serum, incubation with corresponding the primary and secondary antibodies. Several specific cell wall-directed probes available from PlantProbes[1] were used as primary antibodies to evaluate the presence of major cell wall polymers including (1→4)-β-galactan (LM5) (Andersen et al., 2016), heteromannan (LM21) (Marcus et al., 2010), and pectin (JIM5, JIM7) (Clausen et al., 2003). Secondary antibodies include Alexa 546 anti-rat antibody (Invitrogen A11081) or Alexa 546 anti-mouse antibody (Invitrogen A11060). Confocal observation was performed with a Zeiss LSM 880 confocal microscope (ZEISS, Germany).

Paraffin sectioning was performed as previously described (Chen et al., 2016), Briefly, *D. catenatum* root samples were fixed with FAA solution for at least 24 h, dehydrated in gradient ethanol/tert-Butyl alcohol solutions, and embedded in paraffin. The paraffin tissue blocks were sectioned to 6 μm by a rotary microtome (Leica, Germany). Then, the paraffin sections were stained by safranin O-fast green staining *via* Safranin O-Fast Green stain kit (Solarbio, China) or PAS staining *via* glycogen periodic acid Schiff (PAS/hematoxylin) stain kit (Solarbio, China).

Differential interference contrast (DIC) observation was performed as previously described (Chen et al., 2016). Plastic sectioning and transmission electron microscopy (TEM) observation were performed as previously described (Yu et al., 2009). All conventional light microscopy was performed with Olympus SZX16 or Olympus BX60 microscopes (Olympus, Japan).

RNA-Seq and Data Analysis

For auxin treatment, the root samples of *D. catenatum* plantlets treated with different NAA concentrations (0, 5, 10, and 20 mg L^{-1}) for 3 months were harvested for RNA-seq. For DcWOX12 function analysis, 2-week-old Arabidopsis seedlings of *DcWOX12-GFP* overexpression transgenic line and Col-0 wildtype were collected for RNA-seq. A total of three independent RNA samples (biological replicates) were performed. RNA was extracted using the MiniBEST Plant RNA Extraction Kit (TaKaRa, Japan) according to the manufacturer's instructions. The paired-end sequencing was performed on an Illumina HiSeq 4000. Reads of above-described samples were aligned to the genomes *D. catenatum* or Arabidopsis using HISAT package (Kim et al., 2015), which initially remove a portion of the reads based on quality information accompanying each read and then maps the clean reads to the reference genome. StringTie was used to assemble the mapped reads of each sample and estimate the expression levels of all genes by calculating FPKM (Pertea et al., 2015). The differentially expressed genes were defined with $|\log2(\text{fold change})| \geq 1$ and with an adjusted p-value (q-value) < 0.05 by R package (Anders and Huber, 2010). Heatmap was generated using TBtools software (Chen et al., 2020). Differentially expressed genes (DEGs) by auxin treatments were clustered with STEM (Ernst and Bar-Joseph, 2006) based on OmicShare platform for data analysis[2]. Gene Ontology (GO) enrichment in DEGs was identified compared to the reference transcriptome background using OmicShare tools.

Statistical Analysis

All data were statistically analyzed using IBM SPSS Statistics 26 software and presented as the mean ± SE. Independent-samples t-test was used to calculate significant differences between two groups of data, which are denoted by asterisks ($^{*}p < 0.05$; $^{**}p < 0.01$).

RESULTS

Unique Biological Characteristics of *D. catenatum* Root

Dendrobium catenatum is a typical epiphytic orchid, which usually grows on the surface of forest trunk or cliff in natural environment (**Figures 1A,B**). Due to its unique lifestyle, the root of *D. catenatum* is highly specialized. When exposed to the air, the root tip (representing meristem zone and transition zone) keeps green, and the rest of root (representing mature zone) looks white (**Figure 1C**), which is unique to epiphytic orchids. It is caused by the spongy velamen radicum that consists of multilayered epidermal cells, becomes dead at maturity (**Figure 1D**), and has protection and water storage function. However, the internal tissues are still green (**Figure 1D**), indicating that the root of *D. catenatum* has nutritional function and can synthesize organic compounds. *D. catenatum* usually grow without lateral roots and root hairs (**Figure 1E**), which may be due to direct exposure to the air and does not need to increase the contact area with the soil. However, when the root was injured, lateral root primordia would be induced near the wound, and lateral roots would grow to maintain the function of the mother root (**Figures 1F–H**). When sterilely growing in culture bottle, root hairs can be induced by high humidity to enhance the absorption function of roots (**Figure 1I**). In the epiphytic mode of *D. catenatum*, the gravitropism of roots is not obvious. It is speculated that the roots in this case adhere to the surface of trunk or rock by secreting special polysaccharides, such as glucomannan. Iodine staining also showed that the entire root cap region of *D. catenatum* contained starch granules, evidently different from the common terrestrial plants including Arabidopsis and rice, in which starch granules were only confined to columella cells in the root cap (**Figures 1J–M**). In addition, the root of *D. catenatum* was thick and sturdy (**Figure 1N**), with an average diameter of around 1.2 mm, much thicker than that of rice (∼0.3 mm) or Arabidopsis (∼0.1 mm), and even had starch deposition in cortex parenchyma cells (**Figures 1J,K**), and these properties lead to structural rigidity and insensitivity to gravity. So, the specialization of morphological characteristics in *D. catenatum* is closely related to its epiphytic life mode.

Developmental Origin of *D. catenatum* Root

Different from the fact that the root cell pattern of the representative model plant Arabidopsis has been completely established during embryogenesis, the embryos in the mature

[1]http://www.plantprobes.net/

[2]http://www.omicshare.com/tools

seeds of *D. catenatum* are poorly developed and arrest at a very early stage comparable to the globular stage of Arabidopsis embryogenesis (Fang et al., 2016), resulting in no establishment of polarized embryo axis and no production of radicle. Later on, orchid embryos continue to develop *in vitro* without following traditional embryogenesis procedures, and instead proliferate to form a unique protocorm structure that is sometimes considered to be a continuum of zygotic embryogenesis (Jones and Tisserat, 1990), but is actually distinct from zygotic embryonic tissue at molecular level (Fang et al., 2016). To trace the embryonic origin of *D. catenatum* root, we defined the developmental stages during seed germination based on the landmark developmental events. At first, the embryo in the tiny seed (**Figure 2A**) experienced morphologically apolar enlargement, resulting in rupture of the testa and protocorm formation (stage I, **Figure 2B**). Subsequently, appearance of the rhizoid leads to establishment of morphological polarity (stage II, **Figure 2C**). The apical shoot enclosed by coleoptile was developed (stage III, **Figure 2D**), and stage IV is the characteristics of the appearance of the first true leaf due to separate from the unfolded coleoptile (**Figure 2E**). Then, the first adventitious root occurred from the base (stage V, **Figure 2F**), indicating the developmental phase from protocorm to plantlet. The secondary true leaf appeared, suggesting the stem node formation (stage VI, **Figure 2G**). As more roots and leaves appear, tillering occurred (stage VII, **Figures 2H,I**). Further cytological observation (**Figures 2J–O**) found that during the polarity establishment of the protocorm, shoot meristem surrounding by young leaves (named shoot body) was first generated at the shoot pole, and then, root meristem was generated at the opposite pole, where root protruded besides the base of protocorm. Protocorm seems to be a temporary and compensatory organ that appears when orchid seeds lack endosperm or cotyledon to supply energy for germination. It can not only effectively protect shoot body, but also accumulate starch granules, provide habitat for germination promoting endophytes, and produce rhizoids to absorb certain nutrients from the outside. Therefore, the root of *D. catenatum* originates endogenously from the root meristem opposite to shoot body enclosed by the protocorm.

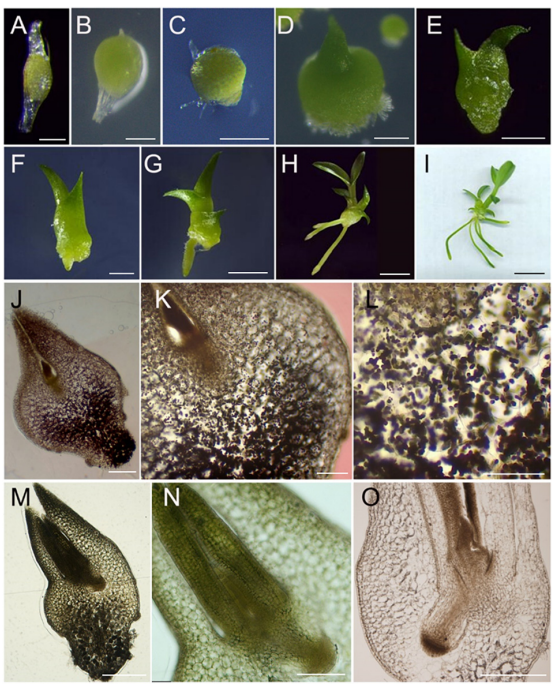


FIGURE 2 | Developmental origin tracing of *D. catenatum* root. **(A–I)** Seed germination and seedling development. **(A)** 0 DAG (day after germination), **(B)** 10 DAG, **(C)** 20 DAG, **(D)** 40 DAG, **(E)** 80 DAG, **(F)** 100 DAG, **(G)** 120 DAG, **(H)** 230 DAG, **(I)** 260 DAG; **(J–O)** Anatomic observation. **(J–L)** 30 DAG, **(M–O)** 40 DAG, **(O)** 80 DAG. Bars = 100 μm in **(A**,**B**,**K)**, 250 μm in **(C)**, 500 μm in **(D**,**M**,**O)**, 0.1 cm in **(E**,**F)**, 0.2 cm in **(G)**, 4 cm in **(H)**, 1 cm in **(I)**, 200 μm in **(J**,**N)**, 50 μm in **(L)**.

Cytological Changes in *D. catenatum* Root Structure During Growth and Development

To investigate the cytological changes in *D. catenatum* root during development, we performed the paraffin section on the root tips of JP966 tissue culture seedlings (**Figure 3**). According to the cross-section observation of different regions in root tip, meristematic zone (**Figure 3A**) and transition zone (**Figure 3B**) in a low differentiation level roughly included root cap, epidermis, ground tissue, and vascular bundle without distinct boundaries between adjacent cell types. The epidermis cells are still living and exodermis is undistinguishable due to the lack of obvious thickening Casparian strip, and the cells are smaller with compact arrangement and distinct nuclei. The mature zone (**Figure 3C**) was clearly divided into velamen, exodermis, cortical parenchyma, endodermis, and stele. The velamen is irregular and consists of multiple layers of dead cells. The exodermis is a single layer of lignified sclerenchyma cells (large) with horseshoe-shaped thickening of the cell wall intercalated by passage cells. The cortex is multilayered, where the cells close to both sides are smaller, and those in the middle are larger, oval in cross-section, rectangular in longitudinal section, with gaps among cells, and inconspicuous nuclei. The endodermis is a single layer of sclerenchyma cells (small) and interspersed with passage cells, and much less developed than exodermis. The stele belongs to radical type, and the structure of the pericycle is not obvious. Longitudinal section showed that the root cap surrounding the RAM is underdeveloped (**Figure 3D**) and well organized cell pattern in mature zone (**Figure 3E**).

To investigate the change rule of the root structure of *D. catenatum* during growth, we observed different segments of mature zones at the same root of the JP966 tissue culture seedlings by vibratome sectioning and phloroglucinol staining (**Figure 4**). Phloroglucinol specifically stains lignin, a hallmark component of the secondary cell wall. The lack of staining in the RAM indicates that there is no secondary wall formation at this time (**Figure 4B**), which is consistent with high cell division activity and low cell differentiation. Overall, the maturation zones at different root segments have similar pattern (**Figures 4C–G**). However, by comparison of the cross-sections

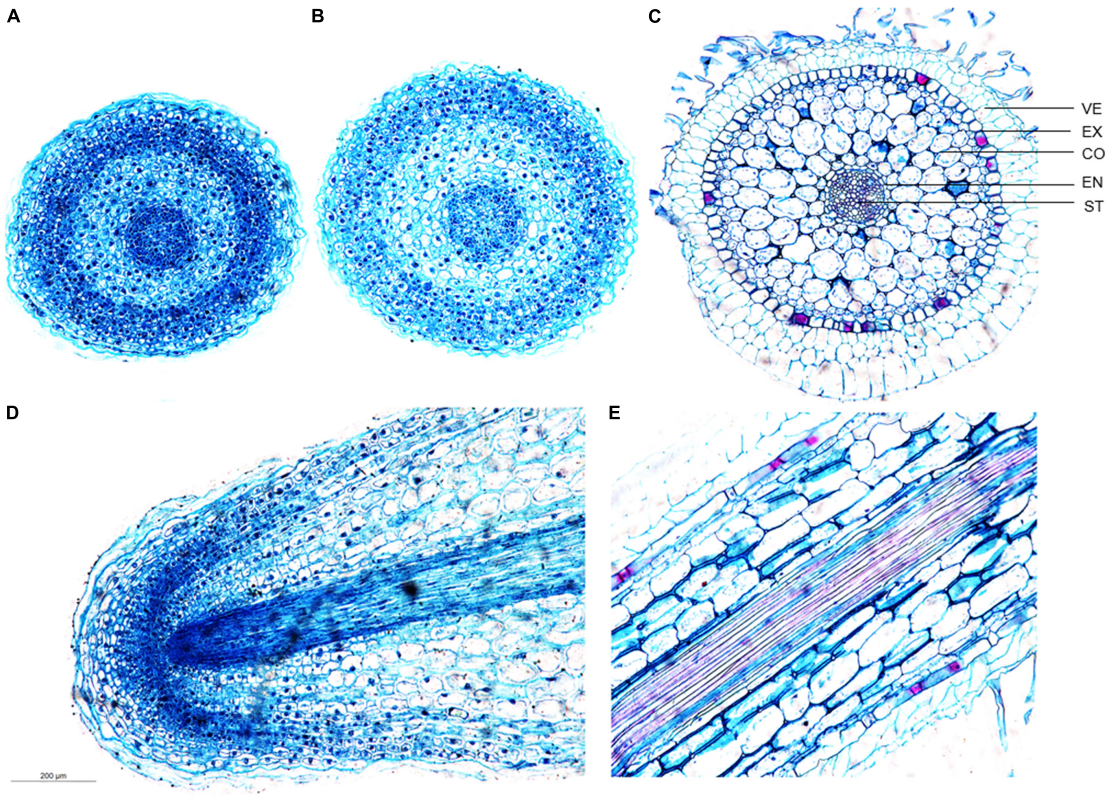


FIGURE 3 | Developmental differences of *D. catenatum* root structure. **(A)** Cross-section of the meristem zone, **(B)** cross-section of the transition zone, **(C)** cross-section of the mature zone, **(D)** longitudinal view of the root apex, **(E)** longitudinal view of the mature zone. VE, velamen; EX, exodermis; EN, endodermis; ST, stele.

of the representative mature zones (MZs) near the RAM (MZ1), in the middle (MZ2∼4) and near the stem base (MZ5), we found that the diameter of the whole root, velamen, cortex, and stele of different segments of a single root were different (**Figures 4C–L**). The cross-sectional widths of MZ2/3/4 were significantly larger than those of MZ1 and MZ5 (**Figure 4M**); correspondingly, the number of cell layers in the velamen and cortex: MZ2/3/4 > MZ1/5 (**Figure 4N**). The number of xylem ridges was the least in the earliest MZ5 (4), which was the most stained and lignified and then slowly stabilized (7) (**Figure 4O**). Therefore, the pattern produced by *D. catenatum* root at the beginning of formation (MZ5) is relatively simple and then tends to be stable (MZ3∼5), whereas the newly differentiated mature zone (MZ1) near the root tip is not yet fully differentiated, indicating that structure of the same root varies during growth.

Immunolocalization of *D. catenatum* Aerial Roots

Plant cell wall mainly comprises cellulose, hemicellulose (e.g., xyloglucans, xylans, mixed-linkage glucans, and mannans) and pectin [e.g., homogalacturonan (HG), rhamnogalacturonan-I (RG-I), and RG-II] (Verhertbruggen et al., 2009; Marcus et al., 2010). To identify whether the chemical composition has changed in the specialized *D. catenatum* root or not, specific polysaccharide molecular probes were used for detection (**Figure 5**). In mature zone, anti-mannan antibody LM21 recognizes the β-linked mannan polysaccharide (Marcus et al., 2010), which was localized on the cell wall of all root cell types of *D. catenatum* (**Figures 5A–C**). Anti-pectin antibodies JIM5 and JIM7, respectively, recognize low and high levels of methyl-esterified HGs (Willats et al., 2000; Clausen et al., 2003; Henry et al., 2021) and had similar immunolabeling pattern in cortex and stele, and no labeling in velamen, exodermis, and endodermis (**Figures 5D–I**). The difference was that JIM5 had a strong signal in the cortex tricellular junction delta region (**Figure 5F**), whereas JIM7 had the strong labeling in the vertices of this delta region (**Figure 5I**), where high methyl-esterified HGs have a strong ability to form gels in the cell wall matrix, increase cell wall porosity, and might facilitate the transport of solutions among cortical cells (Joca et al., 2020). Anti-pectin probe LM5 recognizes (1-4)-β-D-galactan side chains of RG-I (Andersen et al., 2016) and had strong signal in cortex and phloem, moderate signal in exodermis, endodermis, and stele (excluding phloem), but no signal in velamen (**Figures 5J–L**). In addition, LM5 epitope had labeling in the undifferentiated cells of meristem zone and transition zone, but no labeling in the root cap in the root tip (**Figures 5M–R**). Therefore, the content and domain of pectic polysaccharides are closely related to the root tissue types and developmental status.

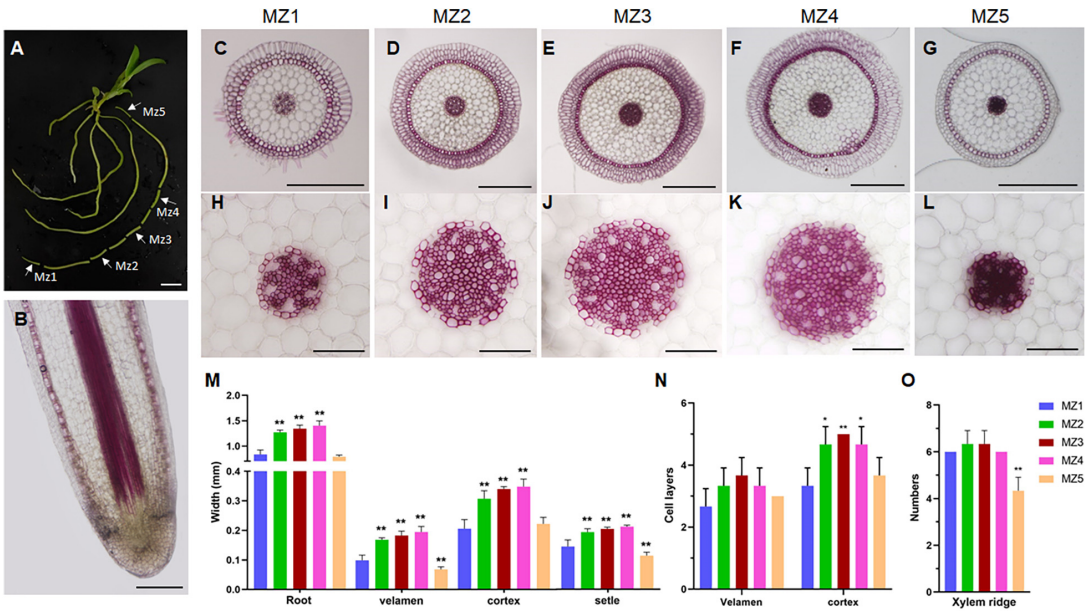


FIGURE 4 | Structural differences of maturation zones during *D. catenatum* root growth. **(A)** Sampling sites of root. White arrows indicated different mature zones (MZ1 to MZ5). **(B)** Longitudinal section of root apex. **(C–L)** Cross-section of mature area of different parts of root. **(H–L)** Steles of different parts of root, Bars = 100 μm. **(M)** The diameter of the root and the width of the velamen, endodermis, and stele. **(N)** The number of cell layers of the velamen and cortex. **(O)** The xylem ridges in different parts. MZ, mature zone. $^{**}p < 0.01$ (two-sample *t*-test, compared with MZ1). Bars = 1 cm in **(A)**, 25 μm in **(B)**, 500 μm in **(C–G)**, 100 μm in **(H–L)**.

Effects of Growth Environment and Germplasm on Root Development

To explore the effects of different growth environments on the root development of *D. catenatum*, we compared the root structure grown in distinct conditions (**Figure 6**). Under pot cultivation mode using pine bark substrate (**Figure 6A**), root hair was absent on the root surface far away from the substrate (**Figure 6B**), similar to root status stretching freely in the air; the velamen cells close to the substrate but not in contact may produce root hairs (**Figures 6C,D**), and the cell wall of velamen and root hair displayed LM5 labeling, indicating that they are living cells; the root closely attached to the pine bark had small and dense cells at the attachment site (**Figure 6C**), which is similar to the specific velamen cells attached to the truck surface (**Figures 6E,F**). It is speculated that they are living cells, responsible for plant attachment under tree/rock epiphytic modes. Under the condition of tissue culture (**Figures 6F–H**), the roots from 4-month-old seedlings displayed many root hairs, but half of the thickness of those from 12-month-old plants. The thickness of each velamen cell layer growing in the substrates (e.g., pine bark and agar medium) is similar (**Figures 6B–H**), but the outermost cell layer is about 1–2 times wider than the other layers in the velamen of exposed root (**Figure 6F**), which may be related to its highly efficient water absorption and retention capacity, and provides mechanical protection for the inner living cells. Therefore, the change in growth environment will lead to the corresponding adjustment of velamen structure to better adapt to the diverse environments.

To identify the germplasm on the root development of *D. catenatum*, we compared the root structures exposed to the air in three *D. catenatum* variants (JP966, JP7869, and D6) with *D. nobile* as a reference (**Figures 6J–O**). The diameter of aerial roots and the width of velamen, cortex, and stele were the largest (∼ 1.5 mm) in *D. nobile*, followed by JP966 and JP7869, and the smallest in D6. These suggested that root cell pattern can serve as an indicator to distinguish different germplasms and species.

Effects of Auxin on Root Development of *D. catenatum*

During *Dendrobium* seedling growth and propagation, 0.1–2 mg·L^{-1} NAA is usually added to the medium during aseptic culture (da Silva et al., 2015). Auxin-producing endophytes in the velamen roots are beneficial for the root colonization of epiphytic orchids (Faria et al., 2013; Pavlova et al., 2017; Shah et al., 2018, 2021, 2022; Deepthi and Ray, 2019; Chand et al., 2020). To explore the effect of auxin on the root development in *D. catenatum*, we transferred JP966 seedlings with the removal of roots to the medium containing different concentrations of NAA (0, 2, 5, 10, and 20 mg L^{-1}), the phenotype was observed and recorded after 90 days of cultivation (**Figure 7**). We found that auxin treatment had two effects: when NAA concentration was lower than 5 mg L^{-1}, the plant biomass, root length, and rooting number of *D. catenatum* seedlings were positively correlated with auxin concentration compared with the control (0 mg L^{-1} NAA). Under the condition of high concentrations of NAA (>5 mg L^{-1}), with the increase of auxin concentration, the root became shorter, but significantly thicker, and even formed tumor-like

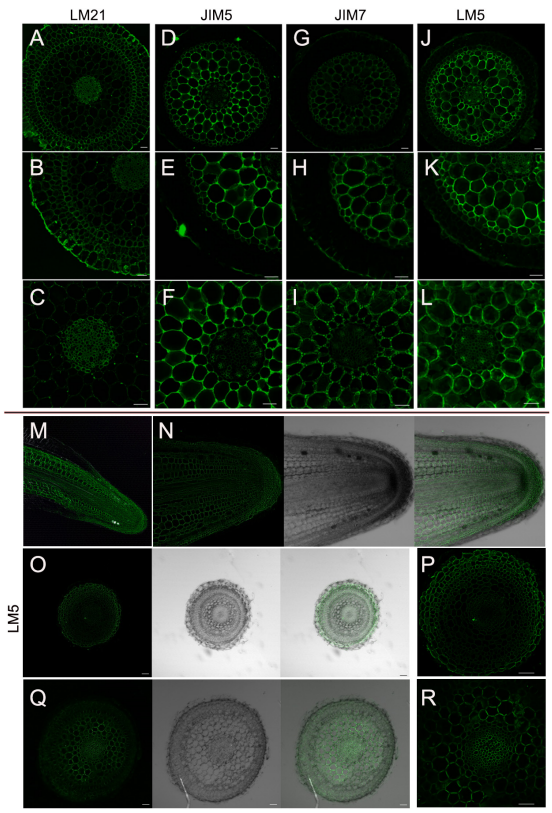


FIGURE 5 | Immunolocalization of *D. catenatum* aerial roots. **(A–L)** Immunolabeling in the mature zone. **(A–C)** LM21 antibody, **(D–F)** LM5 antibody, **(G–I)** JIM5 antibody, **(J–L)** JIM7 antibody. **(M–R)** LM5 epitope in root tip. **(M)** Longitudinal section of root apex, **(N)** longitudinal section of meristem, **(O,P)** transverse section of meristem zone, **(Q,R)** transverse section of transition zone. Bars = 50 μm.

structures under the treatment of 20 mg L^{-1} NAA (**Figures 7A–D**). Cytological analysis further revealed that high concentration of exogenous auxin (10 mg L^{-1} NAA) significantly increased the cortical cell number (∼ 11 layers), compared with ∼ 4 layers in the control, but decreased the width (∼ 0.06 mm) and cell layers (∼2) of velamen, compared with ∼ 0.2 mm width and 3 cell layers in the control, respectively (**Figures 7E–M**). The exodermis under 10 mg L^{-1} NAA treatment also had smaller cells with significantly decreased lignin staining (**Figures 7I–J**). Therefore, high concentration of auxin will cause the rapid proliferation of cortical parenchyma cells and inhibit the differentiation of velamen and exodermis.

Transcriptional Response of *D. catenatum* Root to Auxin Treatment

To analyze the transcriptional response of *D. catenatum* by auxin application, root samples from four NAA treatments (0, 5, 10, and 20 mg L^{-1}) were collected for RNA-seq analysis. Gene expression level in each sample was calculated and normalized to FPKM. The differentially expressed genes (DEGs) among these four NAA treatment groups were identified by a false discovery rate (FDR) < 0.05 and a $|\log2(\text{fold change})| \geq 1$. There was a total of 4,448 DEGs among comparison of four treatments (**Figures 8A–C**).

To further investigate the expression patterns and functions of DEGs among these auxin treatments, the DEGs were clustered by STEM (Ernst and Bar-Joseph, 2006). The total 4,436 DEGs were divided into 20 profiles (**Figure 8D**). A total of 3,308 (74.57%) of these DEGs were clustered into 8 profiles (p-value < 0.05), which included a flat-downward trend for profile 9 (gene number 781), a flat-upward trend (profile 10, 646), a flat-upward-flat trend (profile 12, 495), a flat-downward-flat trend (profile 7, 222), an upward–downward trend (profile 13, 450), a downward–upward trend (profile 6, 227), an upward trend (profile 19, 270), and a downward trend (profile 0, 217).

Since 5 mg L^{-1} NAA significantly promoted root elongation growth, we intersected the differentially upregulated genes (103) in the NAA_5vs0 comparison group with the significantly enriched profiles of trend analysis and found that the genes of interest were concentrated in profiles 9 and 13. Among them, auxin transporters *ABCB21*, *LAX2,* and *LAX3* were significantly upregulated, and cellulose synthase (*CES*) and cellulose synthase-like (*CSL*) genes were significantly enriched (**Figure 8E**), indicating that relatively low level of auxin promoted root elongation by enhancing the expression of cell wall synthesis related genes. Since high level of auxin (20 mg L^{-1} NAA) significantly promoted root cell proliferation, we intersected the differentially upregulated genes (1,573) in the NAA_20vs5 comparison group with the significantly enriched profiles of trend analysis and found that the genes of interest were concentrated in profiles 10, 12, 19, and 6. Among them, auxin-related genes including conjugating (*GH3.1*, *GH3.8*), transportation (*PIN1/2*, *LAX2*), and signal transduction pathway (*ARF5/6/17*, *IAA4/8/10*) were significantly upregulated, and stem cell control and regeneration pathway-related genes (e.g., *WOX11/12/5*) and cell division regulatory genes (*CYC A1;4/B1;1/D2;1/D3;2*) were significantly upregulated (**Figure 8E**), indicating that high level of auxin promotes cell division and excessive proliferation by inducing regeneration pathway genes.

Functional Analysis of DcWOX12 Gene

Since WOX members are highly induced by high level of auxin (**Figure 8E**) and WOX11/12 play an important role in root regeneration (Hu and Xu, 2016), we used 35S promoter to drive the expression of a *D. catenatum WOX11/12* homologous gene *DcWOX12* (**Supplementary Figure 1**) in Arabidopsis for heterologous function analysis (**Figure 9A**). In *35S::DcWOX12* transgenic Arabidopsis, seed germination was delayed 1 day, compared with wild type (**Figure 9B**). The root growth was also inhibited, and the root length of 2 weeks was ∼ 2 cm, which was much shorter than that of wild type (∼ 7.5 cm); the number of lateral roots in 2 weeks was about 1, which was much lower than that of wild type (∼16) (**Figure 9B**). More than 70% of the transgenic plants developed slowly, became chlorosis, and failed to form normal true leaves, but there were trichomes on the cotyledons (**Figures 9C–F**), which reminiscence the characteristics of true leaves. Shoot apexes in some seedlings continued to overproliferate and

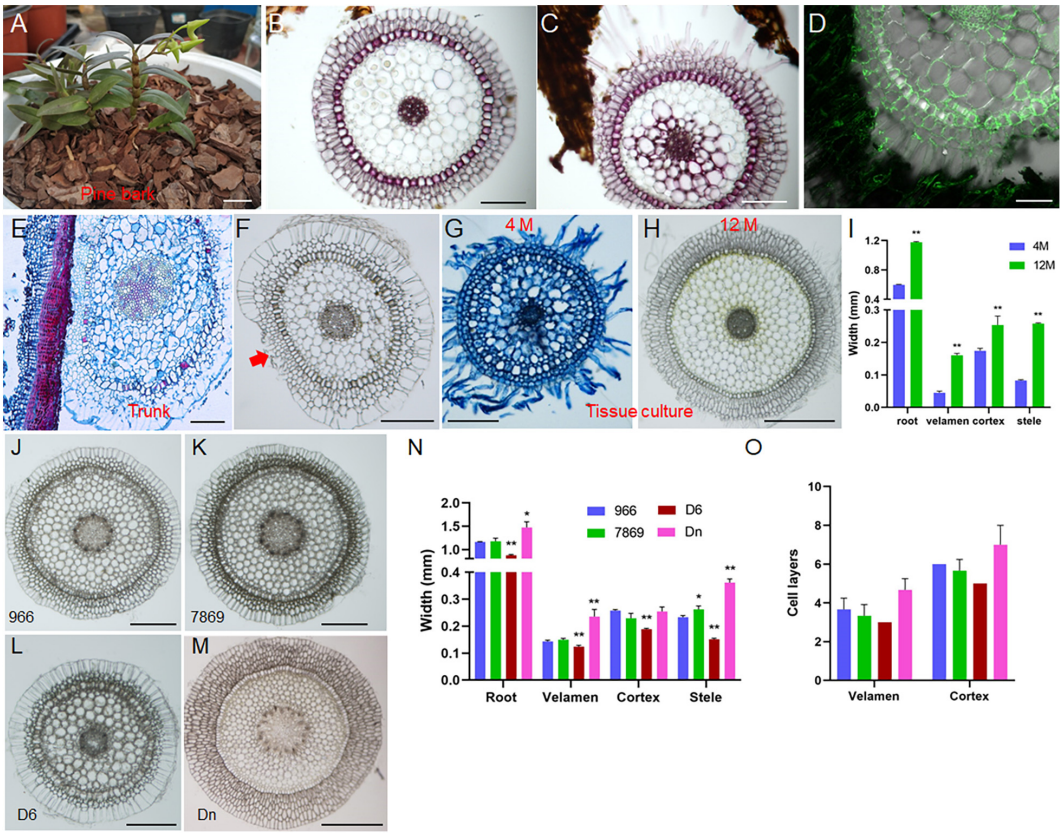


FIGURE 6 | Effects of growth environment and germplasm on root development of *D. catenatum*. **(A–D)** cultivation of *D. catenatum* on bark substrate, **(B)** cross-section of roots away from bark, **(C)** roots close to bark, **(D)** LM5 marker; **(E,F)** cross-cutting of *D. catenatum* root attached to the trunk; red arrow indicated the attachment site. **(G–I)** root cross-cutting of tissue cultured *D. catenatum*, **(G)** 100 D, **(H)** 300 D; $^{**}p < 0.01$ (two-sample *t*-test, compared with 4M). **(J–O)** Cytological observation of aerial root **(J–M)** and corresponding statistical data **(N,O)** of *D. catenatum* varieties JP966, JP7869, D6, and *D. nobile* (Dn). $^{*}P < 0.05$ and $^{**}p < 0.01$ (two-sample *t*-test, compared with JP966). Bars = 1 cm in **(A)**; 200 μm in **(B,C,E–G,J–L)**; 100 μm in **(D)**, 500 μm in **(H,M)**.

form a structure-like embryogenic callus, and occasionally differentiated into true leaves (**Figures 9G–K**); although the cotyledon growth of some seedlings arrested, the roots continued to grow and dedifferentiate to form highly proliferative callus (**Figure 9L**). Therefore, *DcWOX12* overexpression promotes cell dedifferentiation and cell proliferation and inhibits cell and organ differentiation. Transcriptome sequencing showed that stem cell control, regeneration pathway, and embryo-specific gene expression were significantly increased (**Figure 9M**). Transmission electron microscopy showed that the chlorosis tissues of *DcWOX12* overexpressing plants failed to develop the typical thylakoid lamella in the chloroplast (**Figure 9N**). Correspondingly, GO enrichment found that the DEGs related to chloroplast development and photosynthesis pathway were significantly enriched (**Figure 9O**), which further indicated that DcWOX12 inhibits organ differentiation.

DISCUSSION

Rooting structure emergence is a key morphological innovation associated with the plant evolutionary adaption from water to land. The earliest rooting structure is rhizoid, which is either unicellular filament in liverwort and hornwort or multicellular in moss (Tam et al., 2015). Fossil evidence indicates that rhizoids associated with bilaterally symmetric axes still serve as the rooting structures in the common ancestor of vascular plants, which should be initially rootless (Hetherington and Dolan, 2018). Subsequently, during the evolution of vascular plants, true roots with vascular tissues arose to provide mechanical support for macrophyte growth and much better ability for absorption and fixation than shallow growing rhizoids. Roots have similar structure and consist of the epidermis, cortex, vascular cylinder, and root cap in distinct species (Fujinami et al., 2020), but phylogenetic analyses together with Devonian fossil evidence have suggested that roots belong to paraphyletic origin and have evolved independently at least two times, characterized by two distinct root-branching events, including apex dichotomous-branching roots in extant lycophytes and endogenously pericycle/endodermis-derived LRs/ARs in euphyllophytes (ferns and seed plants) (Kenrick and Strullu-Derrien, 2014). The emergence and diversification of LRs/ARs greatly enhanced the environmental adaptability and successful land colonization of plants. *D. catenatum* as a typical

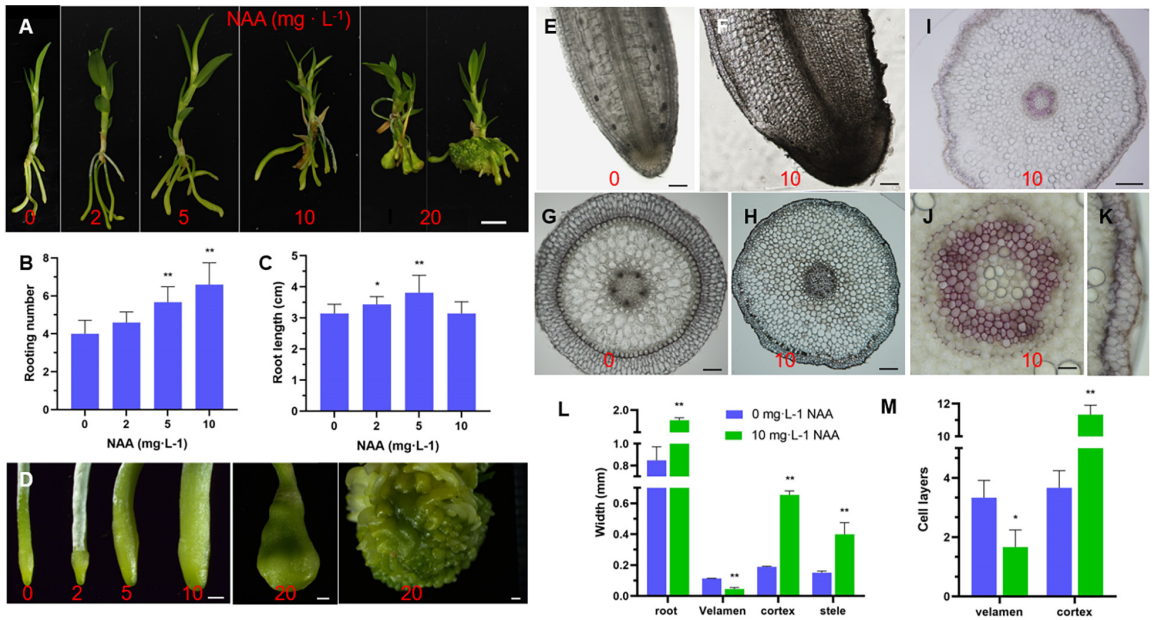


FIGURE 7 | Effect of auxin on root development of *D. catenatum*. **(A–D)** Effect of auxin gradient on phenotype of *D. catenatum*. **(A)** Statistical data of rooting number and root length **(B,C)**, close-up view of roots by gradient auxin treatment **(D)**; **(E–M)** cytological analysis of root thickening caused by high concentration of auxin (10 mg L^{-1}). **(E–H)** Optical observation of vibrating section, **(I–K)** Phloroglucinol staining, **(L,M)** statistical data of thickness and cell layers of different root tissues. $^{*}p < 0.05$ and $^{**}p < 0.01$ (two-sample *t*-test, compared with 0 mg L^{-1}). Bars = 1 cm in **(A)**, 1 mm in **(D)**, 200 μm in **(E**,**F**,**H)**,10 μm in **(I)**, 100 μm in **(G)**, 50 μm in **(J,K)**.

epiphytic orchid represents a novel plant lifestyle transformation from soil growth to tree/aerial growth, and its root system is directly exposed to severe environmental fluctuations, resulting in further specialization of root developmental program. The most significant change is the appearance of a spongy multilayered velamen, which has strong water absorption and retention capability with great possibility to protect roots from being dehydrated and dying when exposed to the air. Moreover, velamen root cells of epiphytic orchids may harbor phototrophic algal associates inside, such as the cyanobacterial-symbiosis supplying host plants with fixed nitrogen (Tsavkelova et al., 2003a,b; Deepthi and Ray, 2020). There are chloroplasts in the root interior, which can synthesize organic matter and store starch, and enhance the osmotic regulation ability of roots. Due to no need to contact the soil to expand the absorption area, lateral roots and root hairs are usually absent. The starch grains in the whole root cap are evenly distributed, and the gravitropism is not obvious, which is consistent with the epiphytic life of aerial roots. Orchid seeds lacking endosperm are very tiny and easy to spread to the trunk or cliff with the wind; they contain underdeveloped embryo arrested at the globular stage without radicle differentiation (Fang et al., 2016). Germination of orchid seeds is not a spontaneous process and requires symbiosis with specific fungi in the wild (Teixeira da Silva et al., 2015). Upon germination, they do not continue to complete the embryogenesis, but produce a unique transitional organ – protocorm, the interior of which first differentiates into the "shoot body" with SAM and young leaves, and root meristem is subsequently produced at their opposite poles. Therefore, *D. catenatum* roots do not originate from radicles, but from *de novo* root organogenesis, which may reflect the characteristics of the ancestors. The protocorm can not only protect the shoot body, store the nutrient starch, but also provide a habitat for the symbiotic microbe, establish a long-term and solid symbiotic relationship, and promote nutrient supply and stress tolerance.

D. catenatum root structure includes velamen, exodermis, cortex, endodermis, and stele. The exodermis and endodermis serve as two lines of defense to control the entry of substances, and the exodermis is more developed, which may be related to the requirement for high-efficiently filtering a large amount of water that are instantaneously captured by the velamen and preventing moisture escaping from the cortex. Thus, the cooperative evolution between velamen and exodermis confers aerial root of epiphytic orchid high tolerance to drought stress. In the cortex, water and minerals bind with storage substances to reside within the cell, bind with pectin polysaccharides to reside in the apoplast, and then further permeate through endodermis and enter the vascular system to transport.

The root development in *D. catenatum* is not only regulated by intrinsic genetic cues, such as germplasm, seedling age, and sampling segment, but also displays corresponding structural adaptability under different cultivation modes, especially for the velamen. When the root was grown in the matrix, the cells of each velamen layer were uniform in size; when exposed to the air directly, the outermost layer of velamen cells widened significantly; the cells closely attached to the bark still maintain viability, which are small and densely arranged. Whether root

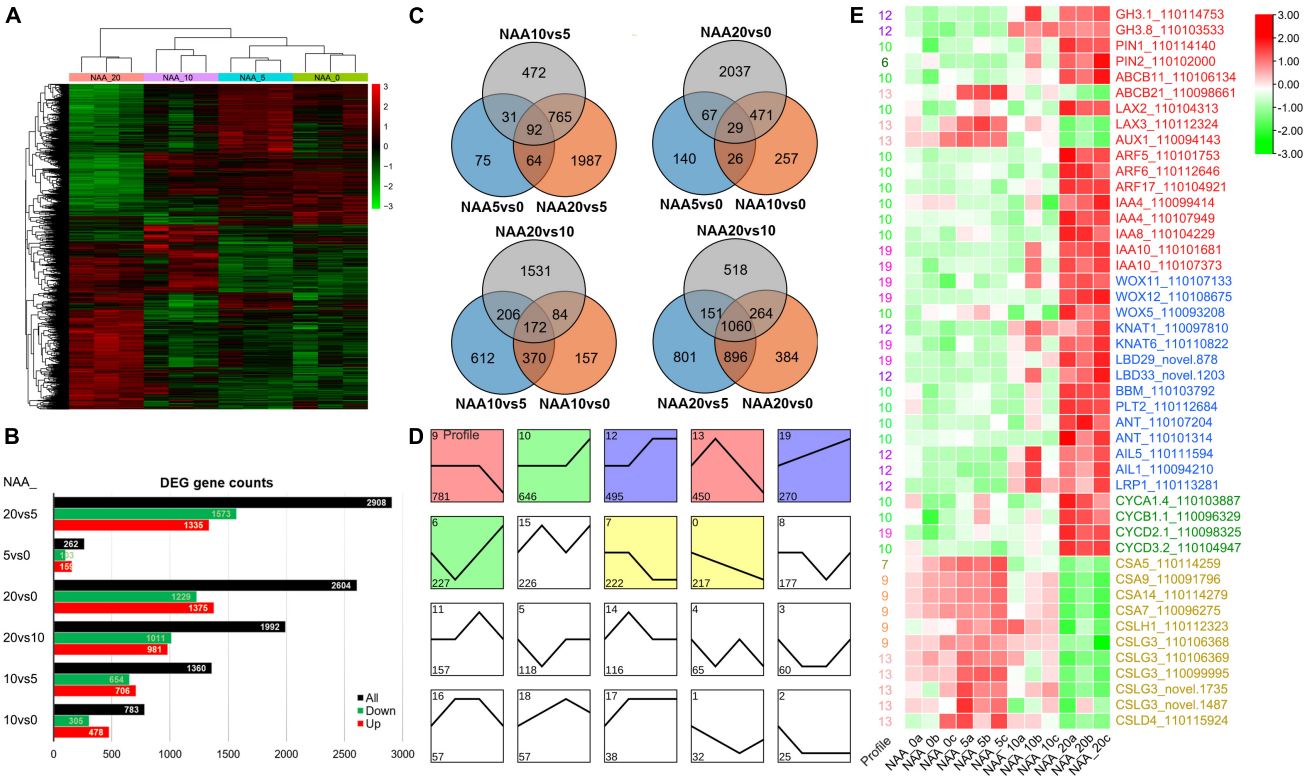


FIGURE 8 | DEGs in different comparisons. **(A)** Cluster of DEGs in different comparisons. Red indicates high expression genes, whereas green indicates low expression genes. **(B)** Statistics of all, downregulated, and upregulated DEGs in different comparisons. **(C)** Venn diagram of DEGs in different comparisons. All DEGs are clustered into three comparison groups represented by three ellipses. The overlapping parts of different ellipses represent the number of DEGs in common from those comparison groups. **(D)** A number of 20 expression profiles of DEGs. **(E)** The heatmap of important DEGs associated with auxin response, stem cell control, cell division, and cell wall synthesis. Red indicates high expression genes, whereas green indicates low expression genes. The FPKM values in each row were processed by *Z*-score normalization.

hairs are produced on the surface of the velamen depends on the availability and efficiency of nutrient acquisition. The roots exposed directly to the air or far from the pine bark substrate do not need to produce root hairs, whereas roots attached to the pine bark substrate or grown in agar medium can expand the absorption area by producing root hairs. Therefore, epiphytic orchid roots have evolved to make extensive elastic adaptations to environmental changes with economical and efficient principles. Under artificial aseptic culture conditions, the seed germination and seedling growth in most of plants can be directly cultured on hormone-free medium, but the seed germination and subsequent growth of *D. catenatum* require additional auxin application (1–2 mg·L^{-1} NAA); otherwise, the seedling becomes weak, indicating that the auxin synthesized by itself cannot meet its own needs. However, in the case of facility cultivation using pine bark substrates, *D. catenatum* grows normally, which may be related to the extra auxin supplied by its symbionts. A wide variety of orchid endophytic bacteria and fungi have been identified to promote plant growth *via* auxin secretion during seedling acclimation and root colonization (Faria et al., 2013; Pavlova et al., 2017; Shah et al., 2018, 2021, 2022; Chand et al., 2020). Therefore, preparing suitable concentration of auxin solution or isolating the auxin-producing symbiotic microorganisms can be directly applied to further improve the production and breeding of medicinal orchids.

It is known that root is usually very sensitive to auxin, and the lowest concentrations of indole-3-acetic acid (IAA) to inhibit the growth of root, shoot, and stem are around 0.01, 1 (=0.175 mg L^{-1}), and 10 mm, respectively (Thimann, 1938). In Arabidopsis, 0.1 mg L^{-1} NAA has severely inhibited root growth, and 1 mg L^{-1} has completely abolished root growth (**Supplementary Figure 2**). However, the root of *D. catenatum* has high tolerance to auxin and grows best at the concentration of 5 mg L^{-1} NAA in this study. We speculated that the high tolerance of *D. catenatum* root to auxin is related to both the barrier function of the developed exodermis and osmotic regulation of storage matter in the cortex. Phylogenetic analysis suggests that IAA biosynthesis evolved independently in bacteria, fungi, and plants; IAA might serve as a widespread signaling molecule that mediates interspecies communication during evolution (Fu et al., 2015). Therefore, the biological significance of the high auxin tolerance of *D. catenatum* may be reflected in its interaction with external microorganisms, which helps to create a microhabitat preferring probiotics and inhibiting pathogens. For instance, high level of auxin treatments (10 and 100 mg L^{-1} IAA) considerably promotes the growth of several

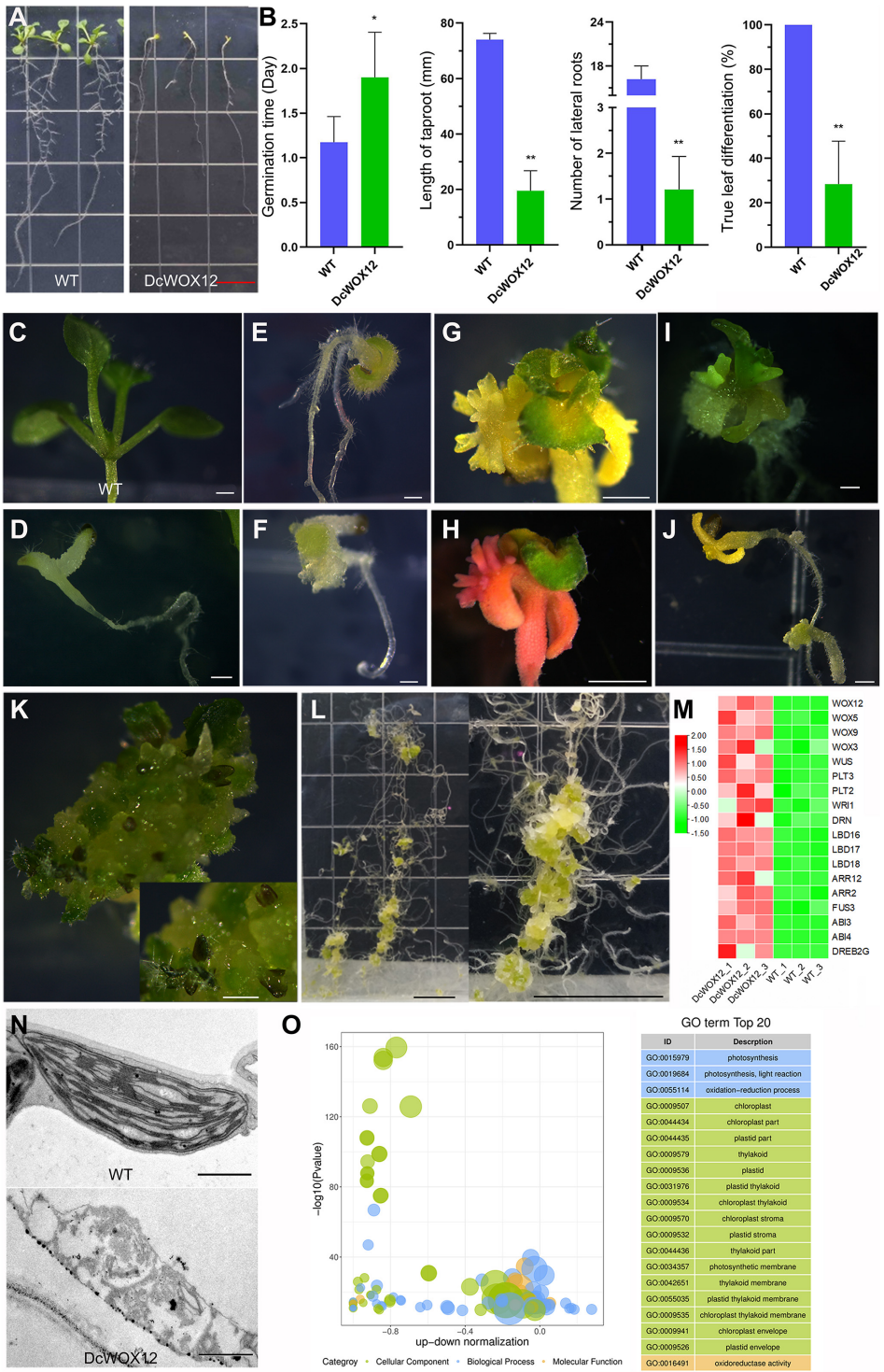


FIGURE 9 | Callus induction by overexpression of *DcWOX12* in Arabidopsis. **(A–L)** The representative phenotype in *DcWOX12* overexpression Arabidopsis; **(A–F)** 2-week-old seedling phenotype, **(B)** showing statistical data of germination time, root length, lateral root number, and true leaf differentiation rate. $^{*}p < 0.05$ and $^{**}p < 0.01$ (two-sample *t*-test, compared with 0 mg·L^{-1}). **(G–J)** 4W, **(K)** 26W, **(L)** 16W. **(M)** The heatmap of important DEGs associated with callus formation and embryo-like structure. Red indicates high expression genes, whereas green indicates low expression genes. **(N)** Electron micrographs of leaf mesophyll cell. **(O)** Bubble plot of top 20 of GO enrichment of DEGs. The ordinate is –log10 (*p*-value), and the abscissa is the up-down normalization value (the ratio of the difference between the number of upregulated genes and the number of downregulated genes in the total differential genes); the yellow line represents the threshold of *p*-value = 0.05; on the right is the term list with the top 20 *p*-values. Different colors represent different ontology classes. Bars = 1 cm in **(A,L)**; 500 μm in **(C–F)** and **(I–K)**, 1 mm in **(G,H)**, and 1 μm in **(N)**.

IAA-producing bacteria associated with *Dendrobium moschatum* (Tsavkelova et al., 2007). Exogenous application of 200 μm IAA inhibits the growth of some plant-associated pathogens, such as *Agrobacterium*, *Erwinia*, *Pseudomonas*, and *Xanthomonas* genera (Liu and Nester, 2006).

Exogenous auxin treatment on *D. catenatum* seedlings exhibited two effects: (i) Relatively low level of auxin promoted rooting and root elongation. Transcriptome analysis showed that the expression of root initiation-related genes, such as WOX family members, was not significantly changed, which may be related to the absence of lateral root in *D. catenatum*. However, cell wall synthesis-related genes (CESs and CSLs) were significantly upregulated, indicating that they play a role in root elongation. (ii) High level of auxin promoted root thickening and overproliferation. Anatomical observation showed that the differentiation of velamen and exodermis was inhibited, and the cortical parenchyma cells proliferated drastically, indicating that high concentrations of auxin can inhibit cell differentiation and promote cell division, which is consistent with the application of high level of auxin to induce explant dedifferentiation and callus formation in tissue culture. Correspondingly, the genes related to stem cell control and regeneration pathways were significantly upregulated, especially WOX11-WOX5 root initiation programs, in line with the established rooting pathway during callus induction (Sugimoto et al., 2010; Hu and Xu, 2016); cell division regulatory genes (*CYCA1;4*, *B1;1*, *D2;1*, *D3;2*) were also significantly induced, consistent with the rapid proliferation of cells.

In Arabidopsis, WOX11/12 are involved in accelerating the formation of AR or/and callus from explants in the presence of auxin (Liu et al., 2014). In rice, OsWOX11 participates in shoot and root development. Both os*wox11* mutant and *OsWOX11* overexpression rice display reduced shoot growth. Moreover, blocking of *OsWOX11* also inhibits crown root development, and overexpression of *OsWOX11* accelerates crown root cell division and dramatically shoot-borne crown root production (Zhao et al., 2009, 2015; Cheng et al., 2016, 2018). Here, overexpression of *DcWOX12* in Arabidopsis inhibits chloroplast and leaf differentiation and reduces root length and number of lateral roots, which is different from Arabidopsis and rice *WOX11/12* overexpression phenotype. Intriguingly, *DcWOX12* overexpression can directly form highly proliferative callus in the absence of auxin, displaying stronger cell dedifferentiation and regeneration ability than *WOX11/12* homologs in Arabidopsis and rice, and somewhat reminiscent of overexpression of SAM master regulator *WUS* (Zuo et al., 2002; Gallois et al., 2004), which provide an exciting new gene resource in breakthrough of high-efficient monocot genetic transformation method in a broad range of genotypes and species (Hofmann, 2016; Lowe et al., 2016). Transcriptome sequencing showed that stem cell control, regeneration pathway, and embryo-specific gene expression were significantly increased in *DcWOX12* overexpression plant. Thus, *WOX11/12* homologous genes in different species not only retained a conservative role, but also underwent functional divergence during evolution. In addition, overexpression of *OsWOX11* in rice improves drought resistance by modulating rice root system development (Cheng et al., 2016). Specific expression of *OsWOX11* in rice roots greatly improves root growth and activity and results in increased K uptake and grain yield in low K soil (Chen et al., 2015). Overexpression of *PagWOX11/12a* enhances salt and drought tolerance in poplar (Wang et al., 2020, 2021). Therefore, different WOX11/12 versions from different species could be exploited in biotechnology to improve plant resistance to abiotic stresses *via* manipulating plant root development and to improve the efficiency of monocot plant transformation *via* promoting cell dedifferentiation and cell fate transition.

CONCLUSION

This study suggests unique developmental characteristics of *D. catenatum* aerial root closely associated with its high environmental adaptability to special epiphytic life mode. Furthermore, *D. catenatum* root displays high tolerance and dual responses to auxin, possibly also related to its adaptability. Heterozygous overexpression system indicated that *D. catenatum* WOX12 confers cell high-efficient pluripotency acquisition properties. Therefore, this study provided not only an insight into ideal root structure breeding of simulated natural cultivation in *D. catenatum*, but also a novel target gene for improving the efficiency of monocot plant transformation, especially in orchid plants.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are publicly available. This data can be found here: NCBI, PRJNA815882.

AUTHOR CONTRIBUTIONS

DC and CL planned and designed the research. JT and WJ performed the experiments. DC, JT, CL, ZH, and WJ analyzed the data. DC and JS wrote the manuscript. All authors approved the manuscript.

FUNDING

This work was funded by the National Natural Science Foundation of China (31870310) and Major Science and Technology Projects of Breeding New Varieties of Agriculture in Zhejiang Province (no. 2021C02074).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.935540/full#supplementary-material

Supplementary Figure 1 | Phylogenetic analysis of WOX proteins from *D. catenatum* and Arabidopsis. The phylogenetic tree was constructed according to the NJ method by MEGA 7.0 with 1000 bootstrap replicates.

Supplementary Figure 2 | Effect of auxin on Arabidopsis root development. Arabidopsis was grown in MS medium containing 0, 0.1, 0.5, and 1 mg L^{-1} NAA, respectively. Bars = 1 cm.

REFERENCES

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

Andersen, M. C., Boos, I., Marcus, S. E., Kracun, S. K., Rydahl, M. G., Willats, W. G., et al. (2016). Characterization of the LM5 pectic galactan epitope with synthetic analogues of beta-1,4-d-galactotetraose. *Carbohydr. Res.* 436, 36–40. doi: 10.1016/j.carres.2016.10.012

Chand, K., Shah, S., Sharma, J., Paudel, M. R., and Pant, B. (2020). Isolation, characterization, and plant growth-promoting activities of endophytic fungi from a wild orchid Vanda cristata. *Plant Signal. Behav.* 15:1744294. doi: 10.1080/15592324.2020.1744294

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

Chen, D., Molitor, A. M., Xu, L., and Shen, W. H. (2016). Arabidopsis PRC1 core component AtRING1 regulates stem cell-determining carpel development mainly through repression of class I KNOX genes. *BMC Biol.* 14:112. doi: 10.1186/s12915-016-0336-4

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

Chen, D., Wang, Q., Feng, J., Ruan, Y., and Shen, W. H. (2019). Arabidopsis ZUOTIN RELATED FACTOR1 proteins are required for proper embryonic and post-embryonic root development. *Front. Plant Sci.* 10:1498. doi: 10.3389/fpls.2019.01498

Chen, G., Feng, H., Hu, Q., Qu, H., Chen, A., Yu, L., et al. (2015). Improving rice tolerance to potassium deficiency by enhancing OsHAK16p:WOX11-controlled root development. *Plant Biotechnol. J.* 13, 833–848. doi: 10.1111/pbi.12320

Cheng, S., Tan, F., Lu, Y., Liu, X., Li, T., Yuan, W., et al. (2018). WOX11 recruits a histone H3K27me3 demethylase to promote gene expression during shoot development in rice. *Nucleic Acids Res.* 46, 2356–2369. doi: 10.1093/nar/gky017

Cheng, S., Zhou, D. X., and Zhao, Y. (2016). WUSCHEL-related homeobox gene WOX11 increases rice drought resistance by controlling root hair formation and root system development. *Plant Signal. Behav.* 11:e1130198. doi: 10.1080/15592324.2015.1130198

Chomicki, G., Bidel, L. P. R., Ming, F., Coiro, M., Zhang, X., Wang, Y., et al. (2015). The velamen protects photosynthetic orchid roots against UV-B damage, and a large dated phylogeny implies multiple gains and losses of this function during the Cenozoic. *New Phytol.* 205, 1330–1341. doi: 10.1111/nph.13106

Clausen, M. H., Willats, W. G., and Knox, J. P. (2003). Synthetic methyl hexagalacturonate hapten inhibitors of anti-homogalacturonan monoclonal antibodies LM7, JIM5 and JIM7. *Carbohydr. Res.* 338, 1797–1800. doi: 10.1016/s0008-6215(03)00272-6

da Silva, J. A., Cardoso, J. C., Dobranszki, J., and Zeng, S. (2015). Dendrobium micropropagation: a review. *Plant Cell Rep.* 34, 671–704. doi: 10.1007/s00299-015-1754-4

Deepthi, A. S., and Ray, J. G. (2018). Endophytic diversity of hanging velamen roots in the epiphytic orchid Acampe praemorsa. *Plant Ecol. Divers.* 11:13. doi: 10.1080/17550874.2019.1610911

Deepthi, A. S., and Ray, J. G. (2019). Applications of endophytic-fungal-isolates from velamen root of wild orchids in floriculture. *Braz. J. Biol. Sci.* 6:13.

Deepthi, A. S., and Ray, J. G. (2020). Algal associates and the evidence of cyanobacterial nitrogen fixation in the velamen roots of epiphytic orchids. *Glob. Ecol. Conserv.* 22:e00946. doi: 10.1016/j.gecco.2020.e00946

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7:191. doi: 10.1186/1471-2105-7-191

Fang, S. C., Chen, J. C., and Wei, M. J. (2016). Protocorms and protocorm-like bodies are molecularly distinct from zygotic embryonic tissues in Phalaenopsis aphrodite. *Plant Physiol* 171, 2682–2700. doi: 10.1104/pp.16.00841

Faria, D. C., Dias, A. C., Melo, I. S., and de Carvalho Costa, F. E. (2013). Endophytic bacteria isolated from orchid and their potential to promote plant growth. *World J. Microbiol. Biotechnol.* 29, 217–221. doi: 10.1007/s11274-012-1173-4

Fu, S. F., Wei, J. Y., Chen, H. W., Liu, Y. Y., Lu, H. Y., and Chou, J. Y. (2015). Indole-3-acetic acid: a widespread physiological code in interactions of fungi with other organisms. *Plant Signal. Behav.* 10:e1048052. doi: 10.1080/15592324.2015.1048052

Fujinami, R., Yamada, T., and Imaichi, R. (2020). Root apical meristem diversity and the origin of roots: insights from extant lycophytes. *J. Plant Res.* 133, 291–296. doi: 10.1007/s10265-020-01167-2

Gallois, J. L., Nora, F. R., Mizukami, Y., and Sablowski, R. (2004). WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem. *Genes Dev.* 18, 375–380. doi: 10.1101/gad.291204

Ge, Y., Fang, X., Liu, W., Sheng, L., and Xu, L. (2019). Adventitious lateral rooting: the plasticity of root system architecture. *Physiol. Plant* 165, 39–43. doi: 10.1111/ppl.12741

Hauber, F., Konrad, W., and Roth-Nebelsick, A. (2020). Aerial roots of orchids: the velamen radicum as a porous material for efficient imbibition of water. *Appl. Phys. A* 126:885. doi: 10.1007/s00339-020-04047-7

Henry, J. S., Ligrone, R., Vaughn, K. C., Lopez, R. A., and Renzaglia, K. S. (2021). Cell wall polymers in the Phaeoceros placenta reflect developmental and functional differences across generations. *Bryophyt. Divers. Evol.* 43, 265–283. doi: 10.11646/bde.43.1.19

Henry, S., Divol, F., Bettembourg, M., Bureau, C., Guiderdoni, E., Perin, C., et al. (2015). Immunoprofiling of rice root cortex reveals two cortical subdomains. *Front. Plant Sci.* 6:1139. doi: 10.3389/fpls.2015.01139

Hetherington, A. J., and Dolan, L. (2018). Bilaterally symmetric axes with rhizoids composed the rooting structure of the common ancestor of vascular plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20170042. doi: 10.1098/rstb.2017.0042

Hofmann, N. R. (2016). A breakthrough in monocot transformation methods. *Plant Cell* 28:1989. doi: 10.1105/tpc.16.00696

Hu, X., and Xu, L. (2016). Transcription factors WOX11/12 directly activate WOX5/7 to promote root primordia initiation and organogenesis. *Plant Physiol.* 172, 2363–2373. doi: 10.1104/pp.16.01067

Joca, T. A. C., de Oliveira, D. C., Zotz, G., Cardoso, J. C. F., and Moreira, A. (2020). Chemical composition of cell walls in velamentous roots of epiphytic Orchidaceae. *Protoplasma* 257, 103–118. doi: 10.1007/s00709-019-01421-y

Jones, D., and Tisserat, B. (1990). Clonal propagation of orchids. *Methods Mol. Biol.* 6, 181–191. doi: 10.1385/0-89603-161-6:181

Kenrick, P., and Strullu-Derrien, C. (2014). The origin and early evolution of roots. *Plant Physiol.* 166, 570–580. doi: 10.1104/pp.114.244517

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

Lakehal, A., and Bellini, C. (2019). Control of adventitious root formation: insights into synergistic and antagonistic hormonal interactions. *Physiol. Plant* 165, 90–100. doi: 10.1111/ppl.12823

Lardon, R., and Geelen, D. (2020). Natural variation in plant pluripotency and regeneration. *Plants* 9:1261. doi: 10.3390/plants9101261

Liu, J., Sheng, L., Xu, Y., Li, J., Yang, Z., Huang, H., et al. (2014). WOX11 and 12 are involved in the first-step cell fate transition during de novo root organogenesis in Arabidopsis. *Plant Cell* 26, 1081–1093. doi: 10.1105/tpc.114.122887

Liu, P., and Nester, E. W. (2006). Indoleacetic acid, a product of transferred DNA, inhibits vir gene expression and growth of *Agrobacterium tumefaciens* C58. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4658–4662. doi: 10.1073/pnas.0600366103

Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M. J., et al. (2016). Morphogenic regulators baby boom and wuschel improve monocot transformation. *Plant Cell* 28, 1998–2015. doi: 10.1105/tpc.16.00124

Marcus, S. E., Blake, A. W., Benians, T. A., Lee, K. J., Poyser, C., Donaldson, L., et al. (2010). Restricted access of proteins to mannan polysaccharides in intact plant cell walls. *Plant J.* 64, 191–203. doi: 10.1111/j.1365-313X.2010.04319.x

Pavlova, A. S., Leontieva, M. R., Smirnova, T. A., Kolomeitseva, G. L., Netrusov, A. I., and Tsavkelova, E. A. (2017). Colonization strategy of the endophytic plant growth-promoting strains of *Pseudomonas* fluorescens and Klebsiella oxytoca on the seeds, seedlings and roots of the epiphytic orchid, Dendrobium nobile Lindl. *J. Appl. Microbiol.* 123, 217–232. doi: 10.1111/jam.13481

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

Shah, S., Chand, K., Rekadwad, B., Shouche, Y. S., Sharma, J., and Pant, B. (2021). A prospectus of plant growth promoting endophytic bacterium from orchid (Vanda cristata). *BMC Biotechnol.* 21:16. doi: 10.1186/s12896-021-00676-9

Shah, S., Shah, B., Sharma, R., Rekadwad, B., Shouche, Y. S., Sharma, J., et al. (2022). Colonization with non-mycorrhizal culturable endophytic fungi enhances orchid growth and indole acetic acid production. *BMC Microbiol.* 22:101. doi: 10.1186/s12866-022-02507-z

Shah, S., Shrestha, R., Maharjan, S., Selosse, M. A., and Pant, B. (2018). Isolation and characterization of plant growth-promoting endophytic fungi from the roots of dendrobium moniliforme. *Plants* 8:5. doi: 10.3390/plants8010005

Shekhar, V., Stckle, D., Thellmann, M., and Vermeer, J. E. M. (2019). The role of plant root systems in evolutionary adaptation. *Curr. Top. Dev. Biol.* 131, 55–80. doi: 10.1016/bs.ctdb.2018.11.011

Sheng, L., Hu, X., Du, Y., Zhang, G., Huang, H., Scheres, B., et al. (2017). Non-canonical WOX11-mediated root branching contributes to plasticity in Arabidopsis root system architecture. *Development* 144, 3126–3133. doi: 10.1242/dev.152132

Si, J. P., Wang, Q., Liu, Z. J., Liu, J. J., and Luo, Y. B. (2017a). [Breakthrough in key science and technologies in Dendrobium catenatum industry]. *Zhongguo Zhong Yao Za Zhi* 42, 2223–2227. doi: 10.19540/j.cnki.cjcmm.2017.0102

Si, J. P., Zhang, Y., Luo, Y. B., Liu, J. J., and Liu, Z. J. (2017b). [Herbal textual research on relationship between Chinese medicine"Shihu" (Dendrobium spp.) and "Tiepi Shihu" (D. catenatum)]. *Zhongguo Zhong Yao Za Zhi* 42, 2001–2005. doi: 10.19540/j.cnki.cjcmm.20170415.002

Sma-Air, S., and Ritchie, R. J. (2020). Photosynthesis in a Vanda sp orchid with photosynthetic roots. *J. Plant Physiol.* 251, 153187. doi: 10.1016/j.jplph.2020.153187

Street, I. H., Mathews, D. E., Yamburkenko, M. V., Sorooshzadeh, A., John, R. T., Swarup, R., et al. (2016). Cytokinin acts through the auxin influx carrier AUX1 to regulate cell elongation in the root. *Development* 143, 3982–3993. doi: 10.1242/dev.132035

Sugimoto, K., Jiao, Y., and Meyerowitz, E. M. (2010). Arabidopsis regeneration from multiple tissues occurs via a root development pathway. *Dev. Cell* 18, 463–471. doi: 10.1016/j.devcel.2010.02.004

Tam, T. H., Catarino, B., and Dolan, L. (2015). Conserved regulatory mechanism controls the development of cells with rooting functions in land plants. *Proc. Natl. Acad. Sci. U.S.A* .112, E3959–E3968. doi: 10.1073/pnas.1416324112

Teixeira da Silva, J. A., Tsavkelova, E. A., Zeng, S., Ng, T. B., Parthibhan, S., Dobranszki, J., et al. (2015). Symbiotic in vitro seed propagation of Dendrobium: fungal and bacterial partners and their influence on plant growth and development. *Planta* 242, 1–22. doi: 10.1007/s00425-015-2301-9

Thimann, K. V. (1938). Hormones and the analysis of growth. *Plant Physiol.* 13, 437–449. doi: 10.1104/pp.13.3.437

Tsavkelova, E. A., Cherdyntseva, T. A., Klimova, S. Y., Shestakov, A. I., Botina, S. G., and Netrusov, A. I. (2007). Orchid-associated bacteria produce indole-3-acetic acid, promote seed germination, and increase their microbial yield in response to exogenous auxin. *Arch. Microbiol.* 188, 655–664. doi: 10.1007/s00203-007-0286-x

Tsavkelova, E. A., Lobakova, E. S., Kolomeitseva, G. L., Cherdyntseva, T. A., and Netrusov, A. I. (2003a). [Associative cyanobacteria isolated from the roots of epiphytic orchids]. *Mikrobiologiia* 72, 105–110.

Tsavkelova, E. A., Lobakova, E. S., Kolomeitseva, G. L., Cherdyntseva, T. A., and Netrusov, A. I. (2003b). [Localization of associative cyanobacteria on the roots of epiphytic orchids]. *Mikrobiologiia* 72, 99–104.

Verhertbruggen, Y., Marcus, S. E., Haeger, A., Ordaz-Ortiz, J. J., and Knox, J. P. (2009). An extended set of monoclonal antibodies to pectic homogalacturonan. *Carbohydr. Res.* 344, 1858–1862. doi: 10.1016/j.carres.2008.11.010

Wang, L. Q., Li, Z., Wen, S. S., Wang, J. N., Zhao, S. T., and Lu, M. Z. (2020). WUSCHEL-related homeobox gene PagWOX11/12a responds to drought stress by enhancing root elongation and biomass growth in poplar. *J. Exp. Bot.* 71, 1503–1513. doi: 10.1093/jxb/erz490

Wang, L. Q., Wen, S. S., Wang, R., Wang, C., Gao, B., and Lu, M. Z. (2021). PagWOX11/12a activates PagCYP736A12 gene that facilitates salt tolerance in poplar. *Plant Biotechnol. J.* 19, 2249–2260. doi: 10.1111/pbi.13653

Willats, W. G., Limberg, G., Buchholt, H. C., van Alebeek, G. J., Benen, J., Christensen, T. M., et al. (2000). Analysis of pectic epitopes recognised by hybridoma and phage display monoclonal antibodies using defined oligosaccharides, polysaccharides, and enzymatic degradation. *Carbohydr. Res.* 327, 309–320. doi: 10.1016/s0008-6215(00)00039-2

Xie, X., Wang, Y., Datla, R., and Ren, M. (2021). Auxin and target of rapamycin spatiotemporally regulate root organogenesis. *Int. J. Mol. Sci.* 22:11357. doi: 10.3390/ijms222111357

Xu, L. (2018). De novo root regeneration from leaf explants: wounding, auxin, and cell fate transition. *Curr. Opin. Plant Biol.* 41, 39–45. doi: 10.1016/j.pbi.2017.08.004

Yu, J., Liu, W., Liu, J., Qin, P., and Xu, L. (2017). Auxin control of root organogenesis from callus in tissue culture. *Front. Plant Sci.* 8:1385. doi: 10.3389/fpls.2017.01385

Yu, Q. B., Jiang, Y., Chong, K., and Yang, Z. N. (2009). AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript accD RNA editing and early chloroplast biogenesis in Arabidopsis thaliana. *Plant J.* 59, 1011–1023. doi: 10.1111/j.1365-313X.2009.03930.x

Zhao, Y., Cheng, S., Song, Y., Huang, Y., Zhou, S., Liu, X., et al. (2015). The interaction between rice ERF3 and WOX11 promotes crown root development by regulating gene expression involved in cytokinin signaling. *Plant Cell* 27, 2469–2483. doi: 10.1105/tpc.15.00227

Zhao, Y., Hu, Y., Dai, M., Huang, L., and Zhou, D. X. (2009). The WUSCHEL-related homeobox gene WOX11 is required to activate shoot-borne crown root development in rice. *Plant Cell* 21, 736–748. doi: 10.1105/tpc.108.061655

Zotz, G., and Winkler, U. (2013). Aerial roots of epiphytic orchids: the velamen radicum and its role in water and nutrient uptake. *Oecologia* 171, 733–741. doi: 10.1007/s00442-012-2575-6

Zuo, J., Niu, Q. W., Frugis, G., and Chua, N. H. (2002). The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J.* 30, 349–359. doi: 10.1046/j.1365-313x.2002.01289.x

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fpls.2022.923000

The Transcriptome Profiling of Flavonoids and Bibenzyls Reveals Medicinal Importance of Rare Orchid *Arundina graminifolia*

Sagheer Ahmad[1], *Jie Gao*[1], *Yonglu Wei*[1], *Chuqiao Lu*[1], *Genfa Zhu*[1,2]* *and Fengxi Yang*[1,2]*

[1] *Guangdong Key Laboratory of Ornamental Plant Germplasm Innovation and Utilization, Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China,* [2] *Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China*

Orchids are very important flowering plants that spend long juvenile phases before flowering. Along with aesthetic importance, they are rich sources of medicinal components. However, their long reproductive cycle is the major hurdle to study the medicinal efficacy. *Arundina graminifolia* is a rare orchid that grows fast, unlike other orchids, and this characteristic makes it an ideal plant to study the medicinal enrichment of orchids. Therefore, this study presents the identification of important medicinal components in various parts of *A. graminifolia*. Transcriptome analysis was performed for five stages (FD1–FD5) of flower development and four tissue types (mature flower, silique, root, and leaf) to ascertain genetic regulators of flavonoids and bibenzyls. Most of the genes showed the highest expression in roots as compared with other tissues. Weighted gene coexpression network analysis (WGCNA) was performed to identify the coexpression modules and the candidate genes involving biosynthesis pathways of these chemicals. MEyellow module contained the highly coexpressed genes. Moreover, the concentrations of phenylpropanoid, bibenzyls, and flavone were ascertained through high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). Phenylpropanoid and bibenzyl were comparatively high in the leaf, while flavone showed a high concentration in the stem. The selected candidate genes [bibenzyl biosynthesis (BIBSY212), CYP84A1, CYP73A4, 4CLL7, UGT88B1, UGT73C3, anthocyanin synthase (ANS), phenylalanine ammonia-lyase (PAL), flavanone synthase FLS, and CHS8] were validated through quantitative real-time PCR (qRT-PCR). Most of these genes showed high expression in leaf and root as compared with other tissue. Therefore, the presence of bibenzyls and flavonoids in different parts of *A. graminifolia* and their molecular regulators can provide a quick source to decipher the medicinal efficacy of orchids.

Keywords: ***Arundina graminifolia*, medicinal constituents, model orchid, HPLC-MS/MS, WGCNA**

OPEN ACCESS

Edited by:
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

Reviewed by:
Paromik Bhattacharyya,
Institute of Himalayan Bioresource Technology (CSIR), India
Phanikanth Jogam,
Kakatiya University, India
Ali Raza,
Fujian Agriculture and Forestry University, China

***Correspondence:**
Fengxi Yang
yangfengxi@gdaas.cn
Genfa Zhu
zhugenfa@gdaas.cn

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: *18 April 2022*
Accepted: *19 May 2022*
Published: *23 June 2022*

Citation:
Ahmad S, Gao J, Wei Y, Lu C, Zhu G and Yang F (2022) The Transcriptome Profiling of Flavonoids and Bibenzyls Reveals Medicinal Importance of Rare Orchid Arundina graminifolia. Front. Plant Sci. 13:923000. doi: 10.3389/fpls.2022.923000

INTRODUCTION

The Orchidaceae family is one of the largest angiosperm families and mainly contains ornamental flowers (Cai et al., 2015b; Wong et al., 2017). About 100,000 million species are grown worldwide, showing the immense horticultural importance of orchids (Ahmad et al., 2021b). Orchids, such as *Cymbidium* and *Phalaenopsis*, bloom after a long vegetative phase of 2–3 years

(Ahmad et al., 2021a). However, a rare orchid *Arundina graminifolia,* completes its vegetative phase in 6 months and then continues flowering (Ahmad et al., 2021a,b, 2022). It is commonly called bamboo orchid due to its shape similar to bamboo plant. Its significantly short vegetative phase makes it an ideal plant to study different aspects and benefits of orchids; especially the medicinal components. Although a few reports describe the identification of different chemicals in bamboo orchids, the molecular regulation remains unattended. Revealing the genetic regulation of flavonoids and bibenzyls in *A. graminifolia* can serve as a useful source to extend the knowledge to chemical identification in other orchids.

Arundina graminifolia is used as a medicinal plant in China because of the presence of flavonoids, stilbenoid, and phenols in its extracts, which exhibit antioxidant, anti-virus, anti-tumor, and other medicinal properties (Ai et al., 2019). The whole plant is a famous Dai medicine in China, curing food poisoning, blood stasis, and liver toxicity (Liu et al., 2007; Zhang et al., 2012; Xiaohua et al., 2015). In India, it is used as an emollient and antibacterial agent (Panda and Mandal, 2013). In Bangladesh, it used to cure rheumatism (Hossain, 2009). Previous studies have reported stilbenoids as the major secondary metabolites in bamboo orchid, revealing the structural diversity of phenanthrenes (Liu et al., 2005), diphenylethylenes (Hu et al., 2013b; Li et al., 2013a; Gao et al., 2014; Meng et al., 2014; Yang et al., 2014), bibenzyls (Majumder and Ghosal, 1993; Du et al., 2014), and other phenolic compounds (Gao et al., 2012; Hu et al., 2013a; Lidan et al., 2013; Li et al., 2013b; Niu et al., 2013). Bibenzyls are effective antitumor agents due to their antioxidant and cell-protective properties (Gong, 2003; Zhang et al., 2007; Barbosa et al., 2009; Li et al., 2010; Su, 2011; Cai et al., 2015a). Moreover, Bibenzyls are used in several drugs and skincare products (Zhang et al., 2007; Hossain, 2011).

In *Dendrobium officinale*, the active medicinal constituents include alkaloids, terpenes, polysaccharides, flavonoids, phenols, and bibenzyl (Zhang et al., 2016; Tang et al., 2017). Bibenzyls have also been identified in *Dendrobium sinense* (Chen et al., 2014). In the orchid *Epidendrum rigidum*, bibenzyls showed phytotoxic activity, suggesting that orchid bibenzyls can be a good lead for developing novel herbicides (Hernández-Romero et al., 2005). The biosynthesis of bibenzyls is regulated by four key enzymes. The biosynthesis of dihydro-m-coumaroyl-CoA begins with phenylalanine and ends with a cinnamate molecule catalyzed by phenylalanine ammonia-lyase (PAL). The cinnamate is incorporated into m-coumaric-CoA with the catalyzation of cinnamate 4-hydroxylase (C4H). Then, dihydro-p-coumaroyl-CoA is produced from p-coumaric-CoA with the catalyzation of 4-coumarate: CoA ligase (4CL). Dihydro-m-coumaric acid is synthesized at the same time from m-coumaric acid with the incorporation of cytochrome P450 (CYP450) (Majumder et al., 2001; Peled-Zehavi et al., 2015; Jeong et al., 2016; Ibdah et al., 2017; Yahyaa et al., 2017).

Cytochrome P450 genes are the important regulators of secondary metabolites and bibenzyls (Adejobi et al., 2021). They play a role in regulating the production of defense-related secondary metabolites (Gomez et al., 2011). The expression of CYP450s significantly affects the quality of bibenzyls in plants. The pathway for secondary metabolite biosynthesis involves different regulatory modifications and physiological factors in response to environmental changes. Usually, the secondary metabolites, such as bibenzyl, accumulate in low concentrations in the tissues (Hussain et al., 2012). However, the market demands high levels of bibenzyl accumulation in plant tissues to facilitate the drug-making. Therefore, it is an immediate need to dissect the molecular mechanisms underlying the biosynthesis of bibenzyls in *A. graminifolia*, which can provide continuous supply of medicinal ingredients much faster than other orchids. The identification of regulatory factors and rate-limiting enzymes, responsible for bibenzyl biosynthesis, is also an area of further exploration in orchids. There is an urgent need of plant-derived sustainable sources of bibenzyl, and *A. graminifolia* has excellent potential to provide a continuous supply of bibenzyl and other medicinal ingredients. Researchers have so far focused on the identification of the chemical components, pharmacological activities, and bioactive substances in *A. graminifolia*. However, the underlying molecular regulation is not discussed.

With the rapid development of RNA-seq technology, transcriptomic data mining offers a great opportunity to discover pivotal genetic regulators or rate-limiting enzymes that control the secondary metabolites in plants (Jia et al., 2015; Zhu et al., 2016). The transcriptome analyses have identified several rate-limiting genes regulating the biosynthesis of lignin in *Apium graveolens* (Pandey et al., 2018), terpenoids in *Eugenia uniflora* (Kumar et al., 2016), and flavonoids in *Solanum viarum*, *Dracaena cambodiana*, and *Phyllanthus embica* (Meng et al., 2016; Lei et al., 2018; Yuan et al., 2018). However, the biosynthesis pathway and potential gene regulators of bibenzyl biosynthesis in *Arundina* plants remain elusive.

Therefore, this study investigates the accumulation of flavonoid, phenylpropanoid, and bibenzyl in various tissues of *A. graminifolia.* We performed the transcriptome analysis from five developmental stages and four tissue types to identify the putative genes associated with the biosynthesis of flavonoid, phenylpropanoid, and bibenzyl. The findings provide new resources for the rapid production of medicinal ingredients in orchids.

METHODS

Plant Materials and Growth Conditions

The bamboo orchid plants were obtained from seeds grown at the greenhouse of the Institute of Environmental Horticulture (Guangdong Academy of Agricultural Sciences, China). The seeds were grown on asymbiotic MS (Murashige and Skoog) media. The seedlings were transferred to a greenhouse and kept at a temperature of 25/20°C day and night, with a 16/8 h of photoperiod. Samples were taken from five stages of flower development (FD), stages 1–5, fully mature flowers, fruits, leaves, and root, as previously mentioned (Ahmad et al., 2021a). Three technical and biological repeats were collected in liquid nitrogen and immediately stored at −180°C for RNA extraction.

RNA-Seq Library Preparation and Sequencing

RNA was extracted using the TaKaRa kit and cDNA libraries were made for RNA. From total RNA, mRNA was filtered using the Oligotex mRNA Midi Kit (QIAGEN, GERMANY), followed by quality and quantity check using a Nano-Drop 2000 spectrophotometer (Thermo Scientific, USA). The cDNA libraries were made using the Illumina manufacturing protocol (Ahmad et al., 2021a). The total mRNA was subjected to first and second strand cDNA synthesis and adapter ligation, and low cycle enrichment was achieved by the TruSeq®RNA HT Sample Prep Kit from Illumina (USA). The products were evaluated with the Agilent 2200 TapeStation and Qubit®2.0 of Life Technologies (USA). Dilutions were made to 10 pM for generating *in situ* clusters on HiSeq2500 pair-end flow cell. The 2 × 100 bp sequencing was performed, resulting in 60 M reads per sample. Transcriptomic *de novo* was done with the Trinity program using default parameters (Grabherr et al., 2011).

Differentially Expressed Gene Analysis

The FPKM (fragments per kilobase of exon model per million reads mapped) expression of genes was calculated using following formula:

$$\text{FPKM} = [\text{total exon reads/mapped reads (millions)}] \times \text{exon length (kb)}$$

The edgeR package was used to identify significant differences among genes (Ahmad et al., 2022). A threshold level of gene significant difference was calculated at a false discover rate (FDR) of <0.05 and a $|\log2 \text{ ratio}| > 1$ (2-fold change). The genes were mapped to public annotation databases, such as Kyoto encyclopedia of genes and genomes (KEGG), gene ontology (GO), non-redundant (NR), and KEGG orthology (KO), using the BLASTX program, as previously documented (Ahmad et al., 2021a). The DEGs were also annotated on the UniProt and protein family (PFAM) databases using default parameters. A hypergeometric test was applied using phyper function in R to identify DEGs with enriched terms. Significantly enriched KEGG or GO terms were filtered at p-value or q-value of ≤ 0.05.

Filtering of Phenylpropanoid and Flavonoid Pathway DEGs

The KEGG pathways were searched to find genes related to flavonoids biosynthesis, phenylpropanoid biosynthesis, and flavone biosynthesis. About 500 DEGs were filtered related to these pathways. Selected pathway genes of flavonoid and phenylpropanoid were used to make a heatmap using the pheatmap package of R. The protein–protein interaction analysis was performed at the online String facility using the protein sequences of DEGs as inputs.

Weighted Gene Coexpression Network Analysis

The weighted gene coexpression network analysis WGCNA package of R was used to perform coexpression and to identify module of coexpressed genes (Ahmad et al., 2021a). Initially, the removal of unqualified genes was performed through the function of goodSamplesGenes. After this, the criterion of scale-free topology was used to select a suitable soft-threshold power using the function of pickSoftThreshold. With the incorporation of one-gene-to-all relationship, the adjacency matrix was converted into topological matrix (TOM) (Yip and Horvath, 2007). TOM-based dissimilarity (1-TOM) function was used to identify genes showing hierarchical clustering. Finally, the modules with highly interconnected gene clusters were created (Ravasz et al., 2002).

The Quantitative Real-Time PCR Analysis

RNA was extracted from flowers, capsules, leaves, and roots to check the expression of 10 selected candidate genes in the flavonoid and bibenzyl pathway. Total RNA was extracted and cDNA was obtained using the Fermentas protocol. The quantitative real-time PCR (qRT-PCR) was performed in a mixture of 20 μl containing 10 μl SYBR premix Ex-taq™ (Takara, Japan) with the Bio-Rad CFX-96 RealTime PCR system (Bio-Rad, USA). *Actin* was the internal standard used to normalize expression data. The primers are shown in **Supplementary Table 3**.

Quantification of Flavonoids and Bibenzyls

The concentrations of flavonoids, bibenzyls, and phenylpropanoids were ascertained for leaf, root, stem, flower, and fruit. The chemical contents were determined following the protocol of high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) (Aglient), as previously documented (Pan et al., 2010).

Statistical Analysis

Chemical and qRT-PCR data were analyzed using one-way ANOVA on SPSS software (SPSS Inc., Chicago, IL, USA; ver. 16.0). A significant difference is shown at $p < 0.05$ or $p < 0.01$ level.

RESULTS

Transcriptome Analysis and Annotation

The transcriptome analysis produced 71.2 billion high quality reads. Each sample produced about 7.8 billion reads consisting of 10.8–12.8 Gb data (**Supplementary Table 1**). The data were filtered into 25,353 unigenes and 94,317 transcripts (**Supplementary Table 2**). The data were annotated to obtain KEGG, GO, Pfam, eggNOG, NR, and SwissProt enrichments. These data have been used to mine genes related to flowering time regulation in our previous researches (Ahmad et al., 2021a,b). However, the abundant data about chemical constituents were not analyzed. Therefore, the DEGs were filtered to obtain genes specific to flavonoids, bibenzyls, and phenylpropanoids pathways.

Major Enzyme Classes and Pathways

All the DEGs related to different chemicals were further analyzed to filter different classes of enzymes that may play roles in the regulation of phytochemicals, such as flavonoids and bibenzyls.

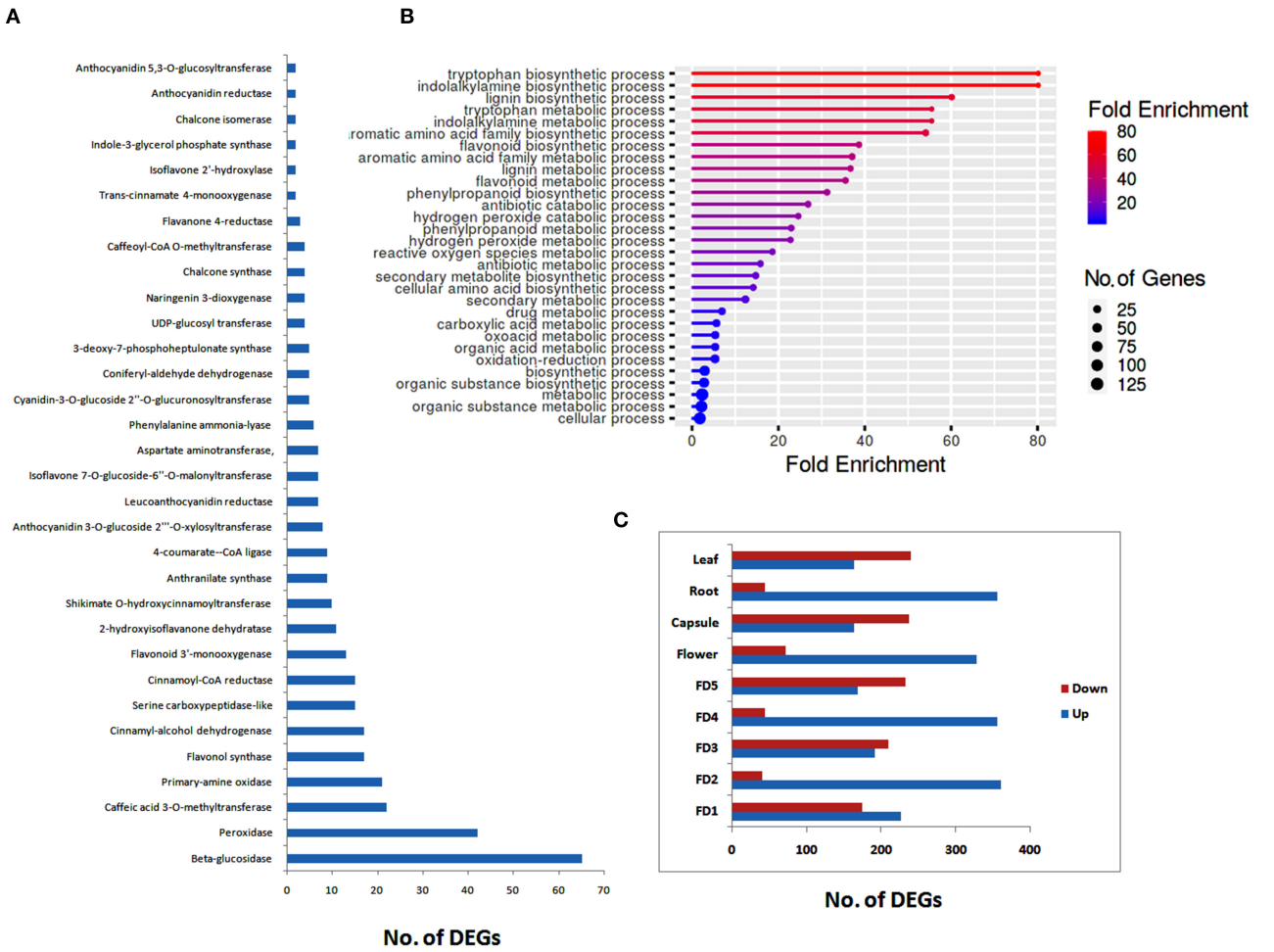


FIGURE 1 | Distribution of secondary metabolite data into enzyme classes **(A)**, major biological process enrichments **(B)** and up- and downregulated differentially expressed genes (DEGs) for different tissues **(C)**.

The highest numbers of genes were related to beta-glucosidases, followed by peroxidases, caffeic acid 3-O-methyltransferases, and primary-amine oxidase (**Figure 1A**). We identified 17 DEGs related to flavonol synthase, an important enzyme regulating flavonoid biosynthesis. Other major enzymes identified in the biosynthesis pathways of flavonoids and bibenzyls included cinnamoyl-CoA reductase, flavonoid 3'-monooxygenase, 4-coumarate–CoA ligase, UDP-glucosyl transferase, and chalcone synthase (**Figure 1A**).

Enrichment of different biological processes of DEGs was found using the ShinyGO online enrichment tool. A considerable numbers of DEGs were enriched in flavonoid metabolic process, flavonoid biosynthetic process, phenylpropanoid biosynthetic process, secondary metabolite biosynthetic process, and antibiotic metabolic process (**Figure 1B**). The highest numbers of genes were related to cellular process and metabolic process.

Stage-specific up- and down-regulated DEGs were also ascertained. The highest numbers of up regulated DEGs were found in root as compared to other stages of flower development or tissue types (**Figure 1C**). However, in floral development stages, there was no considerable difference between up-regulated and down-regulated DEGs, except for FD2, were a large number of DEGs were up-regulated (**Figure 1C**).

Major Regulators of Flavonoids and Phenylpropanoids

A number of genes were involved in the biosynthetic process of flavonoids and phenylpropanoids (**Figure 2**). A number of enzymes, such as anthocyanin reductase (ANR), anthocyanin synthase (ANS), bibenzyl synthases (BIBSY), chalcone isomerase (CHI), phytochrome P450 (CYPs), and flavanone synthase (FLS) were identified in the flavonoid and bibenzyl biosynthesis pathways (**Figure 2A**). In the phenylpropanoid biosynthesis, the major genes identified included 4CLs, BGLUs, CADs, CYPs, PALs, PERs, and SCPLs (**Figure 2B**).

Most of the flavonoid and bibenzyl biosynthesis pathway genes showed the highest expression in root as compared with other tissues (**Figure 2A**). The phenylpropanoid pathway genes were highly expressed in the early stages of flower development and root (**Figure 2B**).

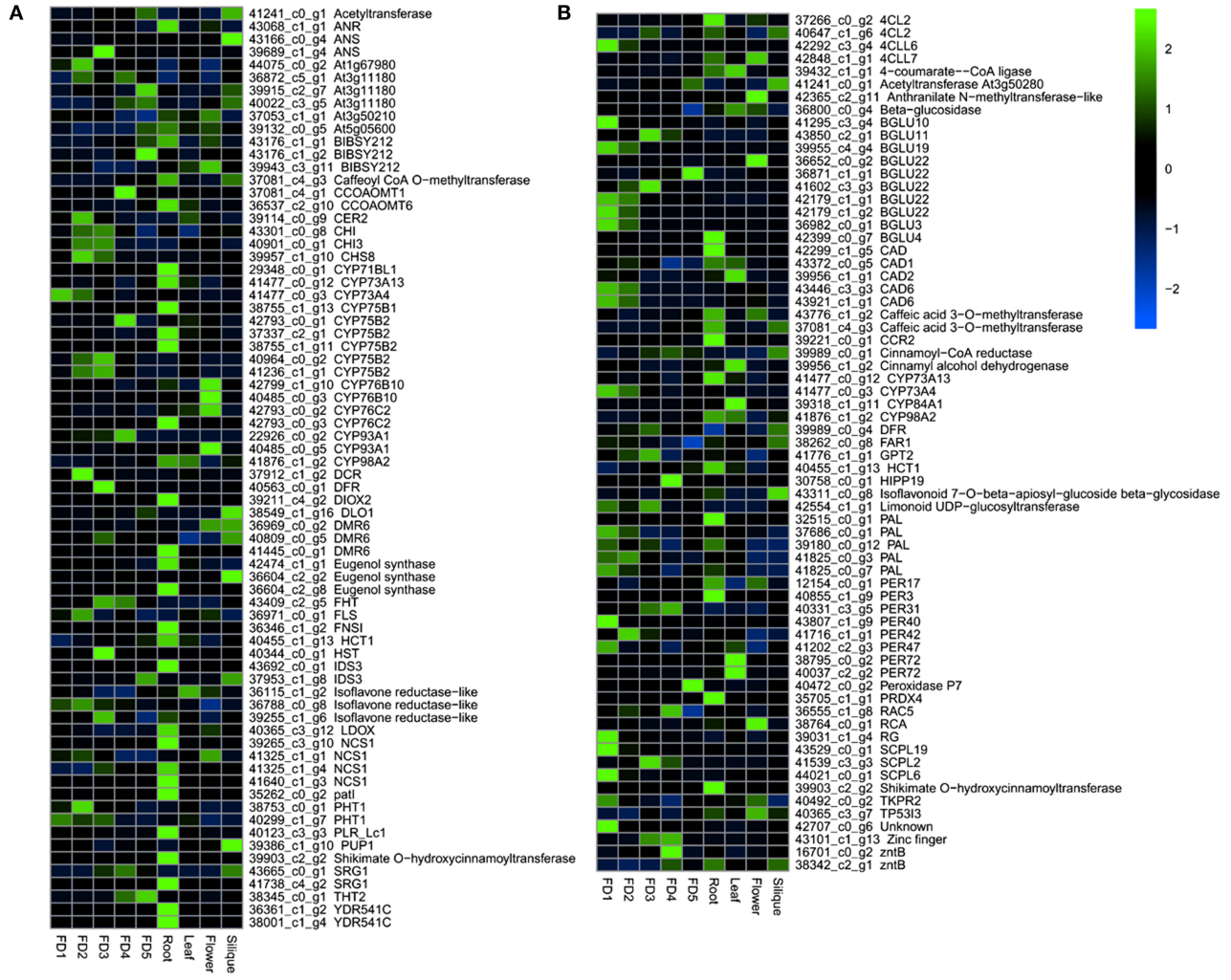


FIGURE 2 | The heatmap of major DEGs involved in the biosynthesis of flavonoids **(A)** and phenylpropanoids **(B)**.

Root Specificity and Protein–Protein Interaction of Genes

Plenty of DEGs related to flavonoids, anthocyanins, bibenzyls, terpenoids, and other chemicals, were found in our data. A heatmap clustering of these DEGs showed that most of them were expressed in roots (**Figure 3A**), which were mostly downregulated in other tissues. In addition, the highest numbers of genes were specifically expressed in root as compared with other tissues. Phenylpropanoid biosynthesis and flavonoid biosynthesis were the abundant pathways along with the biosynthesis of amino acids (**Figure 3B**). Moreover, the same pathways can be found in the protein–protein interactions among DEGs involving chemical homeostasis (**Figure 3C**).

WGCNA Modules in Flower and Non-Reproductive Tissues

A WGCNA was performed to further understand the gene coexpression for phytochemical regulation in *A. graminifolia* (**Figure 4**). The cluster dendrogram shows the possibility of different modules in the specification of flavonoids and phenylpropanoids (**Figure 4A**). The modules based on this clustering show two expression sets. The one set contains clusters of upregulated genes, MEturquoise, and MEyellow; while MEblue, MEbrown, and MEgreen contain the clusters of downregulated genes (**Figure 4B**).

MEyellow showed highly coexpressed genes with greater tissue specificity than flowers (**Figure 4C**). The MEblue contained the most downregulated set of genes, with more downregulation in flower as compared with tissues. The MEyellow module was further analyzed to know its significance for tissue specificity (**Figure 4D**). A significant correlation can be seen in the genes for tissues.

Candidates of Flavonoids and Bibenzyls

We identified DEGs potentially involved in the biosynthesis of phenylpropanoid, flavonoid, and bibenzyls (**Figure 5A**).

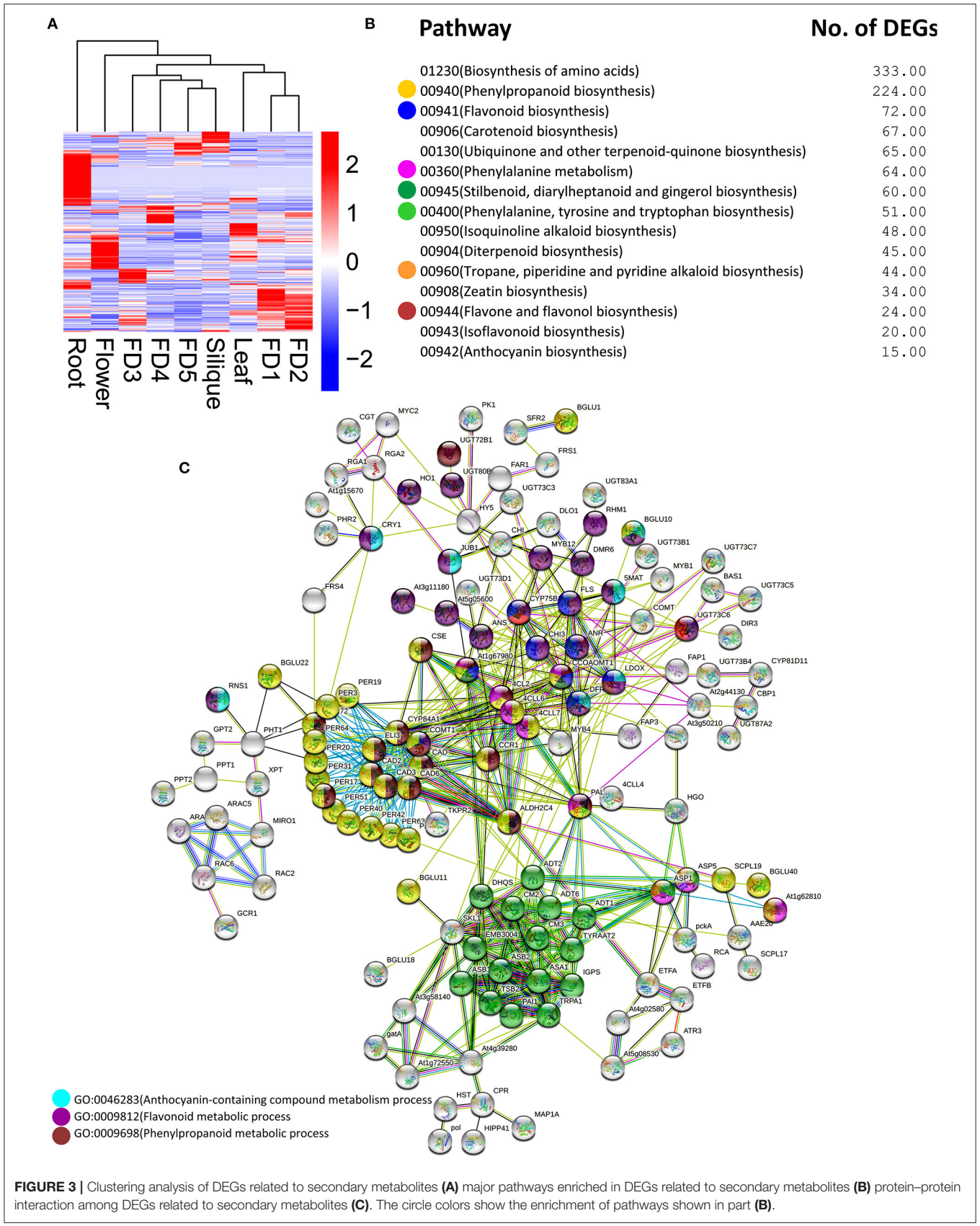


FIGURE 3 | Clustering analysis of DEGs related to secondary metabolites **(A)** major pathways enriched in DEGs related to secondary metabolites **(B)** protein–protein interaction among DEGs related to secondary metabolites **(C)**. The circle colors show the enrichment of pathways shown in part **(B)**.

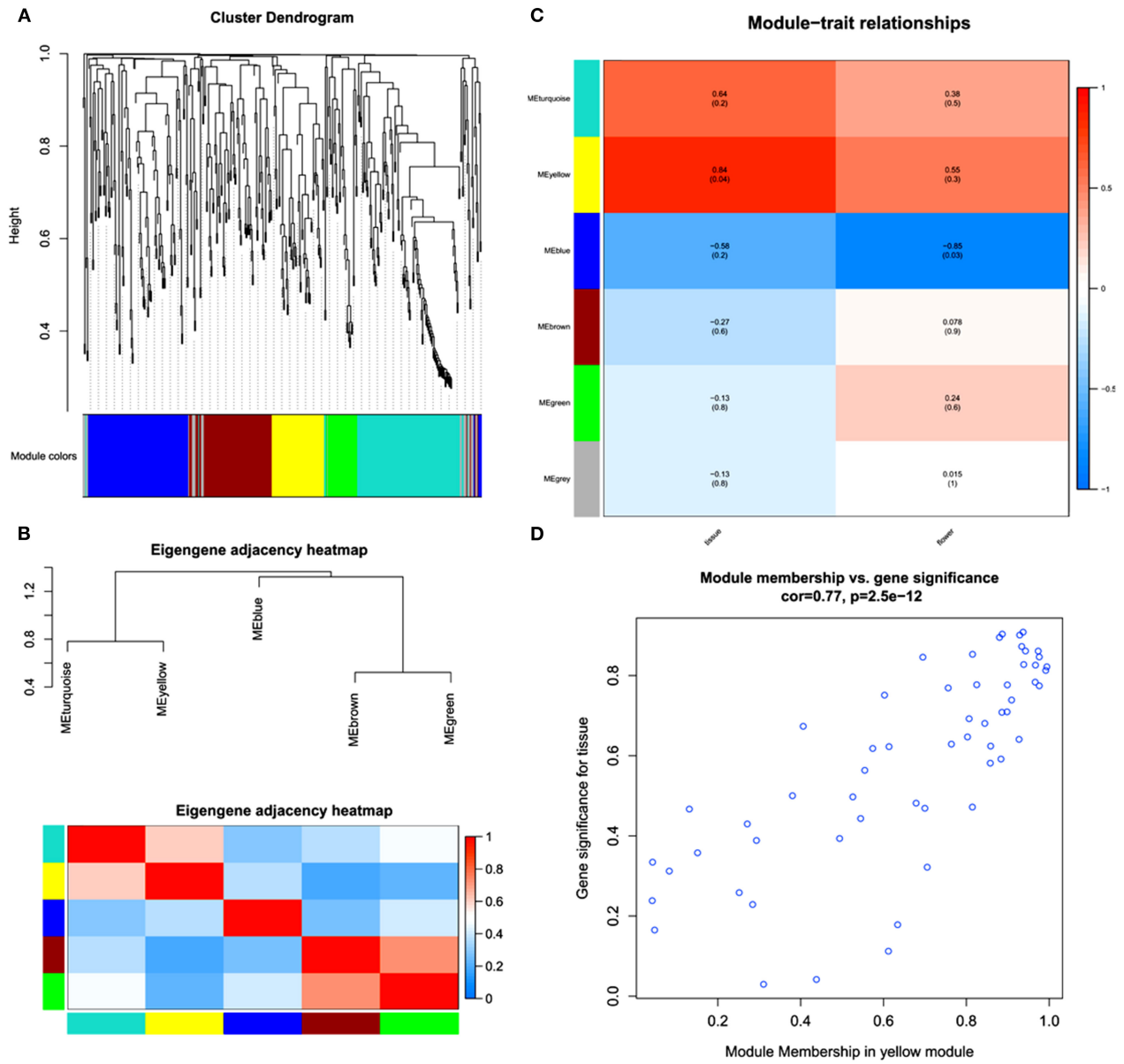


FIGURE 4 | Weighted gene coexpression network analysis (WGCNA); cluster dendrogram **(A)**, eigengene adjacency heatmap of modules **(B)**, key modules with expression intensities **(C)**, and module membership of yellow module for tissue-specific gene significance **(D)**.

Bibenzyl, belonging to sesquiterpenes (Adejobi et al., 2021), is a downstream product of methylerythritol 4-phosphate (MEP) and MVA (mevalone) biosynthesis pathways in plants. Our data also included enzymes for this pathway, such as hydroxymethylglutaryl-CoA synthase (HGMS), mevalonate kinase (MK), 1-deoxy-d-xylulose-5-phosphate reductoisomerase (DXR), 1-deoxy-d-xylulose-5-phosphate synthase (DXS), phosphomevalanote kinase (PMK), 2-C-methyl-d-erythritol 2,4-cyclodiophosphate synthase (MDS), and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (MDS). L-Phenylalanine is the usual substrate to generate bibenzyl *via* cinnamic acid with the catalysis of PAL (**Figure 5B**). The C4H catalyzation generates two isomers of m-coumaric acid and p-coumaric acid. Then, using m-coumaric acid as substrate, dihydro-m-Coumaric acid, dihydro-m-Coumaroyl-CoAic, and 3,3'5-Trihydrobibenzyl are synthesized with the catalysis of CYP450, 4CL, and BBS, respectively. We identified two C4H, two PALs, two important CYP450s (CYP84A1 and CYP98A2), three BBS, and two 4CLs (Majumder et al., 2001; Peled-Zehavi et al., 2015; Jeong et al., 2016; Ibdah et al., 2017; Yahyaa et al., 2017).

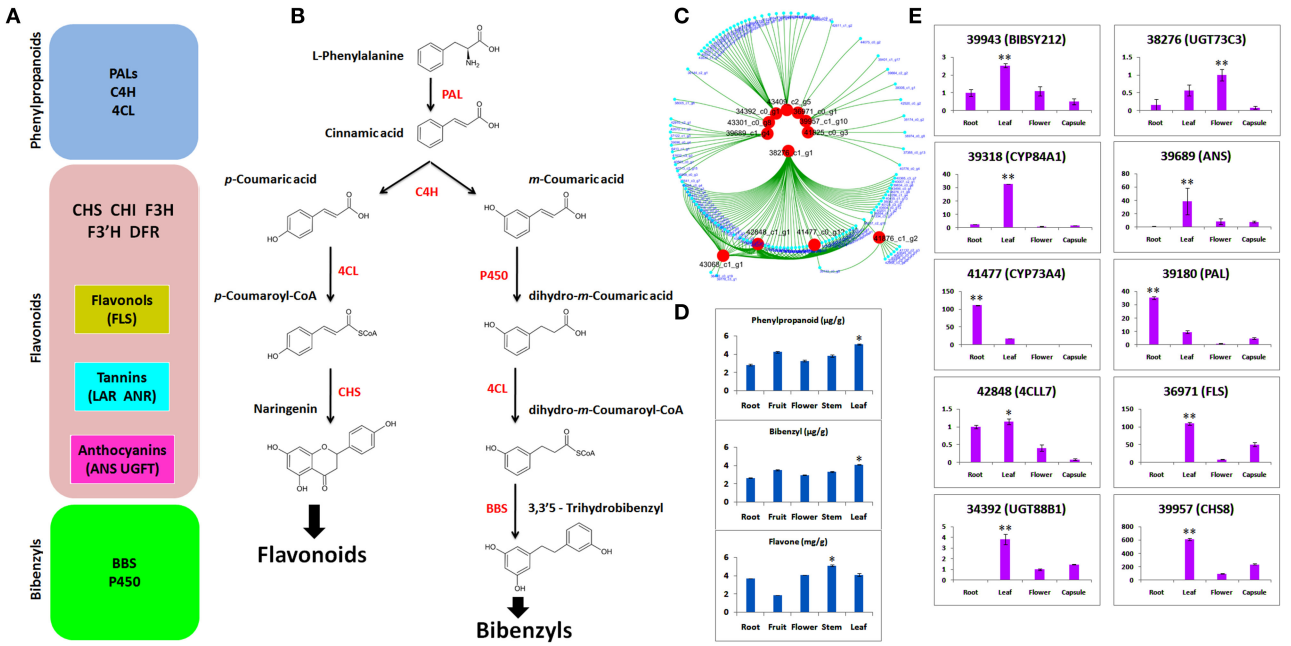


FIGURE 5 | (A) Gene regulators of phenylpropanoids, flavonoids, and bibenzyls; **(B)** biosynthesis route of flavonoids and bibenzyls. Colored in red are the enzymes that regulate different steps of biosynthetic pathways; **(C)** selection of candidates that involve biosynthesis of flavonoids and bibenzyls through coexpressed modules; **(D)** concentrations of phenylpropanoids, bibenzyls, and flavonoids through high-performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS); **(E)** quantitative real-time PCR (qRT-PCR) expression of 10 selected DEGs related to bibenzyl and flavonoid biosynthesis. The statistical significance is shown at $p \leq 0.05$ (*) or $p \leq 0.01$ (**).

Based on WGCNA analysis, the clustering of co-expressed modules suggested 12 important genes that are the important regulators of flavonoids and bibenzyls (**Figure 5C**). These hub genes are the important regulators of bibenzyl and flavonoid biosynthesis pathways shown in **Figure 5B**.

Concentration of Major Phytochemicals

Concentrations of phenlypropanoids, bibenzyls, and flavonoids were ascertained through HPLC-MS/MS in five tissues, such as flower, fruit, leaf, stem, and root (**Figure 5D**). Phenylpropanoid and bibenzyl were abundant in leaf. Flavone is a flavonoid and it was abundant in stem and low in fruit. The higher concentration of bibenzyl was observed in leaf as compared with other tissues.

The qRT-PCR Validation of Selected Genes

To empirically certify the expressions of DEGs obtained from RNA-seq, we selected 10 important genes responsible for bibenzyl biosynthesis (BIBSY212, 4CCL7, CYP84A1, and CYP73A4) and flavonoid biosynthesis (UGT73C3, PAL, FLS, UGT88B1, and CHS8). The qRT-PCR validation of candidate genes for bibenzyls and flavonoids showed that most of the genes were highly expressed in leaf, as compared with other tissues, few expressed in root and just one expressed in the flower (**Figure 5E**).

DISCUSSION

Orchidaceae contains the world's most beautiful flowers with unique shapes, colors, and forms. Potted and cut flower orchids make a huge business (Tokuhara and Mii, 2003; Bhattacharyya et al., 2014). In addition to their economic value, the presence of phytochemicals has been exploited recently for the preparation of vital drugs (Moin et al., 2012; Bhattacharyya et al., 2014). Bioactive compounds, such as polysaccharides, flavonoids, alkaloids, and bibenzyls are integral components of complex processes of drug development (Li et al., 2011; Ng et al., 2012). Polysaccharides possess hepato-protective and immunomodulatory functions, while bibenzyls exhibit anticancer, immunomodulatory, and antioxidant characteristics (Gong et al., 2004; Tang et al., 2017; Li et al., 2018). Several studies have identified genes regulating polysaccharide biosynthesis (He et al., 2015; Zhang et al., 2016; Shen et al., 2017). However, little is documented on the molecular underpinning of bibenzyl biosynthesis in plants, especially in orchis. Bamboo orchid is a highly prized unique orchid and has been playing an important role in traditional Chinese medicine. It is a rare orchid in the Orchidaceae with very short reproductive cycle as compared to others with life span of more than 3 years. Therefore, it can be used as a model to study various phenomena in orchids in a short time, especially the medicinal importance. This is the first report on the genetic regulation of bibenzyls and flavonoids in *A. graminifolia*. In addition to transcriptome dissection, the

concentrations of bibenzyl, flavonoid, and phenylpropanoid were ascertained, showing the enormous medicinal importance of bamboo orchid.

Orchids are a rich source of important phytochemicals. Various active compounds, such as dendrobine, gigantol, nobiline, and moscatilin, have been found in the leaves and stems of *Dendrobium nobile* (Suzuki et al., 1973; Miyazawa et al., 1997; Zhao et al., 2001). The leaf extracts of *Coelogyne stricta* contained alkaloids, terpenoids, and phenols (Minh et al., 2016). The *Dendrobium pandurantum* is a rich source of phenols, alkaloids, tannins, flavonoids, and triterpenoids (Johnson and Janakiraman, 2013). Different parts of *Cymbidium aloifolium* (seeds, roots, leaves, and capsules) contained alkaloids, phenols, tannins, and cardiac glycosides (Shubha and Chowdappa, 2016). Other orchid species, such as *Monodora tenuifolia* (Ezenwali et al., 2010), *Eria pseudoclavicaulis* (Moin et al., 2012), *Rhynchostylis retusa* (Bhattacharjee and Islam, 2015), and *Vanda tessellate* (Bhattacharjee et al., 2015), have been shown to contain phytochemicals, such as alkaloids, phenols, tannins, coumarins, terpenoids, and flavonoids. Recent studies detected alkaloids and stilbenoids from the aerial parts of *D. officinale* and *A. graminifolia*, respectively (Auberon et al., 2016; Chen et al., 2019). High concentrations of phenylpropanoid, flavone, and bibenzyl were found in the aerial parts as well as roots of *A. graminifolia* (**Figure 5D**).

Cytochrome 450 is a vital regulator of secondary metabolites biosynthesis, such as bibenzyl (Adejobi et al., 2021). CYP84A1 (39318) is a critical enzyme in the phenylpropanoid biosynthetic pathway (Anderson et al., 2015). It was specifically expressed in the leaf both in qRT-PCR and transcriptome data (**Figure 5E**). However, CYP73A4 (41477) was expressed in the root, suggesting multiple accumulation points for bibenzyls in *A. graminifolia*. UDP-glycosyltransferases (UGTs) catalyze the transfer of glycosyl groups to acceptors, such as secondary metabolites and hormones (Bowles et al., 2005). High expression of UGT73C3 (38276) and UGT88B1 (34392) was observed in flower and leaf, respectively (**Figure 5E**). Regulators of flavonoids biosynthesis pathway (ANS, PAL, FLS, and 4CCL7) were highly expressed in leaf, suggesting leaf as the potent source of flavonoids other than stem (**Figures 5D,E**). Most of the genes important in the regulation of bibenzyls and flavonoids, presented in this study, have been identified in orchids and model crops (**Supplementary Table 4**). Sesquiterpenes are usually the product of farnesyl diphosphate (FPP) regulated by MVA and MEP pathways (Schwab and WüSt, 2015). Recently, several genes have been identified in FPP biosynthesis pathway (Chen et al., 2019; Adejobi et al., 2021). Most of the genes functioning in the initial biosynthesis stages of sesquiterpenes, such as DXR, DXS, MK, PMK, HMGS, HDR, and MDS, were present in our data. This signifies the involvement of these genes in the initial biosynthesis of bibenzyl in *A. graminifolia*. Genetic modification of their expression can boost bibenzyl content through *A. graminifolia*. Moreover, the CYP450 genes, such as CYP84A1 and CYP73A4, may be the critical genes regulating bibenzyl contents along with BIBSY212.

Orchids are useful pharmaceutical plants possessing anti-inflammatory, diuretic, anti-rheumatic, antiviral, neuroprotective, anti-carcinogenic, relaxation, anticonvulsive, antitumor, wound healing, hypoglycemic, anti-aging, antimicrobial, antioxidant, antibacterial, and anti-diarrheal uses (Ghanaksh and Kaushik, 1999; Moin et al., 2012; Islam et al., 2013; Pant, 2013; Rokaya et al., 2014; Bhattacharjee et al., 2015; Dalar et al., 2015). Although many studies have reported the medicinal components and uses of orchids, this study exploits the plant parts of an easily available orchid with large size. Large sized plants of *A. graminifolia* with profound vegetative growth and short reproductive cycle can generate enough plant waste even after the production of commercial flowers, as compared with any other orchid species. While discarding the orchid waste may cause environmental and health issues (Johnson et al., 1999; Srivirojana et al., 2005), it can be effectively used to obtain precious medicinal ingredients, such as bibenzyls. We found phenylpropanoid, flavone, and bibenzyl in multiple plant parts, such as stem, leaf, root, and flower (**Figure 5D**). This suggests whole plant of *A. graminifolia* is important to extract bioactive compounds for medicinal purposes than other orchid species. The short life cycle and accumulation of bioactive ingredients in multiple plant parts make bamboo orchid a favorite plant to study the efficacy of orchids to obtain pharmaceutical products.

CONCLUSION

The bamboo orchid (*Arundina graminifolia*) is a special representative of the Orchidaceae family due to its short vegetative phase and continuous flowering pattern. Its medicinal importance is yet to be revealed at molecular levels. It grows big with long stems and large leaves, as compared with other orchids, and therefore, produces large waste after cut flower selection. This waste can be a valuable source to obtain different ingredients for drug development. Bibenzyl is a useful medicinal ingredient with numerous health benefits. Therefore, this first report discusses the molecular regulation of bibenzyls, flavonoids, and phenylpropanoids along with their multiple accumulation points in the plant body. The candidate genes included BIBSY212, CYP84A1, CYP73A4, 4CLL7, UGT88B1, UGT73C3, ANS, PAL, FLS, and CHS8, which play important role in the biosynthesis of bibenzyl, and flavonoids in *A. graminifolia*. Prominent concentrations of bibenzyl, phenylpropanoid, and flavone were found in the major plant parts, such as leaf, root, stem, flower, and fruit, suggesting multiple accumulation sites of phytochemicals in bamboo orchid. These interesting outcomes, therefore, broaden our current understanding on the medicinal efficacy of orchids.

DATA AVAILABILITY STATEMENT

The transcriptome data is deposited on the NCBI GeneBank with accession number: PRJNA844531.

AUTHOR CONTRIBUTIONS

SA: conceptualization, software, and writing-original draft. JG: data curation, formal analysis, and investigation. YW: visualization and investigation. CL: data curation and formal analysis. GZ: supervision, conceptualization, and funding acquisition. FY: supervision, conceptualization, funding acquisition, and writing-reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Laboratory for Lingnan Modern Agriculture Project (NZ2021010), the Natural Science Foundation of Guangdong province (2017A030312004), grants from the National Key R&D Program (2018YFD1000400 and 2019YFD1001003), the Guangzhou Science and Technology Project (201707010307), the Innovation Team of Modern Agricultural Industry Technology System in Guangdong Province (2021KJ121), and the Guangdong Academy of Agricultural Sciences Discipline Team Construction Project (202127 TD, BZ202006, and R2020 PY-JX018).

ACKNOWLEDGMENTS

We are thankful to funding agencies for funding and Plant Editors for providing professional services for language editing and final polishing of our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.923000/full#supplementary-material

REFERENCES

Adejobi, O.I., Guan, J., Yang, L., Hu, J.-M., Yu, A., Muraguri, S., et al. (2021). Transcriptomic analyses shed light on critical genes associated with bibenzyl biosynthesis in *Dendrobium officinale*. *Plants* 10, 633. doi: 10.3390/plants10040633

Ahmad, S., Lu, C., Gao, J., Ren, R., Wei, Y., Wu, J., et al. (2021a). Genetic insights into the regulatory pathways for continuous flowering in a unique orchid Arundina graminifolia. *BMC Plant Biol.* 21, 587. doi: 10.1186/s12870-021-03350-6

Ahmad, S., Lu, C., Wei, Y., Gao, J., Jin, J., Zheng, C., et al. (2022). The de novo transcriptome identifies important zinc finger signatures associated with flowering in the orchid *Arundina graminifolia*. *Sci. Hortic.* 291, 110572. doi: 10.1016/j.scienta.2021.110572

Ahmad, S., Lu, C., Wu, J., Wei, Y., Gao, J., Jin, J., et al. (2021b). Transcriptional cascade in the regulation of flowering in the bamboo orchid *Arundina graminifolia*. *Biomolecules* 11, 771. doi: 10.3390/biom11060771

Ai, Y., Xie, T.-X., Liu, D.-K., Tu, X.-D., Zhou, J., and Liu, Z.-J. (2019). Complete chloroplast genome of *Arundina graminifolia* (Orchidaceae). *Mitochondrial DNA Part B* 4, 2898–2899. doi: 10.1080/23802359.2019.1660281

Anderson, N.A., Bonawitz, N.D., Nyffeler, K., and Chapple, C. (2015). Loss of ferulate 5-hydroxylase leads to mediator-dependent inhibition of soluble phenylpropanoid biosynthesis in Arabidopsis. *Plant Physiol.* 169, 1557–1567. doi: 10.1104/pp.15.00294

Auberon, F., Olatunji, O.J., Krisa, S., Antheaume, C., Herbette, G., Bont,é, F., et al. (2016). Two new stilbenoids from the aerial parts of *Arundina graminifolia* (Orchidaceae). *Molecules* 21, 1430. doi: 10.3390/molecules21111430

Barbosa, E.G., Bega, L.A., Beatriz, A., Sarkar, T., Hamel, E., Do Amaral, M.S., et al. (2009). A diaryl sulfide, sulfoxide, and sulfone bearing structural similarities to combretastatin A-4. *Eur. J. Med. Chem.* 44, 2685–2688. doi: 10.1016/j.ejmech.2008.12.018

Bhattacharjee, B., and Islam, S.S. (2015). Assessment of antibacterial and antifungal activities of the extracts of *Rhynchostylis retusa* blume-A medicinal orchid. *World J. Pharm. Pharm. Sci.* 4, 74–87.

Bhattacharjee, B., Islam, T., Rahman, Z., and Islam, S. (2015). Antimicrobial activity and phytochemical screening of whole plant extracts of *Vanda tessellata* (Roxb.) Hook. *Ex. G. Don. World J. Phar. Pharmace. Sci* 4, 72–83.

Bhattacharyya, P., Kumaria, S., Diengdoh, R., and Tandon, P. (2014). Genetic stability and phytochemical analysis of the *in vitro* regenerated plants of *Dendrobium nobile* Lindl., an endangered medicinal orchid. *Meta Gene* 2, 489–504. doi: 10.1016/j.mgene.2014.06.003

Bowles, D., Isayenkova, J., Lim, E.-K., and Poppenberger, B. (2005). Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant Biol.* 8, 254–263. doi: 10.1016/j.pbi.2005.03.007

Cai, H.-L., Huang, X.-J., Nie, S.-P., Xie, M.-Y., Phillips, G.O., and Cui, S.W. (2015a). Study on *Dendrobium officinale* O-acetyl-glucomannan (Dendronan®): part III–Immunomodulatory activity *in vitro*. *Bioact. Carbohyd. Dietary Fibre* 5, 99–105. doi: 10.1016/j.bcdf.2014.12.002

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W., et al. (2015b). The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* 47, 65–72. doi: 10.1038/ng.3149

Chen, X.-J., Mei, W.-L., Cai, C.-H., Guo, Z.-K., Song, X.-Q., and Dai, H.-F. (2014). Four new bibenzyl derivatives from *Dendrobium sinense*. *Phytochem. Lett.* 9, 107–112. doi: 10.1016/j.phytol.2014.04.012

Chen, Y., Wang, Y., Lyu, P., Chen, L., Shen, C., and Sun, C. (2019). Comparative transcriptomic analysis reveal the regulation mechanism underlying MeJA-induced accumulation of alkaloids in *Dendrobium officinale*. *J. Plant Res.* 132, 419–429. doi: 10.1007/s10265-019-01099-6

Dalar, A., Guo, Y., Esim, N., Bengu, A.S., and Konczak, I. (2015). Health attributes of an endemic orchid from Eastern Anatolia, *Dactylorhiza chuhensis* Renz&Taub.–*In vitro* investigations. *J. Herbal Med.* 5, 77–85. doi: 10.1016/j.hermed.2015.02.001

Du, G., Shen, Y., Yang, L., Shu, L., Wen, M.-L., and Hu, Q.-F. (2014). Bibenzyl derivatives of *Arundina graminifolia* and their cytotoxicity. *Chem. Nat. Comp.* 49, 1019–1022. doi: 10.1007/s10600-014-0813-3

Ezenwali, M., Njoku, O., and Okoli, C. (2010). Studies on the anti-diarrheal properties of seed extract of *Monodora tenuifolia*. *Int. J. App. Res. Nat. Prod.* 2, 20–26.

Gao, X., Yang, L., Shen, Y., Shu, L., Li, X., and Hu, Q.-F. (2012). Phenolic compounds from *Arundina graminifolia* and their anti-tobacco mosaic virus activity. *Bull. Korean Chem. Soc.* 33, 2447–2449. doi: 10.5012/bkcs.2012.33.7.2447

Gao, Y., Jin, Y., Yang, S., Wu, J., Gao, X., Hu, Q., et al. (2014). A new diphenylethylene from *Arundina graminifolia* and its cytotoxicity. *Asian J. Chem.* 26, 3903. doi: 10.14233/ajchem.2014.16010

Ghanaksh, A., and Kaushik, P. (1999). Antibacterial effect of *Aerides multiflora* Roxb: a study *in vitro*. *J. Orchid Soc. India* 1, 65–68.

Gomez, C., Conejero, G., Torregrosa, L., Cheynier, V., Terrier, N., and Ageorges, A. (2011). *In vivo* grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST. *Plant J.* 67, 960–970. doi: 10.1111/j.1365-313X.2011.04648.x

Gong, Y. (2003). *Mechanisms of Erianin Anti-Tumor Angiogenesis*. Nanjing: China Pharm University.

Gong, Y.Q., Fan, Y., Wu, D.Z., Yang, H., Hu, Z.B., and Wang, Z.T. (2004). *In vivo* and *in vitro* evaluation of erianin, a novel anti-angiogenic agent. *Eur. J. Cancer* 40, 1554–1565. doi: 10.1016/j.ejca.2004.01.041

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data

without a reference genome. *Nat. Biotechnol.* 29, 644. doi: 10.1038/nbt.1883

He, C., Zhang, J., Liu, X., Zeng, S., Wu, K., Yu, Z., et al. (2015). Identification of genes involved in biosynthesis of mannan polysaccharides in *Dendrobium officinale* by RNA-seq analysis. *Plant Mol. Biol.* 88, 219–231. doi: 10.1007/s11103-015-0316-z

Hernández-Romero, Y., Acevedo, L., Sánchez Mde, L., Shier, W.T., Abbas, H.K., and Mata, R. (2005). Phytotoxic activity of bibenzyl derivatives from the orchid Epi*dendrum rigidum*. *J. Agric. Food Chem.* 53, 6276–6280. doi: 10.1021/jf0508044

Hossain, M.M. (2009). Traditional therapeutic uses of some indigenous orchids of Bangladesh. *Med. Aromat. Plant Sci. Biotechnol.* 42, 101–106.

Hossain, M.M. (2011). Therapeutic orchids: traditional uses and recent advances—an overview. *Fitoterapia* 82, 102–140. doi: 10.1016/j.fitote.2010.09.007

Hu, Q.-F., Zhou, B., Huang, J.-M., Gao, X.-M., Shu, L.-D., Yang, G.-Y., et al. (2013a). Antiviral phenolic compounds from *Arundina gramnifolia*. *J. Nat. Prod.* 76, 292–296. doi: 10.1021/np300727f

Hu, Q.-F., Zhou, B., Ye, Y.-Q., Jiang, Z.-Y., Huang, X.-Z., Li, Y.-K., et al. (2013b). Cytotoxic deoxybenzoins and diphenylethylenes from *Arundina graminifolia*. *J. Nat. Prod.* 76, 1854–1859. doi: 10.1021/np400379u

Hussain, M.S., Fareed, S., Saba Ansari, M., Rahman, A., Ahmad, I.Z., and Saeed, M. (2012). Current approaches toward production of secondary plant metabolites. *J. Pharm. Bioallied Sci.* 4, 10. doi: 10.4103/0975-7406.92725

Ibdah, M., Martens, S., and Gang, D.R. (2017). Biosynthetic pathway and metabolic engineering of plant dihydrochalcones. *J. Agric. Food Chem.* 66, 2273–2280. doi: 10.1021/acs.jafc.7b04445

Islam, M., Mehraj, H., Roni, M., Shimasaki, K., and Jamal Uddin, A. (2013). Correlation between cane growth and flowering behavior of *Dendrobium orchid* cultivars. *J. Bangladesh Acad. Sci.* 37, 205–209. doi: 10.3329/jbas.v37i2.17561

Jeong, Y.J., An, C.H., Woo, S.G., Park, J.H., Lee, K.-W., Lee, S.-H., et al. (2016). Enhanced production of resveratrol derivatives in tobacco plants by improving the metabolic flux of intermediates in the phenylpropanoid pathway. *Plant Mol. Biol.* 92, 117–129. doi: 10.1007/s11103-016-0497-0

Jia, X.-L., Wang, G.-L., Xiong, F., Yu, X.-R., Xu, Z.-S., Wang, F., et al. (2015). *De novo* assembly, transcriptome characterization, lignin accumulation and anatomic characteristics: novel insights into lignin biosynthesis during celery leaf development. *Sci. Rep.* 5, 1–14. doi: 10.1038/srep08259

Johnson, M., and Janakiraman, N. (2013). Phytochemical and TLC Studies on Stem and Leaves of the Orchid Dendrobium panduratum subsp. Villosum Gopalan & AN Henry. *Indian Journal of Natural Products and Resources.* 4, 250–254.

Johnson, M.K., Alexander, K.E., Lindquist, N., and Loo, G. (1999). A phenolic antioxidant from the freshwater orchid, Habenaria *repens*. *Compar. Biochem. Physiol. Part C Pharmacol. Toxicol. Endocrinol.* 122, 211–214. doi: 10.1016/S0742-8413(98)10109-3

Kumar, A., Kumar, S., Bains, S., Vaidya, V., Singh, B., Kaur, R., et al. (2016). De novo transcriptome analysis revealed genes involved in flavonoid and vitamin C biosynthesis in *Phyllanthus emblica* (L.). *Front. Plant Sci.* 7, 1610. doi: 10.3389/fpls.2016.01610

Lei, Z., Zhou, C., Ji, X., Wei, G., Huang, Y., Yu, W., et al. (2018). Transcriptome analysis reveals genes involved in flavonoid biosynthesis and accumulation in *Dendrobium catenatum* from different locations. *Sci. Rep.* 8, 1–16. doi: 10.1038/s41598-018-24751-y

Li, M., He, Y., Peng, C., Xie, X., and Hu, G. (2018). Erianin inhibits human cervical cancer cell through regulation of tumor protein p53 via the extracellular signal-regulated kinase signaling pathway. *Oncol. Lett.* 16, 5006–5012. doi: 10.3892/ol.2018.9267

Li, Y., Li, F., Gong, Q., Wu, Q., and Shi, J. (2011). Inhibitory effects of Dendrobium alkaloids on memory impairment induced by lipopolysaccharide in rats. *Planta Med.* 77, 117–121. doi: 10.1055/s-0030-1250235

Li, Y., Wang, C., Wang, F., Dong, H., Guo, S., Yang, J., et al. (2010). Chemical constituents of *Dendrobium candidum*. *China J. Chinese Mater. Medica* 35, 1715–1719. doi: 10.4268/cjcmm20101314

Li, Y., Yang, L., Shu, L., Shen, Y., Hu, Q., and Xia, Z. (2013b). Flavonoid compounds from *Arundina graminifolia*. *Asian J. Chem.* 25, 4922. doi: 10.14233/ajchem.2013.14143

Li, Y.-K., Zhou, B., Ye, Y.-Q., Du, G., Niu, D.-Y., Meng, C.-Y., et al. (2013a). Two new diphenylethylenes from *Arundina graminifolia* and their cytotoxicity. *Bull. Korean Chem. Soc.* 34, 3257–3260. doi: 10.5012/bkcs.2013.34.11.3257

Lidan, S., Yanqiong, S., Liying, Y., Xuemei, G., and Qiufen, H. (2013). Flavonoids derivatives from *Arundina graminifolia* and their cytotoxicity. *Asian J. Chem.* 25, 8358–8360. doi: 10.14233/ajchem.2013.14743A

Liu, M., Ding, Y., and Du, L. (2007). Chemical components of traditional Dai medicine *Arundina graminifolia* (D. Don) Hochr. *Zhong Cao Yao* 38, 676–677.

Liu, M., Ding, Y., and Zhang, D. (2005). Phenanthrene constituents from rhizome of *Arundina graminifolia*. *China J. Chinese Mater. Medica* 30, 353–356.

Majumder, P., and Ghosal, S. (1993). Two stilbenoids from the orchid *Arundina bambusifolia*. *Phytochemistry* 32, 439–444. doi: 10.1016/S0031-9422(00)95011-8

Majumder, P., Sen, S., and Majumder, S. (2001). Phenanthrene derivatives from the orchid *Coelogyne cristata*. *Phytochemistry* 58, 581–586. doi: 10.1016/S0031-9422(01)00287-4

Meng, C.-Y., Niu, D.-Y., Li, Y.-K., Zhou, B., Ye, Y.-Q., Du, G., et al. (2014). A new cytotoxic stilbenoid from *Arundina graminifolia*. *Asian J. Chem.* 26, 2411–2413. doi: 10.14233/ajchem.2014.16082

Meng, Y., Yu, D., Xue, J., Lu, J., Feng, S., Shen, C., et al. (2016). A transcriptome-wide, organ-specific regulatory map of *Dendrobium officinale*, an important traditional Chinese orchid herb. *Sci. Rep.* 6, 1–13. doi: 10.1038/srep18864

Minh, T.N., Khang, D.T., Tuyen, P.T., Minh, L.T., Anh, L.H., Quan, N.V., et al. (2016). Phenolic compounds and antioxidant activity of Phalaenopsis orchid hybrids. *Antioxidants* 5, 31. doi: 10.3390/antiox5030031

Miyazawa, M., Shimamura, H., Nakamura, S.-I., and Kameoka, H. (1997). Antimutagenic activity of gigantol from *Dendrobium nobile*. *J. Agric. Food Chem.* 45, 2849–2853. doi: 10.1021/jf9603902

Moin, S., Sahaya, B., Servin, P., and Chitra, B. (2012). Bioactive potential of Coelogyne stricta (D. Don) Schltr: an ornamental and medicinally important orchid. *J. Phar. Res* 5, 2191–2196.

Ng, T.B., Liu, J., Wong, J.H., Ye, X., Sze, S.C.W., Tong, Y., et al. (2012). Review of research on Dendrobium, a prized folk medicine. *Appl. Microbiol. Biotechnol.* 93, 1795–1803. doi: 10.1007/s00253-011-3829-7

Niu, D.-Y., Han, J.-M., Kong, W.-S., Cui, Z.-W., Hu, Q.-F., and Gao, X.-M. (2013). Antiviral fluorenone derivatives from *Arundina gramnifolia*. *Asian J. Chem.* 25, 9514–9516. doi: 10.14233/ajchem.2013.15052

Pan, X., Welti, R., and Wang, X. (2010). Quantitative analysis of major plant hormones in crude plant extracts by high-performance liquid chromatography–mass spectrometry. *Nat. Protoc.* 5, 986. doi: 10.1038/nprot.2010.37

Panda, A.K., and Mandal, D. (2013). The folklore medicinal orchids of Sikkim. *Anc. Sci. Life* 33, 92. doi: 10.4103/0257-7941.139043

Pandey, S., Goel, R., Bhardwaj, A., Asif, M.H., Sawant, S.V., and Misra, P. (2018). Transcriptome analysis provides insight into prickle development and its link to defense and secondary metabolism in *Solanum viarum* Dunal. *Sci. Rep.* 8, 1–12. doi: 10.1038/s41598-018-35304-8

Pant, B. (2013). Medicinal orchids and their uses: Tissue culture a potential alternative for conservation. *Afr. J. Plant Sci.* 7, 448–467. doi: 10.5897/AJPS2013.1031

Peled-Zehavi, H., Oliva, M., Xie, Q., Tzin, V., Oren-Shamir, M., Aharoni, A., et al. (2015). Metabolic engineering of the phenylpropanoid and its primary, precursor pathway to enhance the flavor of fruits and the aroma of flowers. *Bioengineering* 2, 204–212. doi: 10.3390/bioengineering2040204

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555. doi: 10.1126/science.1073374

Rokaya, M.B., Uprety, Y., Poudel, R.C., Timsina, B., Münzbergov,á, Z., Asselin, H., et al. (2014). Traditional uses of medicinal plants in gastrointestinal disorders in Nepal. *J. Ethnopharmacol.* 158, 221–229. doi: 10.1016/j.jep.2014.10.014

Schwab, W., and WüSt, M. (2015). Understanding the constitutive and induced biosynthesis of mono-and sesquiterpenes in grapes (*Vitis vinifera*): a key to unlocking the biochemical secrets of unique grape aroma profiles. *J. Agric. Food Chem.* 63, 10591–10603. doi: 10.1021/acs.jafc.5b04398

Shen, C., Guo, H., Chen, H., Shi, Y., Meng, Y., Lu, J., et al. (2017). Identification and analysis of genes associated with the synthesis of bioactive constituents in *Dendrobium officinale* using RNA-Seq. *Sci. Rep.* 7, 187. doi: 10.1038/s41598-017-00292-8

Shubha, J., and Chowdappa, S. (2016). Phytochemical analysis and antibacterial activity of *Cymbidium aloifolium* L. a medicinal orchid from Western Ghats of Karnataka, India. *Int. J. Adv. Sci. Res. Pub.* 2, 19–23.

Srivirojana, N., Theptepa, T., Punpuing, S., Guest, P., Tun, K., Chankham, O., et al. (2005). "Population pressure, utilization of chemicals in agriculture, health outcomes and solid waste management," in *Proceedings of the International Conference on Integrated Solid Waste Management in Southeast Asian Cities, Siem Reap. Cambodia)*, 5–7.

Su, P. (2011). *Research on the Molecular Mechanism of Erianin Anti-Hepatoma Effect*. Beijing: University of Chinese Academy of Sciences.

Suzuki, M., Hayakawa, Y., Aoki, K., Nagase, H., Nakamura, H., Yamada, K., et al. (1973). Stereochemistry of intermediates in the syntheses of Dendrobium alkaloids. *Tetrahedron Lett.* 14, 331–334. doi: 10.1016/S0040-4039(01)95654-0

Tang, H., Zhao, T., Sheng, Y., Zheng, T., Fu, L., and Zhang, Y. (2017). *Dendrobium officinale* Kimura et Migo: a review on its ethnopharmacology, phytochemistry, pharmacology, and industrialization. *Evid. Based Complement. Alter. Med.* 2017, 7436259. doi: 10.1155/2017/7436259

Tokuhara, K., and Mii, M. (2003). Highly-efficient somatic embryogenesis from cell suspension cultures of Phalaenopsis orchids by adjusting carbohydrate sources. *In Vitro Cell. Dev. Biol. Plant* 39, 635–639. doi: 10.1079/IVP2003466

Wong, D.C., Pichersky, E., and Peakall, R. (2017). The biosynthesis of unusual floral volatiles and blends involved in orchid pollination by deception: current progress and future prospects. *Front. Plant Sci.* 8, 1955. doi: 10.3389/fpls.2017.01955

Xiaohua, D., Jin, Z., Hui, W., Haifeng, C., Chao, Z., and Zepu, Y. (2015). Effect of Yajieshaba, a preparation of Dai indigenous medicine, on enhanced liver detoxification. *J. Trad. Chinese Med.* 35, 197–205. doi: 10.1016/S0254-6272(15)30028-5

Yahyaa, M., Ali, S., Davidovich-Rikanati, R., Ibdah, M., Shachtier, A., Eyal, Y., et al. (2017). Characterization of three chalcone synthase-like genes from apple (Malus x domestica Borkh.). *Phytochemistry* 140, 125–133. doi: 10.1016/j.phytochem.2017.04.022

Yang, J.X., Wang, H., Lou, J., Li, L., Liu, G.Y., Gao, X., et al. (2014). A new cytotoxic diphenylethylene from *Arundina graminifolia*. *Asian J. Chem.* 26, 4517–4518. doi: 10.14233/ajchem.2014.16494

Yip, A.M., and Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinform.* 8, 1–14. doi: 10.1186/1471-2105-8-22

Yuan, Y., Yu, M., Jia, Z., Song, X.E., Liang, Y., and Zhang, J. (2018). Analysis of *Dendrobium huoshanense* transcriptome unveils putative genes associated with active ingredients synthesis. *BMC Genomics* 19, 1–16. doi: 10.1186/s12864-018-5305-6

Zhang, C., Chongsuvivatwong, V., Keawpradub, N., and Lin, Y. (2012). Analysis of prescription database extracted from standard textbooks of traditional Dai medicine. *J. Ethnobiol. Ethnomed.* 8, 1–6. doi: 10.1186/1746-4269-8-34

Zhang, J., He, C., Wu, K., Teixeira Da Silva, J.A., Zeng, S., Zhang, X., et al. (2016). Transcriptome analysis of *Dendrobium officinale* and its application to the identification of genes associated with polysaccharide synthesis. *Front. Plant Sci.* 7, 5. doi: 10.3389/fpls.2016.00005

Zhang, X., Xu, J.-K., Wang, J., Wang, N.-L., Kurihara, H., Kitanaka, S., et al. (2007). Bioactive bibenzyl derivatives and fluorenones from *Dendrobium nobile*. *J. Nat. Prod.* 70, 24–28. doi: 10.1021/np060449r

Zhao, W., Ye, Q., Tan, X., Jiang, H., Li, X., Chen, K., et al. (2001). Three new sesquiterpene glycosides from *Dendrobium nobile* with immunomodulatory activity. *J. Nat. Prod.* 64, 1196–1200. doi: 10.1021/np0102612

Zhu, J.-H., Cao, T.-J., Dai, H.-F., Li, H.-L., Guo, D., Mei, W.-L., et al. (2016). *De novo* transcriptome characterization of *Dracaena cambodiana* and analysis of genes involved in flavonoid accumulation during formation of dragon's blood. *Sci. Rep.* 6, 38315. doi: 10.1038/srep38315

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fpls.2022.942591

Transcriptional Proposition for Uniquely Developed Protocorm Flowering in Three Orchid Species: Resources for Innovative Breeding

Sagheer Ahmad[1], Jinliao Chen[1], Guizhen Chen[1], Jie Huang[1], Yang Hao[1], Xiaoling Shi[1], Yuying Liu[1], Song Tu[1], Yuzhen Zhou[1], Kai Zhao[1,2], Siren Lan[1], Zhongjian Liu[1] and Donghui Peng[1]**

[1] Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, China, [2] College of Life Sciences, Fujian Normal University, Fuzhou, China

OPEN ACCESS

Edited by:
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

Reviewed by:
Sarfraz Shafiq,
Western University, Canada
Tairq Pervaiz,
University of California, Riverside, United States
Arshad Iqbal,
University of Swat, Pakistan

***Correspondence:**
Zhongjian Liu
zjliu@fafu.edu.cn
Donghui Peng
fjpdh@126.com

Specialty section:
This article was submitted to Plant Development and EvoDevo, a section of the journal Frontiers in Plant Science

Received: 12 May 2022
Accepted: 01 June 2022
Published: 28 June 2022

Citation:
Ahmad S, Chen J, Chen G, Huang J, Hao Y, Shi X, Liu Y, Tu S, Zhou Y, Zhao K, Lan S, Liu Z and Peng D (2022) Transcriptional Proposition for Uniquely Developed Protocorm Flowering in Three Orchid Species: Resources for Innovative Breeding. Front. Plant Sci. 13:942591. doi: 10.3389/fpls.2022.942591

During orchid seed culture, seeds germinate as protocorms, and protocorms normally develop into plant with leaves and roots. Orchids require many years of vegetative development for flowering. However, under a certain combination of growth cultures, we observed that protocorms can directly flower without leaves and roots. Therefore, we performed comparative transcriptome analysis to identify the different transcriptional regulators of two types of protocorms of *Cymbidium ensifolium*, *Cymbidium sinense,* and *Cymbidium goeringii*. Zinc finger, MYB, AP2, and bHLH were the most abundant transcription factor (TF) families in the transcriptome. Weighted gene coexpression network analysis (WGCNA) was performed to identify hub genes related to leaf and flower development. The key hubs included *SPL6*, *SVP*, *SEP2*, *KNOX1*, *AP2*, *OFP1*, *COL12*, *MYB13*, *MYB36*, *MYB59*, *bHLH086*, and *ARF7*. The hub genes were further validated through statistical tools to propose the roles of key TFs. Therefore, this study initiates to answer that why there is no leaf initiation and root development and how can protocorm bypass the vegetative phase to flower? The outcomes can direct future research on short-span flowering in orchids through protocorms.

Keywords: surprised flowering, non-vegetative, WGCNA, transcription factors, orchid development

INTRODUCTION

Thew orchid industry provides the aesthetic nutrition to mankind. However, the most beautiful orchids spend a vegetative period of 2–3 years to reach flowering (Ahmad et al., 2021a, 2022a). Therefore, flowering time regulation is the most important topic in the orchid research. Orchids are grown through protocorms. Protocorm is the first stage of embryo development after seed

Abbreviations: AP, APETALA; ARF, auxin response factor; ATH1, Arabidopsis thaliana Homeobox 1; BHLH, basic helix–loop–helix; CO, CONSTANS; COL12, CONSTANS LIKE12; DOH1, Dendrobium Orchid Homeobox1; FT, Flower locus T; FUL: FRUITFUL; KNOX1, KNOTTED1-like homeobox; LFY, LEAFY; MYB, v-myb avian myeloblastosis viral oncogene homolog; OFP1, Ovate Family Protein1; PLB, protocorm like body; POH1, Phalaenopsis Orchid Homeobox1; SAM, Shoot apical meristem; SEP, SEPALLATA; SOC, SUPPRESSION OF OVEREXPRESSION OF CONSTANS; SPL, SQUAMOSA PROMOTER BINDING LIKE; SVP, SHORT VEGETATIVE PHASE; TCP, Teosinte branched1/Cycloidea/Proliferating cell factor; TOM, topology matrix; WGCNA, weighted gene coexpression network analysis.

germination. This term was first introduced in 1890 by Treub to describe the early stage of lycopod germination (Arditti, 1992; Yeung, 2017). Actually, orchids develop protocorm to make a symbiotic relationship with fungus to obtain nutrients for the development of shoot apical meristem (SAM). Therefore, protocorms are molecularly distinct from zygotes, as shown in *Phalaenopsis aphrodite* (Fang et al., 2016).

After fertilization, the plant embryo becomes a miniature sporophyte (Fang et al., 2016). Two distinct phases called morphogenesis and maturation occur during embryogenesis (Bentsink and Koornneef, 2008; Braybrook and Harada, 2008). Plant body components, such as functionally organized domains, apical-basal polarity, cell differentiation, and tissue specification, are organized during morphogenesis (Steeves and Sussex, 1989). However, sometimes the plants skip some phases of growth and exhibit strange phenotypes. These are abnormal changes caused by intrinsic and extrinsic factors. Genetic changes appear as strange phenotypes. Transcription factors (TFs) play important roles in the regulation of cell division planes and axis polarity (Jeong et al., 2012; Ueda and Laux, 2012).

Protocorm-like bodies (PLBs) are similar to protocorms but emerge from calluses or explants (Jones and Tisserat, 1990; Chugh et al., 2009). During PLB initiation, callus cells construct compact regions composed of meristemoids called promeristems (Lee et al., 2013). PLBs can be multiplied by cuttings, providing a quick source of obtaining orchids through clonal propagation (Yam and Arditti, 2009). The *KNOTTED1*-like homeobox (*KNOX*) genes play important role in PLB regeneration (Fang et al., 2016).

Homeobox genes play vital roles during the plant developmental process (Semiarti et al., 2014). A homeobox gene in maize codes for SAM and inflorescence meristem localized protein (Ritter et al., 2002). *KNOX* genes are recognized as TFs involving the aboveground organ development and SAM maintenance (Ritter et al., 2002; Scofield et al., 2007; Semiarti et al., 2014). KNOX1 proteins maintain a high level of cytokinin and a low level of gibberellin (GA) in SAM (Semiarti et al., 2014). In the orchids, *Dendrobium Orchid Homeobox1 (DOH1)* strongly expresses in vegetative SAM but moderately expresses in transitional SAM and floral buds (Yu et al., 2000). Its overexpression completely suppresses shoot organization in orchids. It is also an upstream regulator of *DOMADS1* (Yu et al., 2000). The *Phalaenopsis Orchid Homeobox1 (POH1)* involves the regulation of protocorm, seedling development, and the floral transition in the *in vitro* cultures of *Phalaenopsis* orchids (Semiarti et al., 2008). Abnormal leaf and shoot phenotype were observed in *Phalaenopsis amabilis* mutants with a defective C-terminal POH1 locus (Rahayu Sulistianingsih, 2013).

However, *DfKN1-4* genes from *Dactylorhiza fuchsia* expressed during floral development (Box et al., 2012; Rudall et al., 2013), suggesting the dual role of *KNOX* genes in vegetative and reproductive development. The expression timing of *KNOX* is crucial to establish a diverse range of floral morphologies (Rudall et al., 2013; Semiarti et al., 2014). Excessive functioning of KNOX proteins in various plant organs, and the misexpression of *KNOX* genes may cause the appearance of adventitious meristems from peripheral cells of explants, and the occurrence of genetic networking, leading to the growth of intact shoots that finally turn into intact plants (Semiarti et al., 2014).

The A-class MADS-box gene *AP1/AGL9* involves floral transition and flower organ development (Theißen, 2001). *AP1/SQUA*-like genes retain the conserved role of meristem identity determination (Chen et al., 2007). A number of *SEP* (*SEPALLATA*)-like E-class genes are required for floral structure formation in orchids (Salemme et al., 2013). *AP1* is a hub between flower inducing *SVP* and *SOC1*, and the determinants of floral organ identity (Honma and Goto, 2001).

R2R3-MYB TFs play important roles in flowering (Chen et al., 2022). The high expression of *FT* causes early flowering in the presence or absence of CO (CONSTANS), a B-box zinc finger TF and the main activator of *FT*. An increased expression of *MYB30* in the phloem of Arabidopsis accelerates flowering through the regulation of FT (Liu et al., 2014). The CO-independent incitement by the high expression of *MYB30* produces less *FT* expression levels as compared with those produced by CO-dependent activation. Moreover, *MYB30* dependent flowering activation is independent of FLC (Chen et al., 2022). *MYB36* regulates the transition of cells from proliferation to differentiation (Liberman et al., 2015). Ovate Family Protein1 (OFP1) regulates the growth and development of plants and together with *Arabidopsis Thaliana* Homeobox1 (ATH1), it integrates flowering time (Zhang et al., 2018). *SQUAMOSA PROMOTER BINDING-LIKE (SPL)* TFs involve flower regulation by activating MADS-box genes, such as *SOC1*, *FUL*, *LFY*, and *AP1* (Zhang et al., 2018).

With the above in mind, this study dissects the transcriptional regulation of a strange flowering that occurred in three orchid species without leaf or root formation. It opens a new era of research for rapid flowering of orchids that usually takes 2–3 years to reach flowering. TFs are the drivers of important genes, and our mining elucidates some key TFs that may serve as an input to decipher the mechanism of rapid flowering for future orchids.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The plants from three species were obtained through protocorm development in a controlled environment. The media concentrations included: 0.5 mg L^{-1} NAA, 8.0 mg L^{-1} 6-BA, 35 g L^{-1} sugar, 1.5 g L^{-1} activated carbon, 7 g L^{-1} agar, Temperature at 26 $\pm$ 2°C, photoperiod at 12 h/day, and light intensity of 2,500–3,000 Lx. The *C. sinense* plant took about 6 months to abnormal flowering without leaves and roots, while *C. ensifolium* and *C. goeringii* both took about 90–120 days to flower.

Each species showed two types of protocorm development; the abnormal flowering plants exiting the vegetative phase and normal plants with vegetative growth. Flowering samples were obtained from abnormally developed floral buds and the leaf samples were obtained from normal plants during vegetative growth. For each species, flower and leaf samples were obtained in triplicates.

RNA-Seq Library Preparation and Sequencing

Total RNA was extracted from 18 tissues (6 samples in 3 repeats) using the TaKaRa RNA extraction kit, and cDNA libraries were prepared. The mRNA was obtained using Oligotex mRNA Midi Kit (Qiagen, Germany). The RNA quality was assessed using a Nano-Drop spectrophotometer (Thermo Fisher Scientific, United States), followed by cDNA library preparation using the Illumina protocol (Ahmad et al., 2021a). The library products were evaluated with the Agilent 2200 TapeStation and Qubit®2.0 (Life Technologies, United States). The products were diluted to 10 pM for the *in situ* generation of clusters on HiSeq2500 pair-end flow cells, followed by pair-end sequencing (2 × 100 bp). Finally, transcriptome *de novo* was performed with the Trinity program using default parameters (Grabherr et al., 2011). Gene expressions were quantified using fragments per kilobase per transcript per million mapped reads (FPKM).

Sample Correlation and Principal Component Analysis

The Pearson correlation coefficient of gene expression between samples was calculated, and the results were shown in the form of a heatmap.

Principal component analysis (PCA) is a multivariate statistical analysis method that reduces multiple variables into a few independent variables (i.e., principal components) through dimensionality reduction, while retaining as much original data information as possible. In the analysis of the transcriptome, PCA reduces the dimensionality of a large number of gene expression information contained in the sample into a few principal components that are not mutually independent, so as to compare samples, and it is convenient to find outlier samples and discriminate samples with high similarity.

The PCA analysis was performed using the princomp function in the R software, and the ggplot2 package in the R software was used to draw the graphics.

Functional Annotation of Transcriptome Data

The assembled unigenes were mapped to public databases for annotation, such as non-redundant (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and KEGG ortholog (KO) databases, using BLASTX with a threshold E-value $\leq 10^{-5}$. The genes were also assessed on the protein family (PFAM) and UniProt databases with default parameters.

According to the GO and KEGG annotations, the differential genes were classified into functional categories and pathways. The phyper function in the R software was used for enrichment analysis. The p value was calculated, and corrected by FDR. The function with Q value ≤ 0.05 was regarded as significant enrichment.

Differential Expression Analysis

We used Bowtie2 to align clean reads to genome sequences, and then used RSEM (v1.2.8) with default parameters to calculate gene expression levels for each sample. RNA-seq by expectation maximization (RSEM) is a software package[1] for RNA-seq reads to calculate the gene and transcript isoform expression levels of *de novo* assembly (Li and Dewey, 2011; Langmead and Salzberg, 2012).

Differentially expressed genes (DEGs) were obtained using DEGseq R package (1.10.1) (Wang et al., 2010). The unigenes with a threshold p-value < 0.001 and the log2FC > 1 were regarded as DEGs.

Identification of TF Families

A TF is a group of protein molecules that can specifically bind to a specific sequence upstream of the 5′ end of a gene, thereby ensuring that the target gene is expressed at a specific time and space with a specific intensity.

For animal TFs, we compared all unigenes to the AnimalTFDB 2.0 database, and obtained the corresponding database links of unigenes corresponding TF family, Ensemble gene ID, and TF family.

For plant TFs, we used getorf to detect unigene's ORF, and then used hmmsearch to align the ORF to the TF protein domain (data from TF), and identified the unigene's ability according to the TF family characteristics described in PlantTFDB.

Co-expression Network Analysis

Weighted gene coexpression network analysis (WGCNA) was performed to obtain coexpression networks, as previously documented (Langfelder and Horvath, 2008; Ahmad et al., 2020, 2021a, 2022a). First, the unqualified genes were removed by using the function of goodSamplesGenes. Then, a suitable soft-threshold power was chosen based on the criteria of scale-free topology, by using a function of pickSoftThreshold. The one-gene-to-all relationship was incorporated, and the adjacency matrix was converted into topological matrix (TOM) (Yip and Horvath, 2007). The genes showing hierarchical clustering according to 1-TOM (TOM-based dissimilarity) were sorted, followed by the clustering of highly interconnected genes in the same module (Ravasz et al., 2002).

Identification of Significant Modules and Hub TFs

From the WGCNA, the module eigengene (ME), gene significance (GS), and module membership (MM) were ascertained. MM represents the gene-module degree of correlation. For highly correlated genes, the MM should be close to 1 or -1. The hub genes represent the highly connected candidates in a module. The exportNetworkToCytoscape function was used in the WGCNA package to create edges file as input for Cytoscape (version 3.9.0). The network displayed by Cytoscape was further analyzed through CytoHubba app in the Cytoscape to filter highly connected hub genes through the degree method (Chin et al., 2014).

[1] http://deweylab.biostat.wisc.edu/rsem/rsem-calculate-expression.html

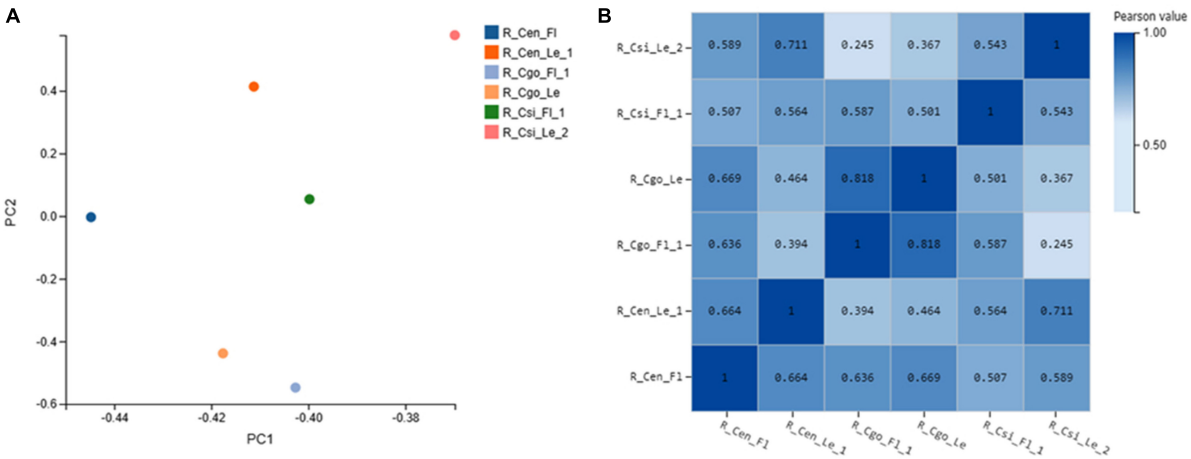


FIGURE 1 | Principal component analysis (PCA) **(A)**; and Pearson's correlation analysis **(B)**, both *X*- and *Y*-axes represent each sample. The color represents the correlation coefficient (the darker the color, the higher the correlation, and the lighter the color, the lower the correlation).

Statistical Analysis to Validate Candidate TFs

One-way ANOVA was used to validate the expression of the most important TFs.

RESULTS

In this project, a total of 6 samples were measured using the DNBSEQ platform, and each sample produced an average of 6.53 Gb of data (**Supplementary Table 1**). The average alignment rate of the samples compared with the genome was 80.47%, and the average alignment rate of the compared gene set was 67.60%. The predicted new genes were 11,539. The total number of expressed genes was 33,499, of which the known genes were 23,637, and 9,862 were predicted to be new genes.

PCA, Pearson's Correlation, and Expression Analysis

Principal component analysis represents the association among sampled tissues (**Figure 1A**). *C. goeringii* samples were closely associated as compared with other tissues or species.

The correlation coefficient reflects the overall gene expression between each sample. The higher the correlation coefficient, the more similar are the gene expression levels. The Pearson correlation coefficient between each two samples was calculated using the cor function in the R software (**Figure 1B**).

We observed the expression pattern for each tissue using the empirical cutoff values of positively expressed genes. The data are shown as boxplots (**Figure 2**). The boxplot distribution of FPKM values curtails the quartile and median values of DEGs among samples.

Pairwise Comparison of DEGs

The DEGs were compared between the flower and leaf tissues of each species and then among the flowers and leaves of three species separately (**Figure 3**). The maximum tissue specific expression difference can be seen in *C. ensifolium*; wherein considerable numbers of genes were expressed in the leaf as compared with flower. The ratio of common genes between flower and leaf was almost equal in all the comparisons. The comparison of flowers among the three species showed that more genes were expressed in *C. sinense* than *C. ensifolium* and *C. goeringii*. While in leaf comparison, more numbers of genes were expressed in *C. ensifolium*, although the difference was not significant with other species (**Figure 3**).

GO and KEGG Enrichment

Gene Ontology and KEGG enrichment analyses were performed for genes expressed in three species (**Figure 4**). The highly enriched biological processes included cellular process and metabolic process, followed by biological regulation and response to stimulus (**Figure 4A**). These four biological processes are considered the most important in the regulation of unique flower development observed in the four species. The highest numbers of genes were related to cells, followed by membrane and organelle. Regarding the molecular function, the highest numbers of genes involved catalytic activity and binding (**Figure 4A**).

The KEGG pathway enrichment showed that most of the genes were involved in the metabolic pathways and carbohydrate metabolism, which are important pathways in the regulation of unique flowering (**Figure 4B**). Other important pathways included signal transduction, translation, environmental adaptation, biosynthesis of secondary metabolites, and transport (**Figure 4B**).

Transcription Factors

A total of 2,075 TFs were identified from the transcriptome data. After removing the uncharacterized or low/null expression profiles, 1,373 TFs were selected for further analysis (**Supplementary Table 2**). The selected TFs were divided into 40 TF families (**Figure 5A**). The prominent TF families with higher

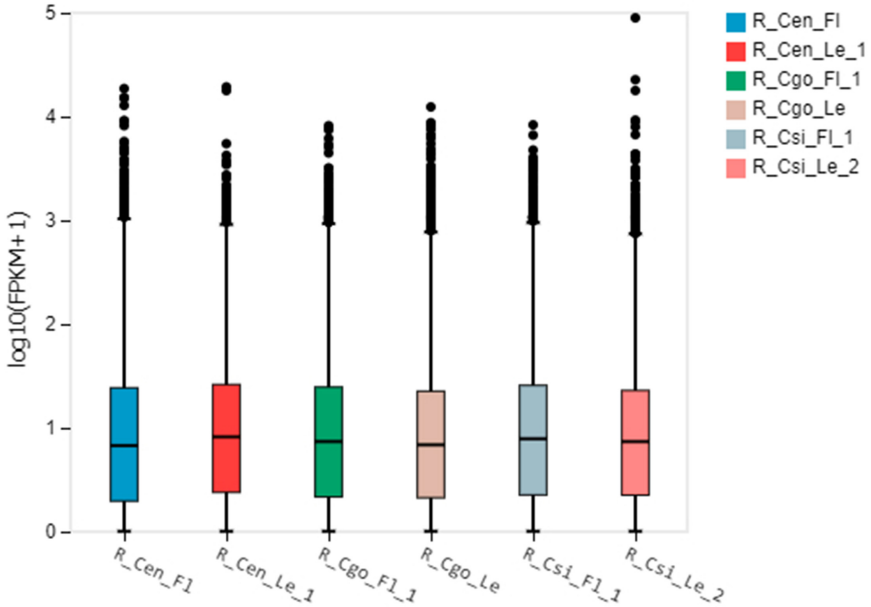


FIGURE 2 | The boxplot shows the distribution of gene expression levels in each sample, and the degree of dispersion of the data distribution can be observed. The *X*-axis is the sample name, the *Y*-axis is log10(FPKM + 1), and the boxplot of each area corresponds to five statistics (from top to bottom are the upper limit, upper quartile, median, and lower quartile, respectively). Number of digits, lower bound, where upper and lower bounds do not account for outliers.

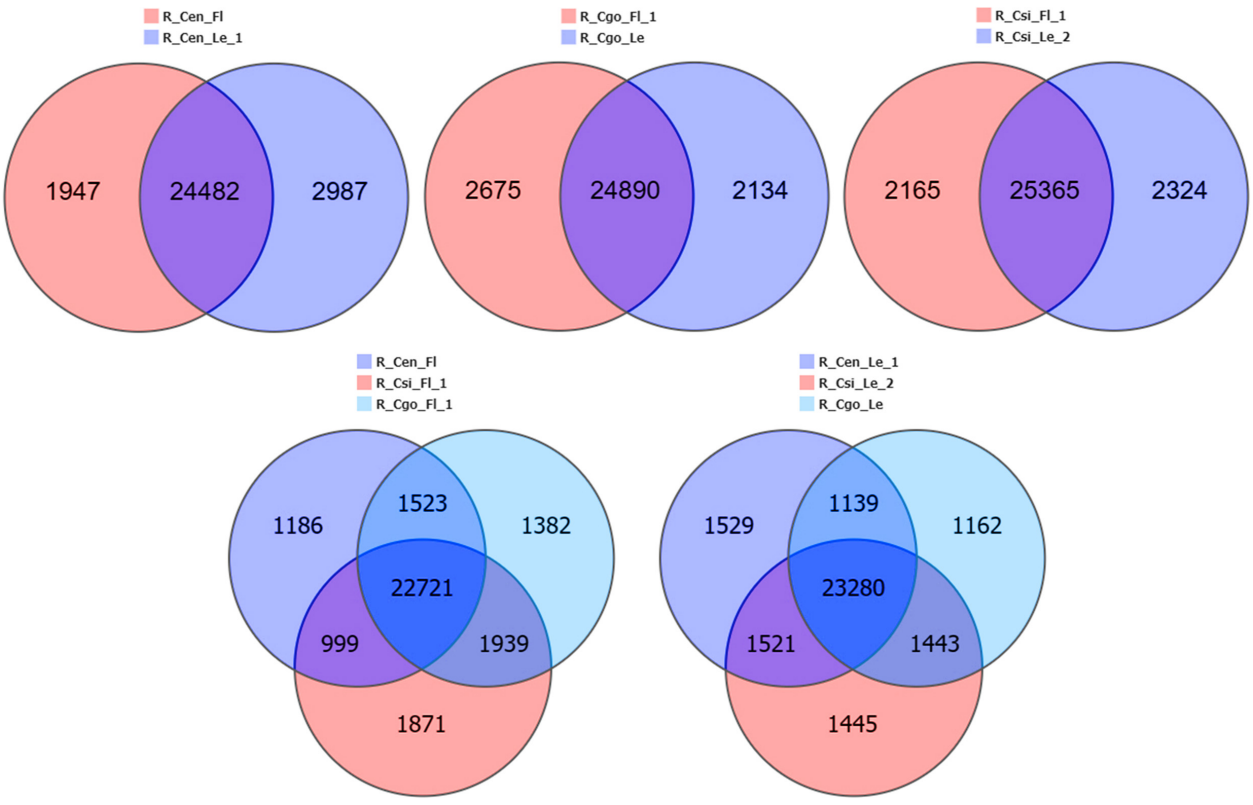


FIGURE 3 | Tissue-specific and common differentially expressed genes (DEGs) among the flower and leaf samples of three orchid species.

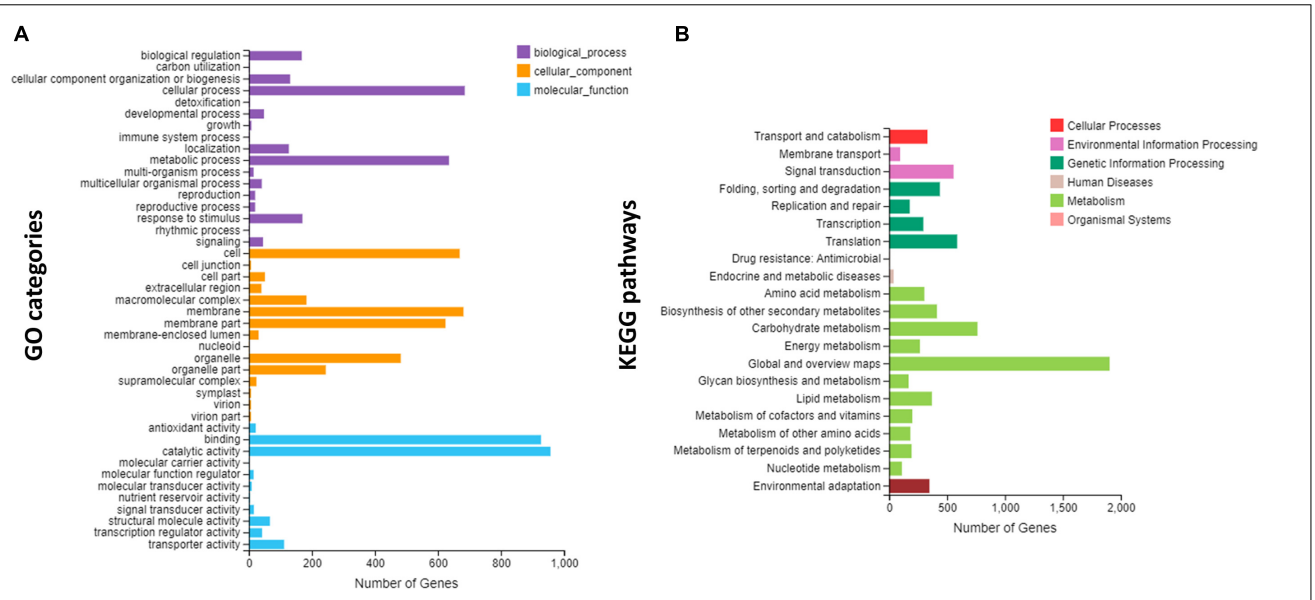


FIGURE 4 | Overview of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichments; **(A)** the GO enrichment, and **(B)** the KEGG enrichment.

number of DEGs included zinc finger, MYB, bHLH, and AP2-EREBP (**Figure 5A**). The most upregulated expression was shown by zinc finger TFs, showing downregulation in *C. ensifolium* and upregulation in *C. goeringii* and *C. sinense*. However, the NAC family showed upregulation in the *C. ensifolium* and *C. goeringii*, and downregulation in *C. sinense*.

A higher downregulation trend can be seen for all the families in the flowers of *C. ensifolium* as compared to other tissues or species (**Figure 5B**). Interestingly, the highest number of downregulated genes can also be seen in the leaf of *C. ensifolium*. The other two species showed an almost equal upregulation and downregulation trend.

Significantly opposite expression profiles between the flowers and leaves can be seen in the *C. goeringii* and *C. sinense* as compared with *C. ensifolium* (**Figure 5C**). A significant number of genes were upregulated in the flowers of *C. goeringii*, while a large number of genes were downregulated in the leaves. However, the trend was opposite for *C. sinense*, wherein a large number of genes were downregulated in the flowers as compared with leaves (**Figure 5C**).

The lowest difference between the upregulated and downregulated genes was observed in the leaves of *C. ensifolium* as compared with other tissues (**Figure 5D**).

Weighted Gene Coexpression Network Analysis

A WGCNA was performed to identify coexpressed gene sets in the flower and leaf transcriptomes of three orchid species (**Figure 6**). From 1,373 TFs, a total of 10 modules were ascertained with opposite expression intensities for flower and leaf (**Figure 6A**). MEturquoise, MEblack, and MEgrey were the most important modules. MEturquoise and MEgrey were upregulated in the flowers, while MEblack was upregulated in the leaf.

The top 100 highly connected hub genes were identified by using the degree method in the CytoHubba plugin in Cytoscape (Chin et al., 2014; **Figure 6B**).

The zinc finger was the most abundant family among the highly connected hubs, followed by AP2 and MYB TFs (**Figure 6C**). However, the expression levels of these TFs were mainly high in the flower of *C. ensifolium* as compared with other tissues. There were 22 AP2 and 19 zinc finger members with higher expression in flowers of three species as compared with leaf tissues. The bHLHs were only four with upregulated expressions in all the tissues as compared with other families. The MYBs were comparatively upregulated in *C. ensifolium* flowers and leaf, while their expression were mainly downregulated in other species (**Figure 6C**).

Flower and Leaf Specific TFs

We identified a number of TFs that upregulated either in the flowers of three species or leaves (**Supplementary Table 3**). There were 24 TFs that showed high expression in the flowers of three species, while they were downregulated in all the leaf samples (**Figure 7A**). These included the famous floral regulators, such as AP2-DREB-like, AP3-2, SEP2, SEP3, AGL6, ARF7, HEC3-like, Knotted-1-like3, and MYB TFs.

A total of 37 TFs were expressed only in the leaves of three species, while they were downregulated in the flowers (**Figure 7B**). These included, 5 MYB TFs, 3 bHLHs, ARF1, COL12, AP1, and 2 SPLs, which are thought to play important roles in the unique flowering pattern of orchid species.

A sub-network was constructed for the 61 TFs that were strictly specific to flower (24) and leaf (37). Through this network, we also identified important hub TFs (**Figure 7C**).

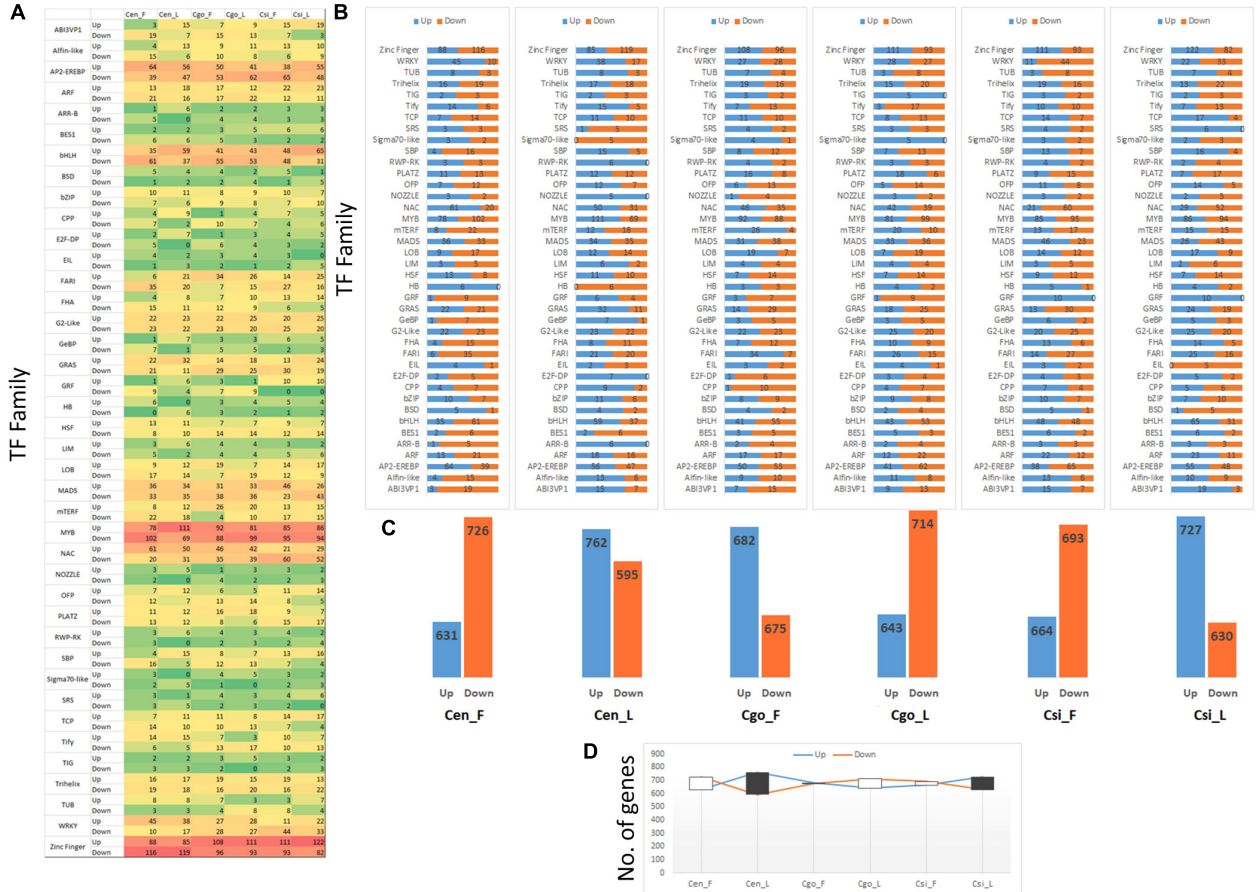


FIGURE 5 | Statistics of 40 transcription factor (TF) families **(A)**, upregulation and downregulation trend of individual TF family **(B)**, difference in the upregulation and downregulation for each tissue **(C)**, and the trend of comparative upregulation and downregulation **(D)**.

DISCUSSION

Cymbidium sinense, *C. ensifolium,* and *C. goeringii* are the most important commercial orchid species, contributing a major source of revenue in the orchid industry owing to their remarkable floral shapes. However, these orchids take 2–3 years to complete their vegetative phase for flowering. Therefore, recent strides have been directed to regulate flowering time in orchids, although nothing is achieved yet. Therefore, any development on quick flowering or narrow vegetative phase can be a great opportunity to induce flowering time regulation in the precious orchids. Luckily, the protocorms of the three orchid species showed abnormal flowering pattern in the controlled environment. Unlike the usual protocorm growth pattern, vegetative-to-reproductive growth cycle, some protocorms skipped the vegetative growth and produced flowers without leaves. This surprising abnormality led to perform a *de novo* transcriptome analysis to isolate commonly expressed genes in three species. Specifically, we concentrated on the TFs that regulate the expression of genes to drive abnormal changes.

Leaf and flower buds are shown to have similar chilling requirements but different requirements of heat, which may cause early flower bud bloom before leaf unfolding (Guo et al., 2014). Abnormal growth changes are mainly induced by physical means, such as irradiation and wounding, by hormonal changes and by affliction of external agents, such as bacteria, viruses, and fungi (Raghavan, 2000). Ectopic expression of *Brassica SHOOT MERISTEMLESS* genes caused abnormal phenotypes in Arabidopsis, which are supposed to be caused by differential levels of endogenous hormones (Elhiti and Stasolla, 2012). However, no particular study is available to reveal that how the reproductive cycle starts by skipping vegetative growth. Genetic changes are supposed to play major role to induce such changes. We have recently identified a number of TFs regulating flowering in the orchids (Ahmad et al., 2020, 2021a, 2022a). The current study identified 33,499 unigenes from six tissues of three orchid species. The DEGs were used as reference to mine 1,373 TFs (**Supplementary Table 2**), which were divided into 40 families. A number of TFs are thought to play significant roles in the regulation of abnormal flower development in the orchids. Zinc finger, MYB, bHLH, and AP2 were the most abundant TF families in the transcriptome data (**Figure 5A**). By comparing the flowers or leaves among the three species, it can be seen that there is no uniformity in the up- and downregulation trends at the

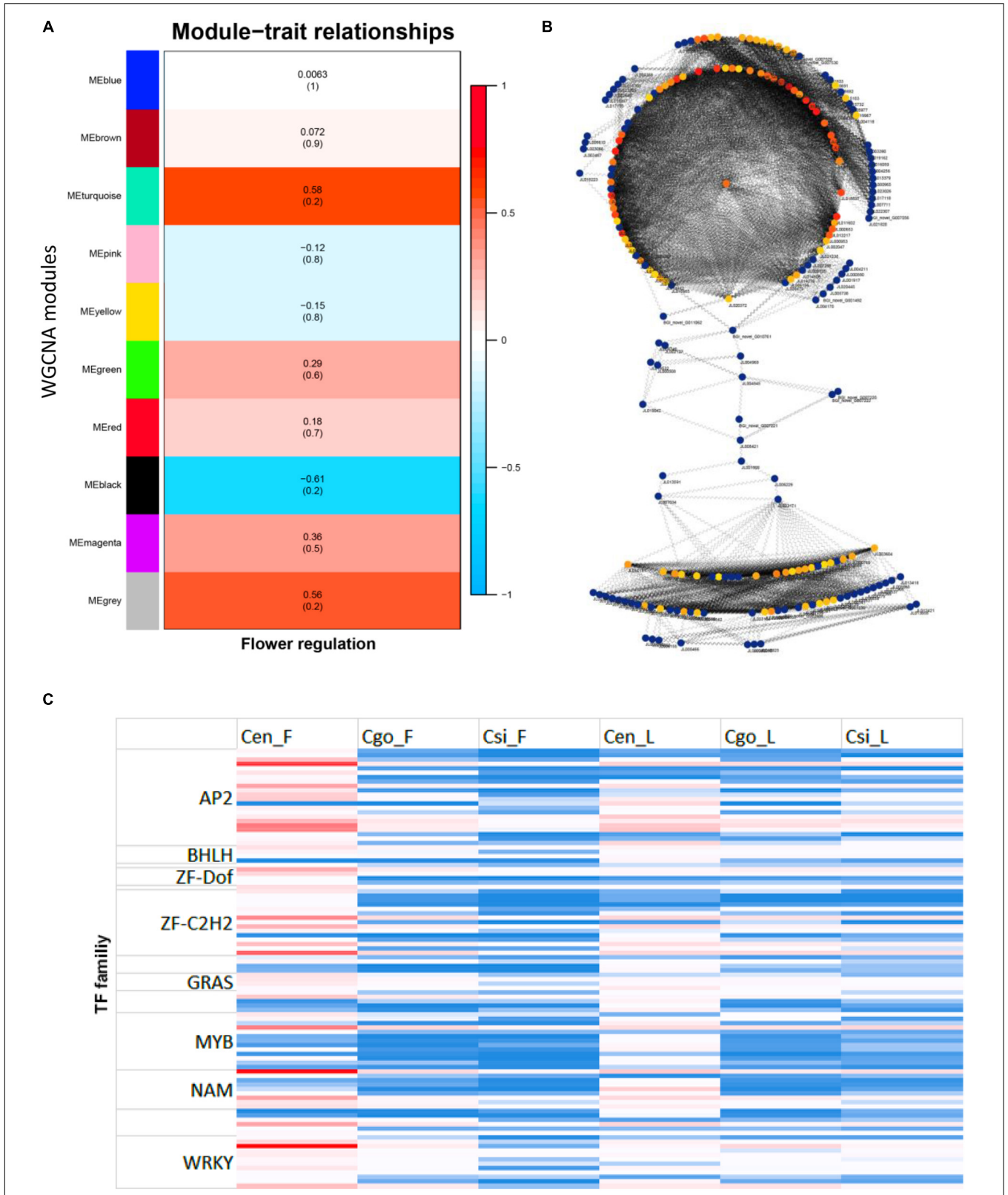


FIGURE 6 | Identification of modules through the weighted gene coexpression network analysis (WGCNA) **(A)**; interaction network by modules **(B)**, the color intensity shows the level of connectivity (red for high connectivity and yellow for low connectivity); and the abundance of TF families for top 100 highly connected TFs identified from modules **(C)**.

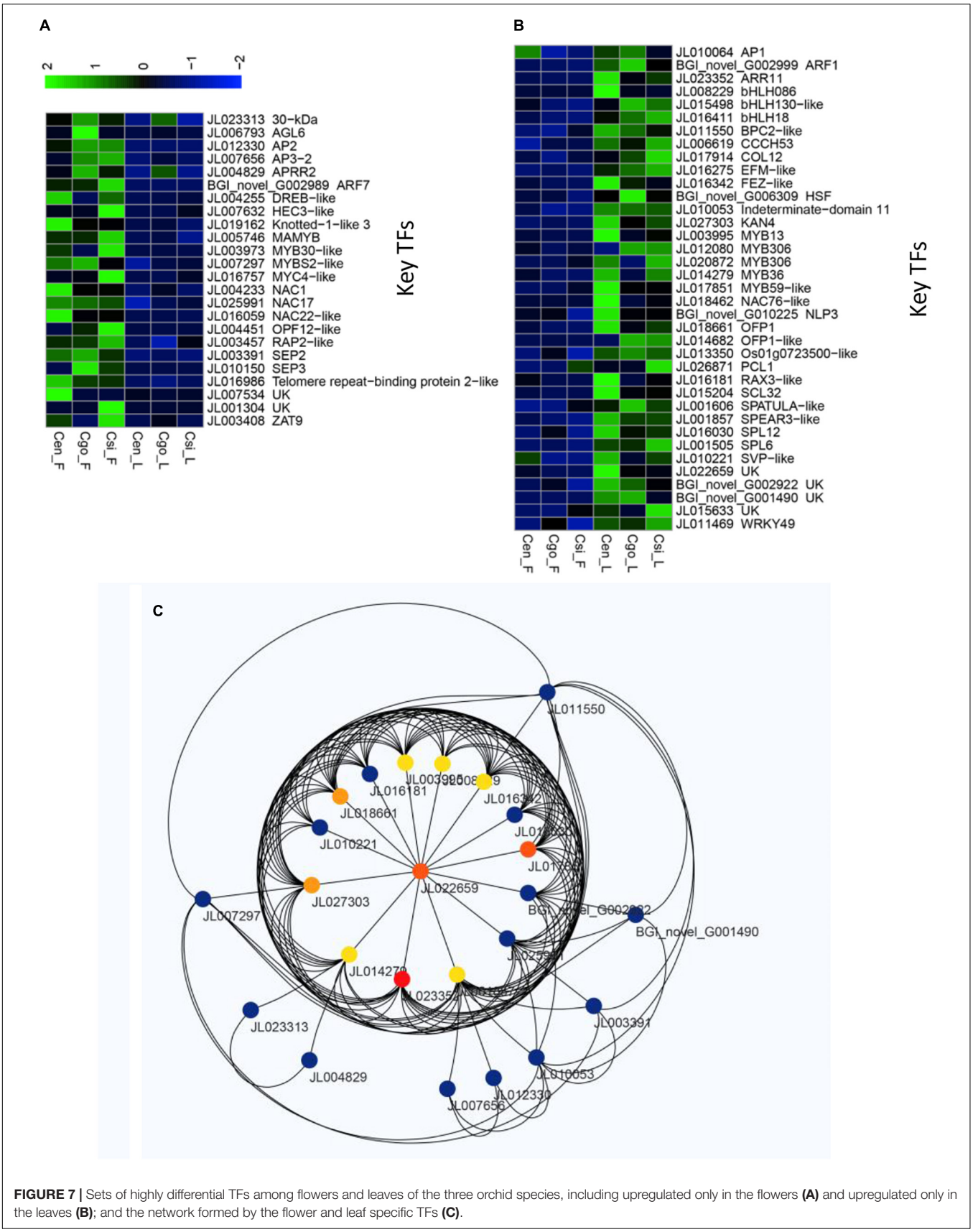


FIGURE 7 | Sets of highly differential TFs among flowers and leaves of the three orchid species, including upregulated only in the flowers **(A)** and upregulated only in the leaves **(B)**; and the network formed by the flower and leaf specific TFs **(C)**.

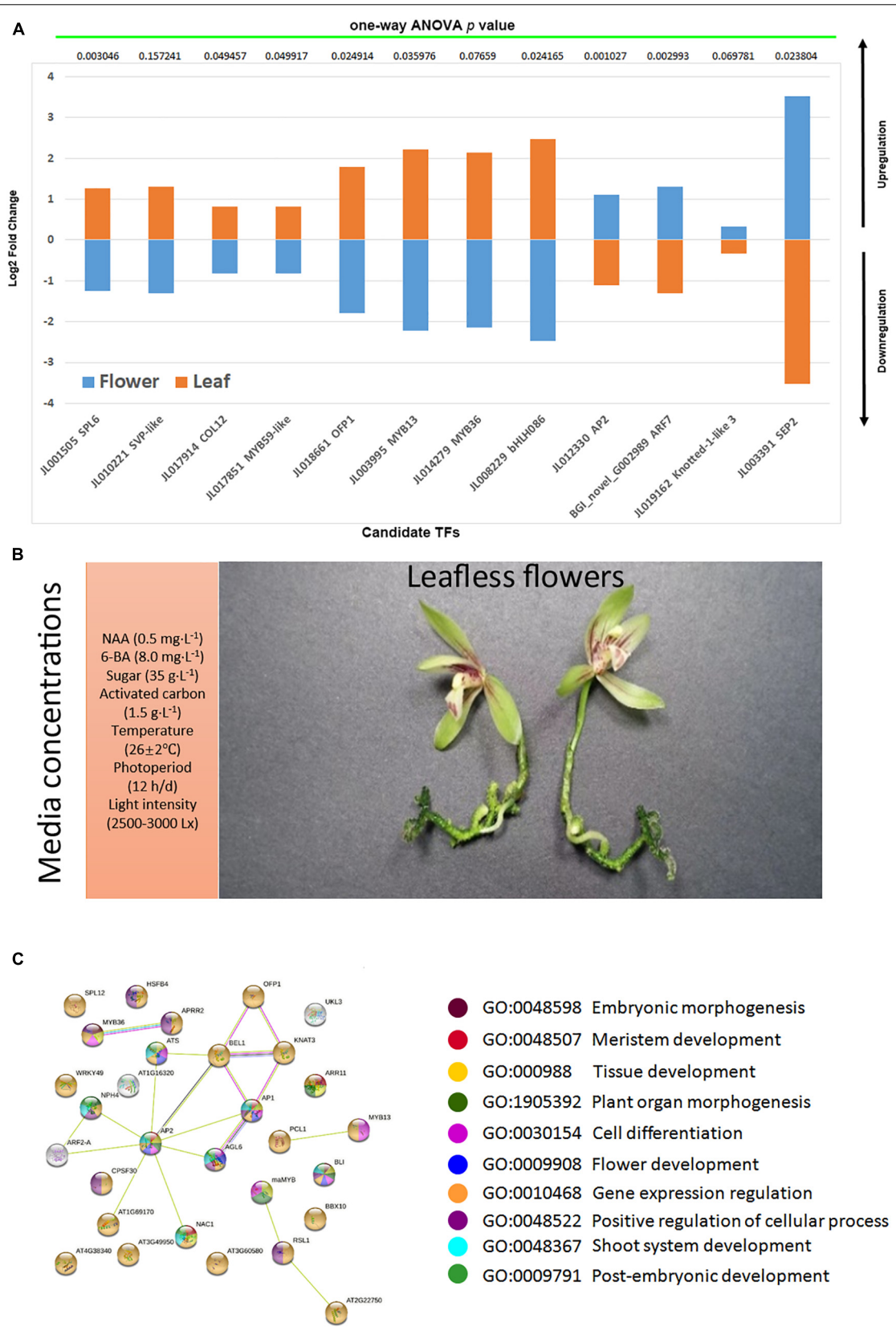


FIGURE 8 | The validation of selected candidate TFs **(A)**; the abnormal flowering along with media compositions **(B)**; the string-based interaction network for the regulation of flowering **(C)**, the ball colors represent biological process enrichment.

family levels (**Figure 5B**). Moreover, the upregulated TFs were comparatively high in *C. goeringii* flowers (**Figure 5C**), suggesting that abnormal flowering might be driven by individual TFs.

Individual TF mining was planned to isolate only those TFs which showed consistently up- or downregulation in either the flowers of three species or leaves (**Figures 7A,B**). A total of 61 potent TFs were isolated (**Figures 7A,B**). This group contained MADS-box TFs, MYBs, bHLHs, zinc fingers, AP2, and SVP TFs with admitted role in flowering regulation in orchids (Teo et al., 2019; Valoroso et al., 2019; Ahmad et al., 2021b, 2022a,b; Chen et al., 2021; Yang et al., 2021). Among the strictly flower or leaf specific TFs, 24 were upregulated in the flowers of three species (**Figure 7A**), while 37 TFs were downregulated in the flowers (**Figure 7B**). From the WGCNA modules and manual mining, we specified 12 potential TFs (*SPL6*, *SVP*-like, *COL12*, *MYB59*-like, *OFP1*, *MYB13*, *MYB36*, *bHLH086*, *AP2*, *ARF7*, *KNOX1*-like, and *SEP2*) that may play key roles in the regulation of abnormal flowering in the orchid species (**Figure 8A**). They were verified through nucleotide blast (**Supplementary Table 3**) and one-way ANOVA.

SHORT VEGETATIVE PHASE (SVP) is known as a flowering time regulator that has been suggested to play a role in continuous flowering in *Arundina graminifolia* (Ahmad et al., 2021b, 2022b). It regulates TCP TFs during bud bread (Singh et al., 2019). *SVP2* in *C. goeringii* interacts with *CgSOC1* and *CgAP1* to regulate flower development as a function of MADS-box TF (Yang et al., 2019). AP1 can act as a hub between flower induction proteins, such as SVP and SOC1, and the proteins of floral organ identity (Honma and Goto, 2001). In the photoperiod and temperature dependent pathways, *SVP* represses *FT* by interacting with *FLM* and *FLC*, thereby regulating flowering (Fujiwara et al., 2008; Gregis et al., 2009; Tao et al., 2012). It also involves hormonal pathways for ABA and GA during the bud break (Singh et al., 2018). *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* TFs play roles in the regulation of flowering time, pollen development, and phase transition (Huijser and Schmid, 2011; Preston and Hileman, 2013). *CONSTANS* (*CO*) TFs regulate flowering time in the photoperiod pathway (Ordoñez-Herrera et al., 2018). The COP1/SPA E3 ubiquitin ligase in *A. thaliana* stimulates the degradation of endogenous developmental pathway regulators. CONSTANS LIKE 12 (COL12) acts as a substrate for COP1/SPA E3 ligases to control flowering time (Ordoñez-Herrera et al., 2018). Overexpressing *COL12* instigates late flowering through suppressing the expression of *FT*. Moreover, it also suppresses flowering by suppressing the function of CO protein (Andrés and Coupland, 2012; Xu et al., 2016; Ordoñez-Herrera et al., 2018).

MYB59 is an R2R3-type TF that mainly involves cell cycle progression in Arabidopsis (Mu et al., 2009). *MYB13* functions as a transcriptional regulator of the genes involving auxin responses (Qian et al., 2020). In Arabidopsis root, *MYB36* controls the transition of cells from proliferation to differentiation (Liberman et al., 2015). *AUXIN RESPONSE FACTOR (ARF7*) is a major TF in the phototropic responses downstream of light and auxin signaling (Han and Hwang, 2018). Ovate Family Protein 1 (*OFP1*) regulates plant growth and development. It interacts with *Arabidopsis Thaliana* Homeobox 1 (*ATH1*) to coordinate flowering time and flower development (Zhang et al., 2018).

The *KNOX* genes play dual roles, regulating the vegetative and reproductive development. The expression timing of *KNOX* is crucial to establish a diverse range of floral morphologies (Rudall et al., 2013; Semiarti et al., 2014). In our case, *KNOX1* was upregulated in the flowers of three species (**Figure 8A**), suggesting a role for abnormal flowering. Along with this, ARF7, AP2, and SEP2 were also upregulated. However, the other MADS-box TFs, MYBs, and bHLH were downregulated (**Figure 8A**). This suggests that abnormal changes also suppressed a number of key floral regulators and integrators, following a different program to skip vegetative phase for rapid flowering. However, extensive functional studies are necessary to unravel the details of this unique flowering pattern.

CONCLUSION

With the above study in mind, a certain ratio of growth media components elicited an interesting flowering phenotype (**Figure 8B**), wherein vegetative phase was skipped to allow a rapid reproductive stage. The *C. sinense* took about 180 days to flower, and the *C. ensifolium* and *C. goeringii* both took 90–120 days to flower. The TF mining elucidated important hubs that may coordinate genes regulating the vegetative development, phase transition, and the fast reproductive development. *KNOX1* along with *MYB36* and *OFP1* could be the key TFs that can drive floral integrators for flowering time regulation. The hub TFs possess important biological processes (**Figure 8C**) that may contribute to such strange phenotypes. These outcomes act as a hypothetical substrate to build the experimental setup for inducing rapid flowering in orchids with short vegetative phase.

DATA AVAILABILITY STATEMENT

The transcriptome data described in this article was submitted to The National Genomics Data Center (NGDC, https://ngdc.cncb.ac.cn) under accession number: PRJCA009885.

AUTHOR CONTRIBUTIONS

SA: conceptualization and writing the original draft. JC and GC: data curation and software. JH and XS: investigation. YH: software. YL: editing the manuscript. ST: data curation. YZ: visualization, investigation, and editing the manuscript. KZ: data curation and conceptualization. SL: software and editing the manuscript. ZL: supervision, conceptualization, and funding acquisition. DP: supervision, conceptualization, funding acquisition, and writing, reviewing, and editing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (32071815); the National Key Research and Development Program of China (2019YFD1001000); the National Key Research and Development Program of China (2018YFD1000401); and the Innovation and Application Engineering Technology Research Center of Ornamental Plant Germplasm Resources in Fujian Province (115-PTJH16005).

ACKNOWLEDGMENTS

We are thankful to funding agencies for funding support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.942591/full#supplementary-material

REFERENCES

Ahmad, S., Lu, C., Gao, J., Ren, R., Wei, Y., Wu, J., et al. (2021a). Genetic insights into the regulatory pathways for continuous flowering in a unique orchid Arundina graminifolia. *BMC Plant Biol.* 21:587. doi: 10.1186/s12870-021-03350-6

Ahmad, S., Lu, C., Wu, J., Wei, Y., Gao, J., Jin, J., et al. (2021b). Transcriptional Cascade in the Regulation of Flowering in the Bamboo Orchid Arundina graminifolia. *Biomolecules* 11:771. doi: 10.3390/biom11060771

Ahmad, S., Lu, C., Wei, Y., Gao, J., Jin, J., Zheng, C., et al. (2022a). The de novo transcriptome identifies important zinc finger signatures associated with flowering in the orchid Arundina graminifolia. *Sci. Hortic.* 291:110572.

Ahmad, S., Peng, D., Zhou, Y., and Zhao, K. (2022b). The Genetic and Hormonal Inducers of Continuous Flowering in Orchids: an Emerging View. *Cells* 11:657. doi: 10.3390/cells11040657

Ahmad, S., Yuan, C., Yang, Q., Yang, Y., Cheng, T., Wang, J., et al. (2020). Morpho-physiological integrators, transcriptome and coexpression network analyses signify the novel molecular signatures associated with axillary bud in chrysanthemum. *BMC Plant Biol.* 20:145. doi: 10.1186/s12870-020-02336-0

Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* 13, 627–639. doi: 10.1038/nrg3291

Arditti, J. (1992). *Fundamentals of Orchid biology*. Hoboken: Wiley.

Bentsink, L., and Koornneef, M. (2008). Seed dormancy and germination. *Arabidopsis Book* 6:e0119.

Box, M. S., Dodsworth, S., Rudall, P. J., Bateman, R. M., and Glover, B. J. (2012). Flower-specific KNOX phenotype in the orchid Dactylorhiza fuchsii. *J. Exp. Bot.* 63, 4811–4819. doi: 10.1093/jxb/ers152

Braybrook, S. A., and Harada, J. J. (2008). LECs go crazy in embryo development. *Trends Plant Sci.* 13, 624–630. doi: 10.1016/j.tplants.2008.09.008

Chen, D., Guo, B., Hexige, S., Zhang, T., Shen, D., and Ming, F. (2007). SQUA-like genes in the orchid Phalaenopsis are expressed in both vegetative and reproductive tissues. *Planta* 226, 369–380. doi: 10.1007/s00425-007-0488-0

Chen, Q., Zhang, X., Wang, B., Xu, S., Zhao, K., Zhang, J., et al. (2022). Genome-wide identification and expression analysis of the R2R3-MYB transcription factor family revealed their potential roles in the flowering process in longan (Dimocarpus longan). *Front. Plant Sci.* 13:820439. doi: 10.3389/fpls.2022.820439

Chen, Y., Xu, Z., Shen, Q., and Sun, C. (2021). Floral organ-specific proteome profiling of the floral ornamental orchid (Cymbidium goeringii) reveals candidate proteins related to floral organ development. *Bot.. Stud.* 62:23. doi: 10.1186/s40529-021-00330-9

Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., and Lin, C.-Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8:S11. doi: 10.1186/1752-0509-8-S4-S11

Chugh, S., Guha, S., and Rao, I. U. (2009). Micropropagation of orchids: a review on the potential of different explants. *Sci. Hortic.* 122, 507–520.

Elhiti, M., and Stasolla, C. (2012). Abnormal development and altered hormone profile and sensitivity in Arabidopsis plants ectopically expressing Brassica shoot apical meristem genes. *J. Genet. Eng. Biotechnol.* 10, 23–32.

Fang, S.-C., Chen, J.-C., and Wei, M.-J. (2016). Protocorms and protocorm-like bodies are molecularly distinct from zygotic embryonic tissues in Phalaenopsis aphrodite. *Plant physiol.* 171, 2682–2700. doi: 10.1104/pp.16.00841

Fujiwara, S., Oda, A., Yoshida, R., Niinuma, K., Miyata, K., Tomozoe, Y., et al. (2008). Circadian clock proteins LHY and CCA1 regulate SVP protein accumulation to control flowering in Arabidopsis. *Plant Cell* 20, 2960–2971. doi: 10.1105/tpc.108.061531

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644. doi: 10.1038/nbt.1883

Gregis, V., Sessa, A., Dorca-Fornell, C., and Kater, M. M. (2009). The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. *Plant J.* 60, 626–637. doi: 10.1111/j.1365-313X.2009.03985.x

Guo, L., Luedeling, E., Dai, J., and Xu, J. (2014). Differences in heat requirements of flower and leaf buds make hysteranthous trees bloom before leaf unfolding. *Plant Divers. Resour* 36, 245–253.

Han, S., and Hwang, I. (2018). Integration of multiple signaling pathways shapes the auxin response. *J. Exp. Bot.* 69, 189–200. doi: 10.1093/jxb/erx232

Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083

Huijser, P., and Schmid, M. (2011). The control of developmental phase transitions in plants. *Development* 138, 4117–4129.

Jeong, S., Volny, M., and Lukowitz, W. (2012). Axis formation in Arabidopsis–transcription factors tell their side of the story. *Curr. Opin. Plant Biol.* 15, 4–9. doi: 10.1016/j.pbi.2011.10.007

Jones, D., and Tisserat, B. (1990). "Clonal propagation of orchids," in *Plant Cell and Tissue Culture*, ed. N. J. Clifton (Heidelberg: Springer), 181–191.

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

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

Lee, Y. I., Hsu, S. T., and Yeung, E. C. (2013). Orchid protocorm-like bodies are somatic embryos. *Am. J. Bot.* 100, 2121–2131. doi: 10.3732/ajb.1300193

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323

Liberman, L. M., Sparks, E. E., Moreno-Risueno, M. A., Petricka, J. J., and Benfey, P. N. (2015). MYB36 regulates the transition from proliferation to differentiation in the Arabidopsis root. *Proc.Natl Acad. Sci. U.S.A.* 112, 12099–12104. doi: 10.1073/pnas.1515576112

Liu, L., Zhang, J., Adrian, J., Gissot, L., Coupland, G., Yu, D., et al. (2014). Elevated levels of MYB30 in the phloem accelerate flowering in Arabidopsis through the regulation of FLOWERING LOCUS T. *PloS One* 9:e89799. doi: 10.1371/journal.pone.0089799

Mu, R.-L., Cao, Y.-R., Liu, Y.-F., Lei, G., Zou, H.-F., Liao, Y., et al. (2009). An R2R3-type transcription factor gene AtMYB59 regulates root growth and cell cycle progression in Arabidopsis. *Cell Res.* 19, 1291–1304. doi: 10.1038/cr.2009.83

Ordoñez-Herrera, N., Trimborn, L., Menje, M., Henschel, M., Robers, L., Kaufholdt, D., et al. (2018). The transcription factor COL12 is a substrate of the COP1/SPA E3 ligase and regulates flowering time and plant architecture. *Plant Physiol.* 176, 1327–1340. doi: 10.1104/pp.17.01207

Preston, J. C., and Hileman, L. C. (2013). Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Front. Plant Sci.* 4:80. doi: 10.3389/fpls.2013.00080

Qian, C., Chen, Z., Liu, Q., Mao, W., Chen, Y., Tian, W., et al. (2020). Coordinated transcriptional regulation by the UV-B photoreceptor and

multiple transcription factors for plant UV-B responses. *Mol. Plant* 13, 777–792. doi: 10.1016/j.molp.2020.02.015

Raghavan, V. (2000). "Abnormal Plant Growth," in *Developmental Biology of Flowering Plants*, (New York, NY: Springer), 323–337.

Rahayu Sulistianingsih, I. (2013). *PENINGKATAN VARIABILITAS GENETIK ANGGREK BULAN ALAM Phalaenopsis amabilis (L.) Blume DENGAN IRRADIASI SINAR GAMMA Co-60*. Gadjah Mada: Universitas Gadjah Mada.

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A.-L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555. doi: 10.1126/science.1073374

Ritter, M. K., Padilla, C. M., and Schmidt, R. J. (2002). The maize mutant barren stalk1 is defective in axillary meristem development. *Am. J. Bot.* 89, 203–210. doi: 10.3732/ajb.89.2.203

Rudall, P. J., Perl, C. D., and Bateman, R. M. (2013). Organ homologies in orchid flowers re-interpreted using the Musk Orchid as a model. *PeerJ* 1:e26. doi: 10.7717/peerj.26

Salemme, M., Sica, M., Gaudio, L., and Aceto, S. (2013). The OitaAG and OitaSTK genes of the orchid Orchis italica: a comparative analysis with other C-and D-class MADS-box genes. *Mol. Biol. Rep.* 40, 3523–3535. doi: 10.1007/s11033-012-2426-x

Scofield, S., Dewitte, W., and Murray, J. A. (2007). The KNOX gene SHOOT MERISTEMLESS is required for the development of reproductive meristematic tissues in Arabidopsis. *Plant J.* 50, 767–781. doi: 10.1111/j.1365-313X.2007.03095.x

Semiarti, E., Ishikawa, T., Yoshioka, Y., Ikezakki, M., Machida, Y., and Machida, C. (2008). "Isolation and charaterization of Phalaenopsis Orchid Homeobox1 (POH1) cDNAS, knotted1-like homeobox family of genes in Phalaenopsis amabilis orchid," in *Proceedings of The 2nd International Conference on Mathematics and Natural Sciences (ICMNS) ITB*, (Bandung, ID), 28–30.

Semiarti, E., Purwantoro, A., and Indrianto, A. (2014). In Vitro culture of orchids: the roles of class-1 knox gene in shoot development. *Berkala Penelitian Hayati* 20, 18–27.

Singh, R. K., Maurya, J. P., Azeez, A., Miskolczi, P., Tylewicz, S., Stojković, K., et al. (2018). A genetic network mediating the control of bud break in hybrid aspen. *Nat. Commun.* 9:4173. doi: 10.1038/s41467-018-06696-y

Singh, R. K., Miskolczi, P., Maurya, J. P., and Bhalerao, R. P. (2019). A tree ortholog of SHORT VEGETATIVE PHASE floral repressor mediates photoperiodic control of bud dormancy. *Curr. Biol.* 29, 128-133.e2. doi: 10.1016/j.cub.2018.11.006

Steeves, T. A., and Sussex, I. M. (1989). *Patterns in Plant Development*. Cambridge: Cambridge University Press.

Tao, Z., Shen, L., Liu, C., Liu, L., Yan, Y., and Yu, H. (2012). Genome-wide identification of SOC1 and SVP targets during the floral transition in Arabidopsis. *Plant J.* 70, 549–561. doi: 10.1111/j.1365-313X.2012.04919.x

Teo, Z. W. N., Zhou, W., and Shen, L. (2019). Dissecting the function of MADS-box transcription factors in orchid reproductive development. *Front. Plant Sci.* 10:1474. doi: 10.3389/fpls.2019.01474

Theißen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4, 75–85. doi: 10.1016/s1369-5266(00)00139-4

Ueda, M., and Laux, T. (2012). The origin of the plant body axis. *Curr. Opin. Plant Biol.* 15, 578–584. doi: 10.1016/j.pbi.2012.08.001

Valoroso, M. C., Sobral, R., Saccone, G., Salvemini, M., Costa, M. M. R., and Aceto, S. (2019). Evolutionary conservation of the orchid MYB transcription factors DIV, RAD, and DRIF. *Front. Plant Sci.* 10:1359. doi: 10.3389/fpls.2019.01359

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612

Xu, D., Zhu, D., and Deng, X. W. (2016). The role of COP1 in repression of photoperiodic flowering. *Research* 5:F1000. doi: 10.12688/f1000research.7346.1

Yam, T. W., and Arditti, J. (2009). History of orchid propagation: a mirror of the history of biotechnology. *Plant Biotechnol. Rep.* 3:1.

Yang, F., Lu, C., Wei, Y., Wu, J., Ren, R., Gao, J., et al. (2021). Organ-Specific Gene Expression Reveals the Role of the Cymbidium ensifolium-miR396/Growth-Regulating Factors Module in Flower Development of the Orchid Plant Cymbidium ensifolium. *Front. Plant Sci.* 12:799778. doi: 10.3389/fpls.2021.799778

Yang, F., Zhu, G., Wei, Y., Gao, J., Liang, G., Peng, L., et al. (2019). Low-temperature-induced changes in the transcriptome reveal a major role of CgSVP genes in regulating flowering of Cymbidium goeringii. *BMC Genomics* 20:53. doi: 10.1186/s12864-019-5425-7

Yeung, E. C. (2017). A perspective on orchid seed and protocorm development. *Bot. Stud.* 58:33. doi: 10.1186/s40529-017-0188-4

Yip, A. M., and Horvath, S. (2007). Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics* 8:22. doi: 10.1186/1471-2105-8-22

Yu, H., Yang, S. H., and Goh, C. J. (2000). DOH1, a class 1 knox gene, is required for maintenance of the basic plant architecture and floral transition in orchid. *Plant Cell* 12, 2143–2159. doi: 10.1105/tpc.12.11.2143

Zhang, L., Sun, L., Zhang, X., Zhang, S., Xie, D., Liang, C., et al. (2018). OFP1 interaction with ATH1 regulates stem growth, flowering time and flower basal boundary formation in Arabidopsis. *Genes* 9:399. doi: 10.3390/genes9080399

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 July 2022
doi: 10.3389/fpls.2022.939843

Recent Acquisition of Functional m6A RNA Demethylase Domain in Orchid Ty3/Gypsy Elements

Luis Alvarado-Marchena[1†], Mireya Martínez-Pérez[1†], Frederic Aparicio[1†], Vicente Pallas[1†] and Florian Maumus[2*†]*

[1] Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas, Universitat Politècnica de València, Ingeniero Fausto Elio, Spain, [2] INRAE, URGI, Université Paris-Saclay, Versailles, France

OPEN ACCESS

Edited by:
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

Reviewed by:
Sergey Morozov,
Lomonosov Moscow State University, Russia
Matej Lexa,
Masaryk University, Czechia

***Correspondence:**
Vicente Pallas
vpallas@ibmcp.upv.es
Florian Maumus
florian.maumus@inrae.fr

†ORCID:
Luis Alvarado-Marchena
orcid.org/0000-0002-1867-4543
Mireya Martínez-Pérez
orcid.org/0000-0003-1864-8473
Frederic Aparicio
orcid.org/0000-0002-1586-9978
Vicente Pallas
orcid.org/0000-0003-4954-989X
Florian Maumus
orcid.org/0000-0001-7325-0527

Specialty section:
This article was submitted to Plant Systematics and Evolution, a section of the journal Frontiers in Plant Science

Received: 18 May 2022
Accepted: 14 June 2022
Published: 04 July 2022

Citation:
Alvarado-Marchena L, Martínez-Pérez M, Aparicio F, Pallas V and Maumus F (2022) Recent Acquisition of Functional m6A RNA Demethylase Domain in Orchid Ty3/Gypsy Elements. Front. Plant Sci. 13:939843. doi: 10.3389/fpls.2022.939843

Long terminal repeats (LTR) retrotransposons are transposable elements (TEs) representing major components of most plant genomes. The fixation of additional conserved protein domains in their genomes is considered a rare event in the course of their evolution. Such changes can bring novel functions and increase their fitness by playing a role in the regulation of their replicative cycle or by affecting their integration landscape so that the detection of new domains can in turn reveal important aspects of host-TE interactions. We have mined angiosperm genomes for the presence of additional domains in LTR retrotransposons. We report a lineage of large (25 kbp) Gypsy-type elements in the genomes of Phalaenopsis orchids that contain an additional open reading frame containing a 2-ODD domain with close similarity to those responsible for m6A RNA demethylase activity in AlkB proteins. By performing *in vitro* assays, we demonstrate the RNA binding capability and the demethylase activity of the Gypsy-encoded AlkB protein, suggesting it could be functional against cognate TE mRNA or any cellular RNA *in planta*. In line with recent literature, we propose that the fixation of an RNA demethylase in this lineage of LTR retrotransposons may reflect an important role for epitranscriptomic control in host surveillance against TEs.

Keywords: epitranscriptomic, LTR retrotransposons, m6A RNA methylation, Phalaenopsis (orchids), plants, transposable elements

INTRODUCTION

Transposable elements (TEs) are genetic sequences whose primary function is to mediate their own replication and integration in host genomes from all domains of life. Owing to their duplicative properties and their repetitive nature, TEs have plethora of impacts on eukaryotic genomes such as influencing their size (e.g., TEs contribute 45 and 80% of the human and maize genomes, respectively), mediating genomic rearrangements, generating genetic diversity, distributing regulatory sequences, shaping the architecture of chromosomes and chromatin, and contributing to the definition of sexual chromosomes (Kazazian, 2004). Two main classes are recognized: class I elements (also called retrotransposons) have a so-called copy-and-paste replication mechanism involving a genomic RNA (gRNA) intermediate, whereas class 2 elements have a cut-and-paste mechanism without gRNA intermediate (Wicker et al., 2007). While TE superfamilies share a conserved backbone of protein domains and key regulatory motifs in a specific organization, they have also evolved significant changes in their architecture along their evolutionary history. Variations can be observed for instance in protein domain content and striking examples of stable acquisitions of domains have been described at variable depth in the

phylogeny of TE superfamilies (Llorens et al., 2009). Such major transitions along the evolution of TE superfamilies can have profound consequences on TE replication and regulation as well as on their interaction with host organisms.

Long terminal repeats retrotransposons are Class I TEs that present long terminal repeats (LTRs) and found throughout eukaryotes. They comprise three superfamilies: Bel/Pao, Ty1/Copia, and Ty3/Gypsy, the latter two being commonly abundant in plant genomes (Wicker et al., 2007). Both superfamilies contain the same assortment of core domains typically encoded by the gag and pol genes, i.e., the GAG, protease, integrase, reverse transcriptase (RT) and ribonuclease H domains. Several examples of protein domain acquisition were reported among LTR retrotransposons. For instance, specific lineages from both Copia- and Gypsy-type elements have acquired envelope-like (ENV) genes sharing similarity to those found in retroviruses (Laten et al., 1998; Wright and Voytas, 1998). As another example, one of the major branches of the Gypsy-type superfamily, the chromoviruses, contains a CHROMO domain that was shown to interact with histone marks and to target integration toward heterochromatin (Marin and Llorens, 2000; Gorinsek et al., 2004; Gao et al., 2008). Furthermore, giant Gypsy-type LTR retrotransposons from the planarian *Schmidtea mediterranea* were found to contain a whole collection of ORFs encoding domains generally absent from this TE superfamily (Arkhipova and Yushenova, 2019). In plants, several additional ORFs were detected across different lineages of Ty3/Gypsy elements although they most often lack known conserved protein domains (Steinbauerova et al., 2011; Neumann et al., 2019; Vicient and Casacuberta, 2020).

The activity of TEs is tightly controlled by host cell machinery at different levels of their replication cycle throughout plant development. For instance, TEs are known to be transcriptionally silenced through epigenetic regulation such as DNA methylation and post-transcriptionally inactivated through different small RNA pathways (Borges and Martienssen, 2015; Matzke et al., 2015).

In different eukaryotes, such as mammals, yeasts and plants, a large variety of RNAs can be the subject of post-transcriptional modifications. N6-methyladenosine (m6A) is one of the most common and abundant modification in RNA molecules present in eukaryotes and may be referred to as the "fifth base" of RNA. Several studies have shown that m6A participates in different aspects of RNA biology such as the regulation of mRNA stability (Wang et al., 2014), translation (Wang et al., 2015), and protein-RNA interactions (Liu et al., 2015) although recently evidence of a direct role in promoting translation has been questioned (Zaccara and Jaffrey, 2020). In contrast to mammals, studies on the function of m6A modification in plants are scarcer (Arribas-Hernandez and Brodersen, 2020). Transcriptome-wide profiles in *Arabidopsis thaliana* detected the m6A modification in over two-thirds of the mRNAs (Wan et al., 2015). m^6A depletion was described to increase the expression of m^6A-targeted mRNAs involved in plant developmental control (Shen et al., 2016) whereas Anderson et al. (2018) found that m6A stabilizes transcripts by inhibiting ribonucleolytic cleavage directly upstream of these modification sites in *Arabidopsis*. This RNA modification has been recently demonstrated to be important to regulate salt (Hu et al., 2021a; Zheng et al., 2021) and drought (He et al., 2021) stress tolerance, the infection cycle of a plant RNA virus (Martínez-Pérez et al., 2017), the stability of flowering-related genes (Duan et al., 2017), the expansion and ripening of tomato fruits (Hu et al., 2021b) or the yield and biomass of rice and potato plants (Yu et al., 2021). Thus, manipulation of plant RNA m6A methylation could be an encouraging strategy to improve plant growth and crop yield.

RNA m6A methylation is a reversible modification. In mammals, the protein complex known to add the methyl group to RNA (writers) includes the catalytic RNA methyltransferase METTL3, the allosteric activator METTL14, and several assisting factors such as WTAP. The demethylase enzymes (erasers) FTO and ALKBH5 have been shown to remove m6A from RNA (Hu et al., 2019). In plants, orthologs of METTL3 and METTL14 (MTA and MTB, respectively) and some of the assisting cofactors were characterized (Shen et al., 2016; Ruzicka et al., 2017). In fact, analysis of *Arabidopsis mta* knockdown mutants revealed that MTA is required for m6A mRNA methylation and essential for normal growth and development (Zhong et al., 2008; Yue et al., 2019). Furthermore, ALKBH5 orthologs found in *A. thaliana*, ALKBH9B and ALKBH10B, were shown to have RNA demethylation activity (Duan et al., 2017; Martínez-Pérez et al., 2017). Interestingly, while *alkbh9b* knockout mutants do not show differences in overall plant RNA m6A methylation level, its depletion results in hypermethylation of alfalfa mosaic virus (AMV) RNA and impairment of systemic infection (Martínez-Pérez et al., 2017, 2021). In addition, several studies have pointed out the varying roles of m6A modifications in the life cycle of animal- and plant-infecting RNA viruses (Gonzales-van Horn and Sarnow, 2017; Williams et al., 2019; Yue et al., 2022). More recently, RNA m6A methylation was also reported to play a role in the control of the replication of mammalian endogenous viruses and retrotransposons elements in contrasting ways. Indeed, mRNAs derived from endogenous retroviruses are methylated and the levels of m6A RNA methylation negatively affect their accumulation in mouse embryonic stem cells (ESCs) (Chelmicki et al., 2021). On the other hand, mRNAs of young LINE-1 retrotransposons are strongly methylated but positively promote LINE-1 retrotransposition in mouse ESCs (Xiong et al., 2021). While the control of TEs by balancing RNA m6A methylation levels appears as an emergent, boiling field of investigation, there is no clue at this stage how this modification may affect different types of TEs and whether it also plays a role in plants.

In this study, we explore the protein domain content of LTR retrotransposons across angiosperm genomes and found several previously unreported domains in Gypsy-type elements. Most intriguingly, we discovered a family of Gypsy-type elements in orchids that contain a functional RNA m6A demethylase acquired from the AlkB protein family.

MATERIALS AND METHODS

Identification of Protein Domains

We first searched for reverse transcriptase in 121 plant genome assemblies (**Supplementary Table 1**) representing about 90 Gb

in total). To this end, we first prepared a query library by clustering the 165 RT amino acid sequences of Gypsy and Copia reference elements taken from Gypsy database version 2 (Llorens et al., 2011) at a depth of 40% identity and we selected one representative sequence per cluster (n = 35). We then used the resulting RT library to search for homologous regions in target genomes with tBLASTn from ncbi blast + package. All overlapping hits on genomes were merged and the corresponding FASTA sequences within the expected size range (520–840 bp) were extracted (n = 1,949,052). To avoid sparing unnecessary computational time, during following steps, RT sequences from each species were clustered at a threshold of 95% identity using MMseqs2 (Steinegger and Soding, 2017) and a maximum of 50 sequences per group were selected for downward analysis. The genomic positions of RT coding regions were extended to 5 kb upstream and downstream and the corresponding sequences were extracted (n = 1,456,640). Extended hits were then clustered using mmseqs2 (with parameters -c 0.5 –max-seq-len 15000), and the groups containing at least 5 sequences were aligned with MAFFT (Katoh et al., 2002). A consensus sequence was then generated for each sequence alignment through the modules "msa2profile" (with parameters –match-mode 1 –match-ratio 0.5) and "profile2consensus," resulting in 25,565 consensus sequences. Each consensus was compared to the CDD database (Marchler-Bauer et al., 2010) using MMseqs2. The classes of domains that are commonly found across LTR retrotransposons were filtered (belonging to Gag, protease, reverse transcriptase, ribonuclease, integrase, and Chromo categories) to obtain the names of uncommon conserved protein domains detected in at least 5 consensuses established from a same species (i.e., corresponding to a TE family). To address the fraction of the consensus sequences representing LTR retrotransposons, we compared each one to a library of reference aa RT sequences from Copia, Gypsy, DIRS, endogenous retroviruses, Caulimoviridae, and LINEs using BLASTx. The consensuses corresponding to LTR retrotransposons were identified from their best hit (highest bit score) against RT from Copia or Gypsy. Intrinsically disordered regions (IDRs) within Gypsy_Pa_2799 ORF2 were predicted using PrDOS with default settings (Ishida and Kinoshita, 2007).

Phylogenetic Analysis

To build the reverse transcriptase tree, we extracted the RT domain from Gypsy_Pa_2799 ORF1 and combined it with those from 96 reference elements obtained from the Gypsy database (Llorens et al., 2011) to produce a multiple sequence alignment using MAFFT (Katoh et al., 2002) with the option "–maxiterate 1000." The resulting alignment was submitted to IQ-TREE (Minh et al., 2020) for model testing (best-fit model: LG + R6 chosen according to BIC) and phylogenetic analysis with 100 bootstrap replicates. For the AlkB tree, we first retrieved the sequences from eight predicted human AlkB proteins (Kurowski et al., 2003). The human proteins were used as queries to identify and collect homologs in *A. thaliana*. We then combined the 2-ODD domain from Gypsy_Pa_2799 ORF2 to these references sequences to produce a multiple sequence alignment using MAFFT with the options "–maxiterate 1000 –dash –originalseqonly." The resulting alignment was trimmed manually at the extremities to keep only the 2-ODD domain and then submitted to IQ-TREE for model testing (best-fit model: LG + R3 chosen according to BIC) and phylogenetic analysis with 100 bootstrap replicates.

Characterization of Gypsy-Type Elements Encoding the 2-ODD Domain

LTRharvest (Ellinghaus et al., 2008) was launched on *P. equestris* and *P. aphrodite* genomes using the options "-maxdistltr 20000 -maxlenltr 3000 -mindistltr 5000." For each genome, the LTRharvest predictions were clustered using MMseqs2 (with parameter -c 0.5) (Steinegger and Soding, 2017) and the clusters containing at least five sequences were aligned with MAFFT (Katoh et al., 2002). A consensus sequence was then generated for each sequence alignment through the modules "msa2profile" (with parameters –match-mode 1 –match-ratio 0.5) and "profile2consensus." Each consensus was compared to the CDD database (Marchler-Bauer et al., 2010) using MMseqs2 and those containing the 2-ODD domain (pfam13532) were selected (n = 15) and aligned using MAFFT. The *P. aphrodite* genome coverage was obtained by using these 15 consensus sequences as a library with RepeatMasker where accepted divergence between target and query was limited to 20% (−div 20) tRNA.

Purification of ALKB Proteins

The open reading frame of the putative Gypsy_Pa_2799 protein and the atALKBH9B gene as well as the paALKBH5-like His448Ala-mutant were chemically synthesized for subsequent periplasmic production strategy (signal peptide present in the N-terminal of the proteins) and cloned into pET28 plasmid fused to a Histidine tag at its C-terminal (ProteoGenix, Schiltigheim, France). Proteins were expressed in BL21 *Escherichia coli* cells and purified by immobilized metal affinity chromatography (IMAC) in native conditions following manufacturer recommendations.

Electrophoretic Mobility Shift Assays

Nucleic acid-protein interactions were analyzed by electrophoretic mobility shift assay (EMSA). For these assays 5 ng of DIG-labeled riboprobe of 5′UTR of AMV RNA 3 transcripts was denatured (5 min at 85°C) and cooled at room temperature for 15 min. Different amounts of purified His:paALKBH5-like fusion protein (0–360 ng) were added and incubated for 30 min at 4°C in 10 μL of binding buffer (10 mM Tris–HCl pH 8.0, 100 mM NaCl and 50% glycerol), and 1U of RiboLock RNase inhibitor (Thermo Fisher Scientific). The samples were then resolved in 1.5% TAE agarose gels, capillary transferred to positively charged nylon membranes (Hybond-N, GE Healthcare, Amersham) and fixed by UV. Finally, localization of the DIG-labeled probes was achieved by development with an anti-DIG system (Roche Diagnostics) according to the manufacturer's protocol. Northwestern assays were carried out as previously described (Pallas et al., 1999). Briefly, purified proteins (GST, His:paALKBH5-like or its mutant variant H448A) were electrophoresed in 12% SDS/PAGE and transferred to

nitrocellulose membranes. Membranes were incubated overnight at 4°C in Renaturing Buffer (10 mM Tris–HCl pH 7.5, 1 mM EDTA, 0.1 M NaCl, 0.05% Triton X-100, 1X Blocking Reagent, Roche). After this, membranes were incubated with 20 mL of the same buffer containing 50 ng/mL of the AMV sgRNA 4 labeled with digoxigenin for 3 h at 25°C.

In vitro Demethylation Assays

The m6A demethylase activity assay was performed by incubating 0.75 μg of m6A monomethylated ssRNA (Dharmacon, Inc., 5′-AUU GUC A[m6A]C AGC AGC-3′) and 2.5 μg of protein for 3 h at 37°C in a reaction mixture containing 50 mM HEPES buffer (pH 7.0), 10 μM α-ketoglutarate, 100 μM L-ascorbic acid ascorbate and 20 μM $(NH_4)2Fe(SO_4)2{\cdot}6H_2O$. Reactions were quenched by heating them at 95°C for 10 min. Digestion of nucleotides to nucleosides and the m6A/Adenosine (m6A/A) quantification by ultra-performance liquid chromatography–photodiode detector–quadrupole/time-of-flight–mass spectrometry (UPLC-PDA-TOF-MS) analysis of ssRNA oligonucleotide and total RNA were performed as described by Martínez-Pérez et al. (2017). Proteins used in this assay were the RNA-demethylase atALKBH9B (Martínez-Pérez et al., 2017) as positive control, the *P. aphrodite* Gypsy-element that codes a hsALKBH5-like (paALKBH5-like) and the paALKBH5-like His448Ala-mutant. This mutation affects the putative motif involved in Fe (II)-binding motif/iron core (H448xD. . .H), characteristic of the 2OG-FeII-dependent dioxygenase superfamily and the AlkB family (Lu et al., 2014). Three independent replicas were made with each of the proteins under study.

RESULTS

Screen for Non-canonical Domains

We searched for non-canonical protein domains embedded in LTR retrotransposons across 121 angiosperm genomes (**Supplementary Table 1**). To this end, we have first established a generic methodology to build a library of consensus sequences representing LTR retrotransposons from each plant genome. In brief, we identified genomic regions containing LTR retrotransposon sequences using the reverse transcriptase domain as a marker. The regions sharing high sequence identity were clustered and the sequences from each cluster were aligned to infer a consensus sequence. In total, we generated 25,565 consensus sequences, including 4,811 and 20,288 corresponding to Copia- and Gypsy-type elements, respectively (**Supplementary Data**). In addition, 423 consensus sequences were classified as endogenous Caulimoviridae and five were classified as LINE-type non-LTR retrotransposons. We have next scanned each consensus sequence for the presence of conserved protein domains by searching for profiles from the CDD database (Marchler-Bauer et al., 2010). After filtering those domains typically found in LTR retrotransposons, including Chromo domain models found in Chromoviruses, we have extracted the list of putative supplementary domains that were detected (**Figure 1**). Here, we focused on domains found in at least five consensus sequences from a given species (hereafter referred to as the "n5" filter) to empirically select evolutionary stable fusion events and to avoid describing potentially less meaningful fusions (charted on **Supplementary Figure 1**).

Applying this method, we detected 15 domains that are not commonly found in LTR retrotransposons, all being detected in Gypsy-type elements as opposed to Copia-type. Among the plant genomes analyzed here, seven atypical domains were found across several species, the most frequent being PMD, G-patch and AIR1.

The PMD domain (Plant Mobile Domain, PF10536), which function is unknown was the most frequent as it was detected in consensuses from ten plant species belonging to monocotyledons and dicotyledons (40 species when skipping the n5 filter). Though not considered a canonical Gypsy-encoded domain, PMD was previously described to associate with plant Gypsy-type elements (Steinbauerova et al., 2011) and MULE DNA transposons (Babu et al., 2006). In *Arabidopsis*, at least two PMD-containing proteins (MAIN and MAIL1) play a role in transcriptional gene regulation and TE silencing (Ikeda et al., 2017; Nicolau et al., 2020).

The G-patch domain (PF01585), containing a glycine-rich motif, is found in several eukaryotic RNA-processing proteins and is hypothesized to mediate binding to RNA. Here, this domain was found in LTR-RT consensus sequences from seven species (16 species when skipping the n5 filter) distributed across different clades of dicotyledons including the Rosids *Hevea brasiliensis* (order Malpighiales), the Asterids *Chenopodium quinoa* (order Caryophyllales), and *Nicotiana sp.* and *Petunia sp.* (order Solanales). Interestingly, the G-patch domain is also found in type D retroviral polyproteins (Aravind and Koonin, 1999).

Similarly, the AIR1 domain (Arginine methyltransferase-interacting protein), which contains a Zn-finger motif, was found in four different species (22 species when skipping the n5 filter) including the Asterids *Daucus carota* (order Apiales) and *Helianthus annuus* (order Asterales) as well as the Rosids *Medicago truncatula* (order Fabales) and *Cephalotus follicularis* (order Oxalidales). The AIR1 domain overlaps with the DNA binding universal minicircle sequence binding protein (UMSBP) domain which was previously described in some giant Gypsy-type LTR retrotransposons (Arkhipova and Yushenova, 2019) as well as in some non-LTR retrotransposons (Saint-Leandre et al., 2019).

Besides these most frequent domains, others were found in consensuses derived from a narrower range of species. For instance, the DUF1409 domain (PF07197), Domain of Unknown Function, was found in consensus sequences derived from seven different species from the genus *Oryza*. The taxonomic distribution of this domain appears to be restricted to the Poaceae family. Likewise, in line with a previous report (Steinbauerova et al., 2011), the YgdH domain was detected in consensus sequences obtained from two species of the *Citrus* genus, *Citrus sinensis* and *Citrus clementina* as well as in *P. trichocarpa* when n5 filter is disabled. YgdH domain is otherwise found in LONELY GUY protein family whose members play a pivotal role in regulating cytokinin activity in *Arabidopsis* (Kuroha et al., 2009) and the cell cycle in the green alga *Chlorella variabilis* (Nayar, 2020).

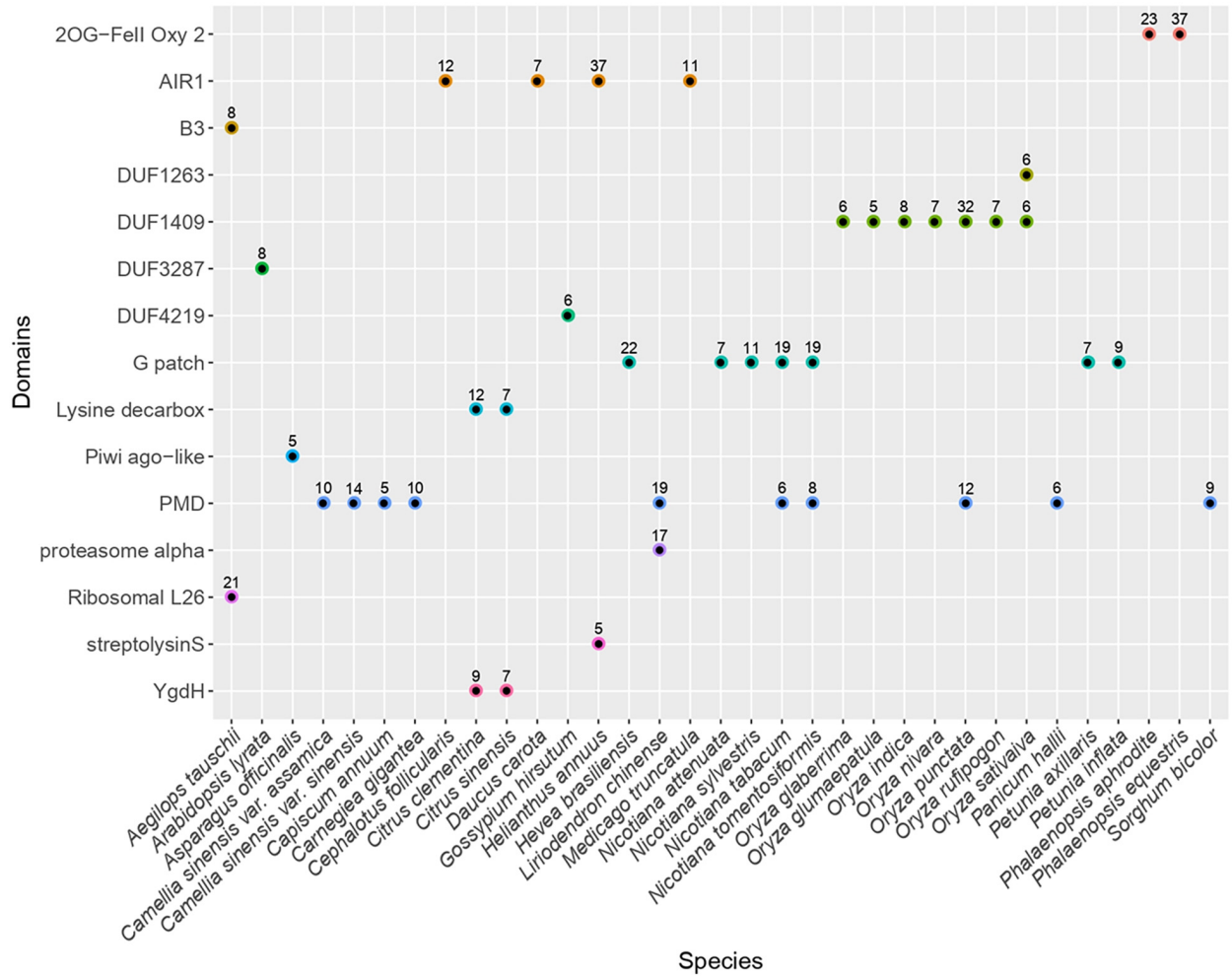


FIGURE 1 | Distribution of frequent extra conserved protein domains across LTR retrotransposons. Plot representing the distribution of the different types of non-canonical conserved protein domains across plant species. Only the domains found in at least five consensus sequences from a given species (passing the n5 filter) are shown. The numbers above each dot indicate the number of consensuses in which each domain was detected in the respective species. The canonical domains found in LTR retrotransposons as well as the hitherto well described non-canonical domains (including chromo and ATHILA domains) are not shown. The redundancy associated to hits against homologous profiles from the CDD database has been collapsed to keep only one representative domain per domain family. The full chart of non-canonical domains detected is available in **Supplementary Figure 1**.

In addition, Ribosomal protein L26 domain was found in consensus sequences from *Aegilops tauschii* only. It contains the KOW RNA-binding motif which is shared so far among several families of ribosomal proteins (Kyrpides et al., 1996). Still from the sole *A. tauschii* genome but in different consensuses, we detected the B3 domain, a DNA binding domain found in plant transcription factors with diverse functions in plant growth and development. We also identified LTR-RT consensus sequences in *Asparagus officinalis* that contain the second linker region (ArgoL2) found in PIWI- and Argonaute-like proteins. It contains residues that were described to contribute to direct the path of the phosphate backbone of nucleotides 7 through 10 in 20-nucleotides small RNAs (Faehnle et al., 2013).

Last but not least, we detected a protein domain with similarity to 2-oxoglutarate/Fe(II)-dependent dioxygenases (2-ODDs) superfamily (cl/PFAM13532) in 38 and 23 consensus sequences established from the genome of the orchids (family Orchidaceae) *Phalaenopsis equestris* and *Phalaenopsis aphrodite*, respectively (**Supplementary Data**). Such domain was not detected in the consensuses derived from any other plant genome selected for this screen (even without the n5 filter), including those from several other orchids (*Dendrobium catenatum*, *Dendrobium huoshanense*, *Gastrodia elata*, *Apostasia shenzhenica,* and *Vanilla planifolia*). 2-ODDs compose a large superfamily of non-heme iron-containing domains that catalyze highly diverse oxidative reactions including hydroxylation, demethylation and desaturations in bacteria, fungi, plants, and metazoa (Herr and Hausinger, 2018). In plants, 2-ODDs are involved in a wide range of biological processes, including DNA repair, RNA and histone demethylation, hormone biosynthesis and the production of secondary metabolites (Farrow and Facchini, 2014; Kawai et al., 2014; Martínez-Pérez et al., 2017).

Phalaenopsis 2-ODD-Containing Elements Are Among Giants

We further investigated the structure and classification of the Phalaenopsis 2-ODD-containing elements. From our initial screen for LTR-RT elements across plant genomes, none of the 2-ODD-containing consensus sequence reconstructed from Phalaenopsis genomes presented long terminal repeats, indicating that those sequences might be truncated. This artifact is methodologically possible in case of large LTR-RT elements as our approach limits the reconstruction of consensus sequence to the regions 5 kb upstream and downstream of RT domains. We used LTRharvest software (Ellinghaus et al., 2008) to investigate the structure of full length corresponding elements. LTRharvest predictions were clustered on a similarity basis and the sequences from each cluster were aligned to derive consensus sequences. Following this approach, we obtained 254 and 178 consensuses from the *P. aphrodite* and *P. equestris* genome assemblies, respectively. After searching for conserved protein domains across these sequences, we identified 15 consensuses encoding the 2-ODD domain in *P. aphrodite* (**Supplementary Data**) but none in *P. equestris*, probably due to lesser contiguity level in this genome. The alignment of the 15 *P. aphrodite* consensuses revealed a core of high sequence conservation with some consensuses presenting structural variations (**Figure 2A**). We selected one sequence which covers the whole core alignment as representative (hereafter referred to as Gypsy_Pa_2799) to analyze its structural features. The Gypsy_Pa_2799 consensus sequence is 25,510 base pairs (bp) long and comprises 6,425 bp LTRs at the extremities (**Figure 2B**). The internal sequence encodes two open reading frames (ORFs): ORF1 (2640 aa) contains all the conserved protein domains typically found in Gypsy-type elements in a single polyprotein while ORF2 (565 aa) carries the 2-ODD domain at its C-terminus (**Figure 2C**). We then addressed the overall contribution of the 2-ODD containing elements to the *P. aphrodite* genome assembly. Using the fifteen selected consensus sequences derived from LTRharvest as input library (**Figure 2A**) for RepeatMasker annotation, we obtained approximately 88 Mb coverage, i.e., 8.6% of the genome assembly size (9.7% when excluding assembly gaps).

We next investigated the phylogenetic relationship of 2-ODD-containing elements with other Gypsy elements. After extracting the aa RT domains from Gypsy_Pa_2799, we found that they are highly conserved, showing above 90% identity with each other, so we selected a single RT sequence as representative. We aligned this sequence with RT sequences taken from a diversity of reference Gypsy-type elements obtained from the Gypsy database to address their evolutionary relationships. The phylogenetic tree indicates that RT domains from 2-ODD-containing elements are sister to the Ogre element from *Pisum sativum* (Neumann et al., 2003) and nested within the Tat family (**Figure 3**). This phylogenetic placement is in line with several features of the orchid elements. First, the relatively large size compared to other

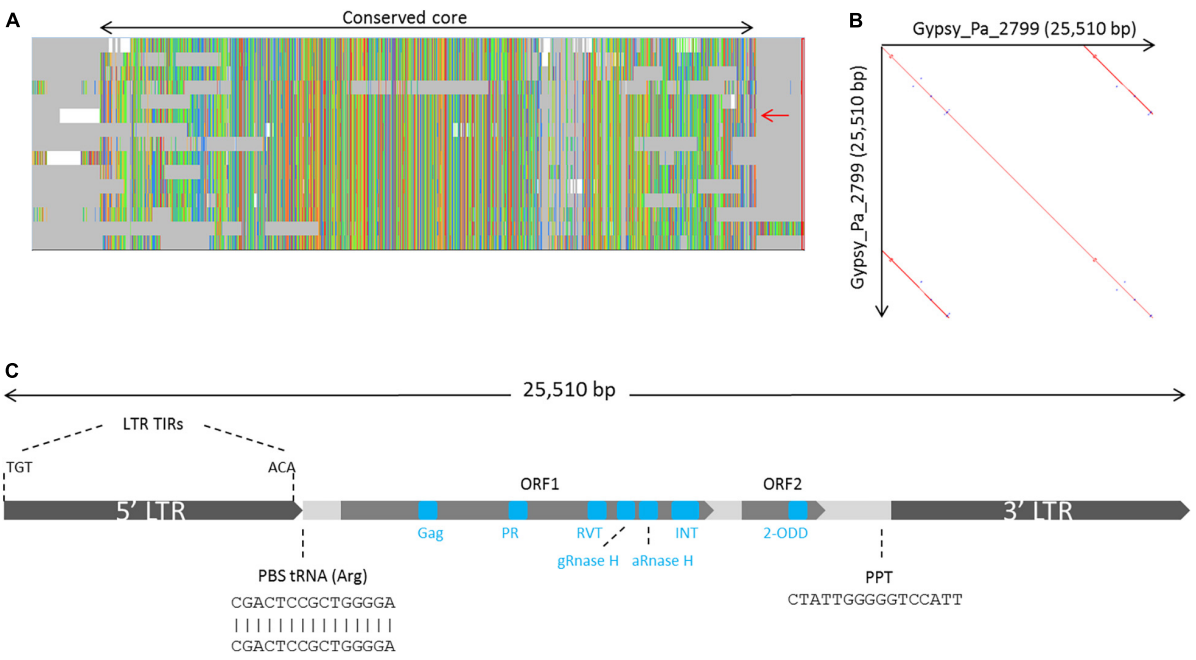


FIGURE 2 | Organization of the family of Gypsy-type elements with 2-ODD domain in *P. aphrodite*. **(A)** Overview of the multiple sequence alignment of the 15 *P. aphrodite* consensus sequences showing their conservation and the main variations between them. The whole alignment spans 33,269 bp. The black, two-headed arrow, spans over the fraction of the alignment which is well conserved across most consensus sequences (representing 28,280 bp). The red arrow indicates the selected reference consensus for this family (Gypsy_Pa_2799). **(B)** Dot plot of the Gypsy_Pa_2799 consensus sequence highlighting its LTRs. **(C)** Schematic representation depicting the structure and main features of the Gypsy_Pa_2799 consensus sequence including the LTRs and their terminal inverted repeats (LTR TIRs), the primer binding site (PBS), the putative polypurine tract (PPT) as well as the two different ORFs and the protein domains they contain (blue boxes, not to scale).

Gypsy-type elements, largely attributed to the size of LTRs, is typical of Ogre elements. Second, the primer binding site (PBS) is similar to that observed across Ogre elements, i.e., antisense to tRNA-Arginine. In addition, as reported across elements from the Tat family (Ustyantsev et al., 2015), the orchid elements also present a dual RNase H domain. However, important structural differences with Ogre elements can be observed as the Gag and Pol polyproteins are separated by a putative intron in the latter while they form a single ORF in Gypsy_Pa_2799 and as Ogre elements typically lack the additional ORF encoding the 2-ODD domain.

Phalaenopsis 2-ODDs Are Homologous to Plant RNA Demethylases

To address its putative function, we first compared the Gypsy-encoded ORF2 containing the 2-ODD domain to the NCBI protein database and noticed the highest similarity against plant proteins belonging to the alpha-ketoglutarate-dependent hydroxylase (AlkB) superfamily, which contains the 2-ODD fold. The TE-encoded 2-ODD domain was translated from consensus Gypsy_Pa_2799 and aligned with those from human and *Arabidopsis* reference proteins to determine its position in the AlkB superfamily using phylogenetic analysis. The resulting tree shows that the Gypsy-encoded AlkB domain is nested within *Arabidopsis* AlkBH5 homologs in a clade being sister to human AlkBH5 (**Figure 4A**). The 2-ODD-containing ORF2 from *Phalaenopsis aphrodite* Gypsy elements therefore represents a hypothetical AlkBH5 homolog and it is named paALKBH5-like hereafter. It is interesting to denote that, similar to hsALKBH5, some of the most closely related orthologs in *Arabidopsis*, atALKBH9B (AT2G17970) and atALKBH10B (AT4G02940), have been shown to bind RNA and to have m6A demethylase activity (Hu et al., 2019). We next addressed whether the

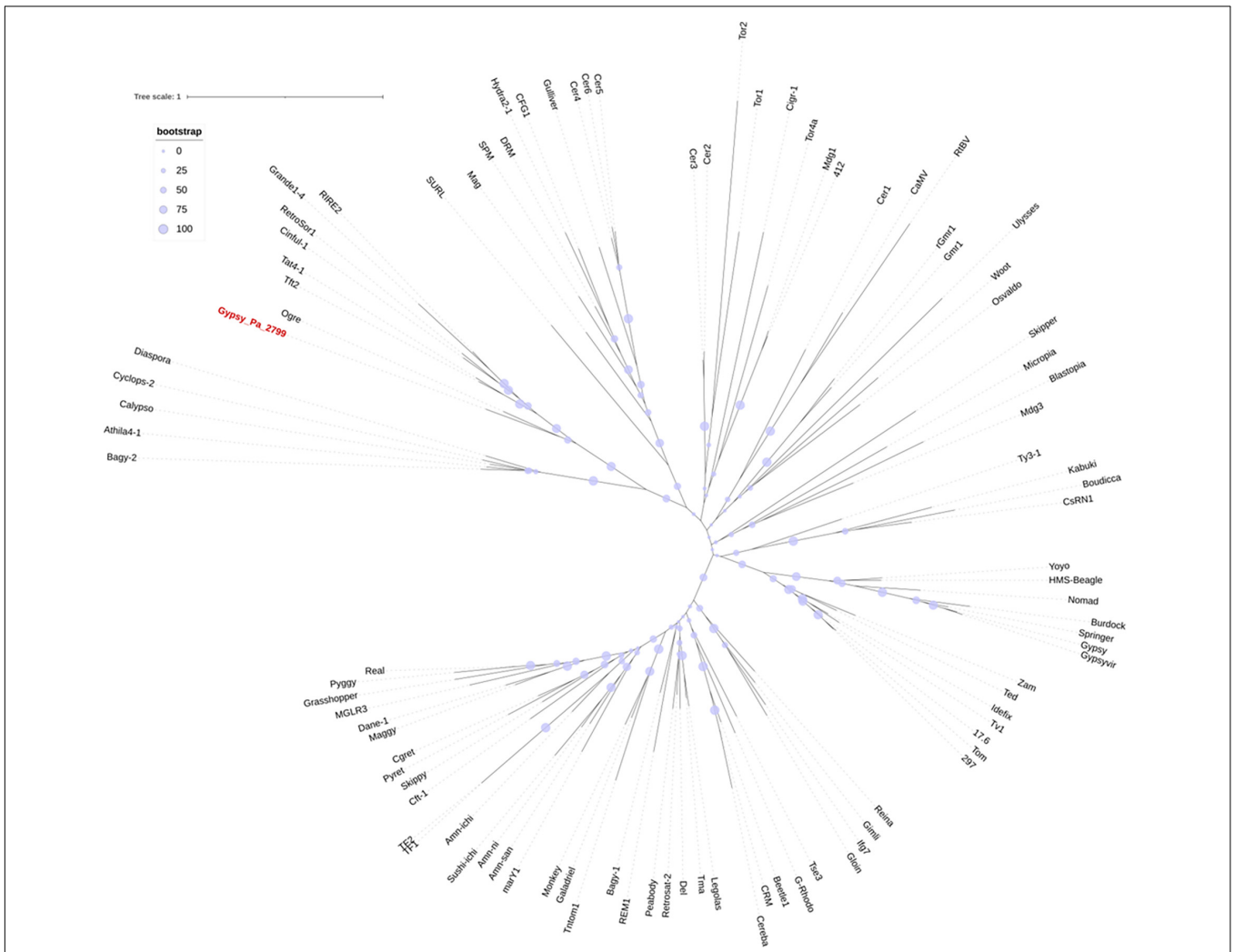


FIGURE 3 | Phylogenetic tree of the Gypsy-encoded reverse transcriptase domain. The tree shows the phylogenetic relationships between reverse transcriptase domains from Gypsy_Pa_2799 (labeled in red) and reference elements from the Gypsy database. The bootstrap values are indicated as circles of size proportional to intervals of values described in the top left panel.

sequences of TE-encoded proteins might conserve the amino acid residues necessary for this activity. The crystal structure of human AlkBH5 protein has revealed the residues binding the metal ion and 2-oxoglutarate (2-OG) cofactors and protein alignment indicated the strong conservation of these residues across the AlkB superfamily (Feng et al., 2014; Xu et al., 2014). Here, aligning the Gypsy-encoded 2-ODD domain with those of reference homologs revealed that all the cofactor-binding residues are conserved in the TE-encoded protein (**Figure 4B**). Thus, we observed that the known HxD/E···H and RxxxxxR motifs (indicated as red and purple arrows, respectively, in **Figure 4B**) involved in the binding to iron [Fe (II)] and 2-OG (Bratlie and Drablos, 2005; Lu et al., 2014) are present in the paALKBH5-like protein. As well, the N and Y residues involved in 2-OG stabilization are also conserved (indicated as yellow arrows in **Figure 4B**). Furthermore, the motif YNF potentially involved in the binding to m6A and an arginine residue essential for hsALKBH5 activity (Feng et al., 2014) are also conserved (residues are highlighted in fuchsia and green, respectively, on **Supplementary Figure 2**).

Gypsy-Encoded 2-ODD Domain Has m6A Demethylase Activity Against ssRNA

To assay the m6A demethylation activity of the *P. aphrodite* Gypsy-encoded 2-ODD domain, the paALKBH5-like protein was amplified and cloned in a bacterial expression plasmid fused to a Histidine tag at its C-terminal part. As positive control, *Arabidopsis* ALKBH9B protein was cloned and expressed in the same conditions. This protein was previously found to present m6A demethylation activity toward ssRNA (Martínez-Pérez et al., 2017). Thus, m6A monomethylated ssRNA were incubated with or without the recombinant proteins. Afterward, ssRNA was digested to single nucleoside and the resultant products were analyzed by ultra-performance liquid chromatography–photodiode detector–quadrupole/time-of-flight–mass spectrometry (UPLC-PDA-TOF-MS) as previously described (Martínez-Pérez et al., 2017). As shown in **Figure 5**, paALKBH5-like protein demethylated m6A on the ssRNA-oligonucleotide substrate by up to 60%, a percentage comparable to that observed for the positive control (ALKBH9B, 75%). Therefore, paALKBH5-like is a protein with ssRNA m6A demethylase activity "*in vitro*." Previously it was found that change of the histidine at position 204 by alanine (H204A) in the characteristic iron ligand domain of the AlkB family (H204xD206. . .H266; Lu et al., 2014) of human ALKBH5 completely abolished the demethylation activity of the protein, confirming that the iron center was implicated in this activity (Zheng et al., 2013). The equivalent H204 is located at position 448 (His448, Asp450, and His496) in paALKBH5-like. To check if the putative iron center in paALKBH5-like was required for its demethylation activity, a recombinant version of the protein having the mutation H448A was generated and biochemically tested. We found that the m6A levels were similar to those of the negative control indicating that the change of H448 by A drastically affects the demethylase capacity of paALKBH5-like (**Figure 5**).

Identification of a Putative RNA Binding Site

Recently, a biochemical analysis to delimit functional domains in *Arabidopsis* ALKBH9B showed two intrinsically disordered regions (IDRs) located at N- and C-terminal parts (Alvarado-Marchena et al., 2021). IDRs are unstructured regions implicated in protein–protein and RNA–protein interactions, which among others, present RGG boxes or R/K basic domains (repeats of arginine/glycine domains or arginine- or lysine-rich regions, respectively) [reviewed in Corley et al. (2020) and Ottoz and Berchowitz (2020)]. In ALKBH9B, an RNA binding domain containing an RGxxxRGG overlaps with the C-terminal IDR (Alvarado-Marchena et al., 2021). Both, C-terminal IDR and the RNA binding domain are absent in the TE-encoded protein (ORF2). However, we could detect an IDR in the N-terminal part of ORF2 and we noticed that it contains a basic domain consisting of a path rich with R and K residues (**Supplementary Figure 2**). Arginine residues have been found to interact with the phosphodiester backbone and in this sense RNA binding domain consisting of rich K and R patches have been identified in several RNA binding proteins (Ottoz and Berchowitz, 2020). Thus, is it tempting to speculate that this internal K/R rich domain would confer RNA binding capabilities to the Gypsy_Pa_2799 protein. To check this possibility we carried out Electrophoretic Mobility Shift Assays (EMSA) by incubating a constant amount (5 ng) of an RNA transcript, corresponding to the 5′ untranslated region (UTR) of AMV RNA3, with increasing concentrations of the paALKBH5-like (Gypsy_Pa_2799) protein (**Figure 6A**). The decrease in the chemiluminescent signal intensity corresponding to free RNA was evident at quantities exceeding 125 ng of the protein. The apparent constant dissociation (Kd) of the RNA–paALKBH5-like interaction from the linear regression of the mean values from at least three technical replicates (Marcos et al., 1999) was estimated to be 0.21 μM (**Figure 6B**). To determine if the mutant lacking demethylase activity retained its RNA-binding capacity, Northwestern experiments were performed with both proteins. As can be seen in **Figure 6C**, the His:paALKBH5-like H448A mutant in which its demethylase activity was abolished maintained its RNA binding capacity, indicating that this capacity is necessary but not sufficient for its function.

DISCUSSION

In this study, we screened the content of conserved protein domains in LTR retrotransposon consensus sequences generated from a selection of plant genome assemblies. Admittedly, the set of domains found is biased by the approach used to build consensus sequences, which is empirically not adapted to the analysis of low-copy number TE families and which efficiency decreases with low contiguity genome assemblies. This analysis is also biased by the selection of the species investigated here. Nevertheless, we generated thousands of consensus sequences representative of Copia-type and Gypsy-type LTR retrotransposons and the analysis of their protein domain content allowed to successfully retrieve several domains that were previously reported in plant Gypsy-type elements such

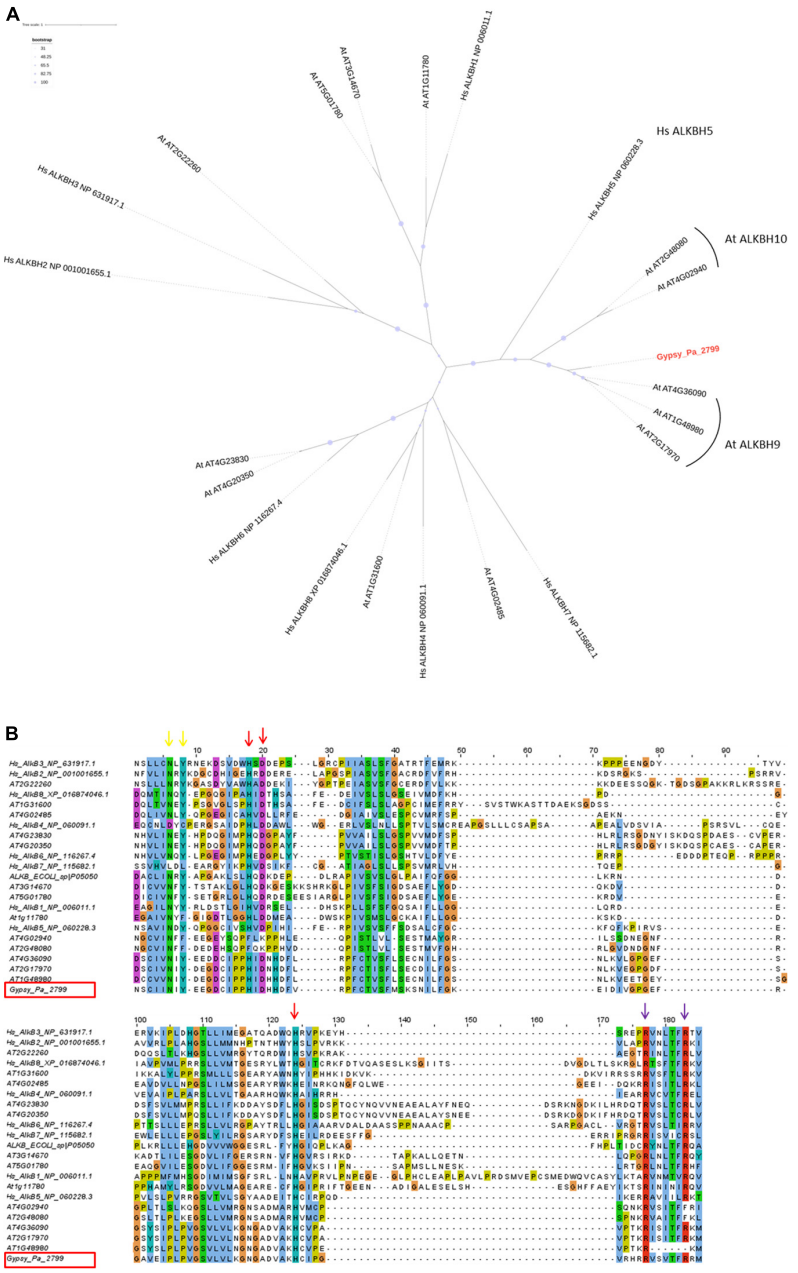


FIGURE 4 | Characterization of the Gypsy-encoded 2-ODD domain. **(A)** Phylogenetic tree inferred from the alignment of Gypsy_Pa_2799 (labeled in red) 2-ODD domain with AlkB homologs from *E. coli*, *A. thaliana*, and *H. sapiens*. The bootstrap values are indicated as circles of size proportional with intervals of values described in the top left panel. **(B)** Conservation of co-factor binding residues in the 2-ODD domain across the alignment used in **(A)**. The alignment view is truncated to positions corresponding to residues 188–285 in human ALKBH5. The columns corresponding to the metal ion binding residues (ALKBH5 His-204, Asp-206, and His-266) are indicated by red arrows. The columns corresponding to the 2-OG binding residues (ALKBH5 Arg-277, and Arg-283) are indicated by purple arrows and the residues involved in 2-OG stabilization (ALKBH5 Asn-193, Tyr-195) are indicated by yellow arrows. The Gypsy_Pa_2799 sequence name is highlighted by a red box.

as CHROMO, PMD, and AIR1. In line with the literature, no domain gain was observed in plant Copia-type elements by contrast to plant Gypsy elements which are more prone to domain acquisition and/or retention.

We noticed that the most frequent protein domains described have functional categories relevant to interactions with RNA and DNA including AlkB, G-patch, KOW, ArgoL2, B3, and potentially AIR1 domain, which overlaps UMSBP DNA binding domain. Given that TEs have compact genomes with limited coding capacity, the retention of these domains through evolution and natural selection suggests that they bring important functions to the cognate TE families that result in

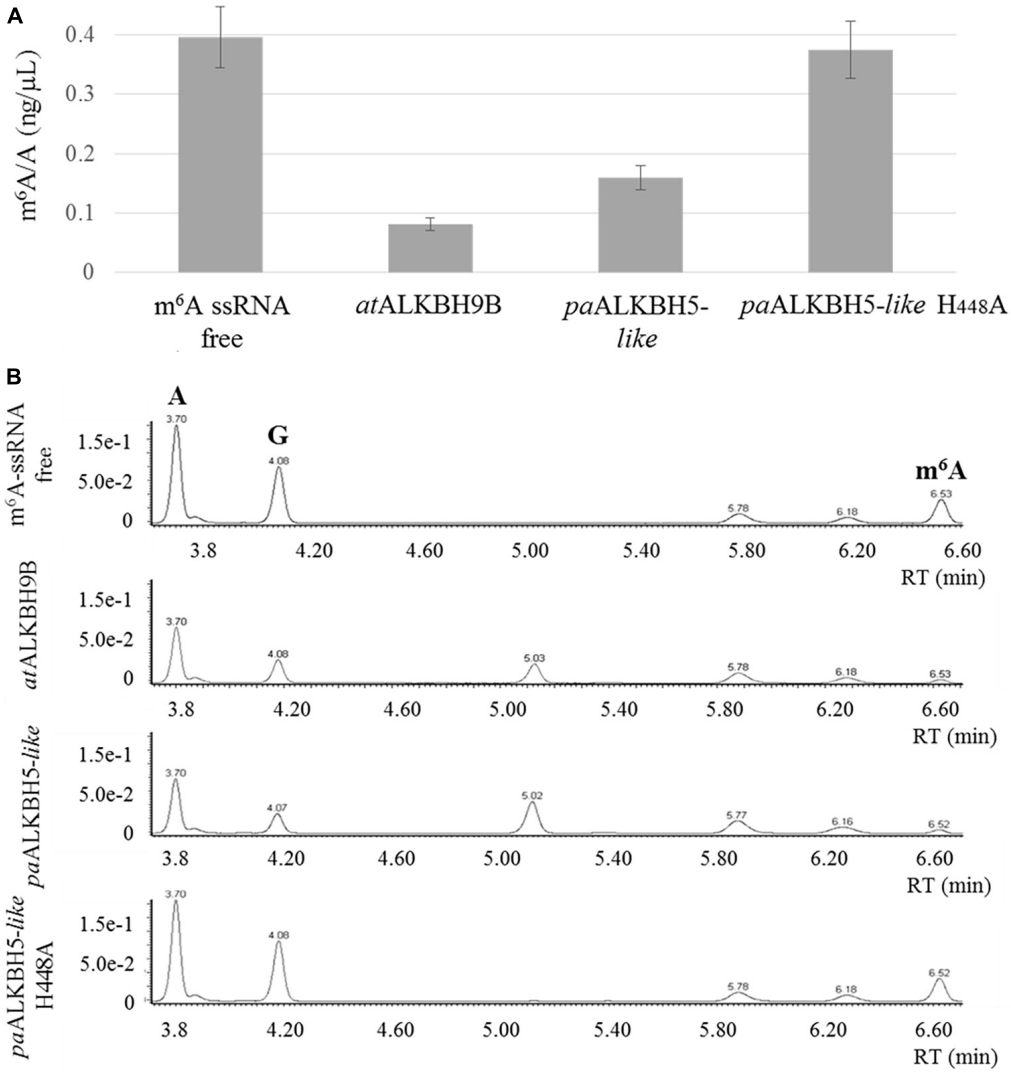


FIGURE 5 | *In vitro* demethylation assay using ssRNA m^{6}A-monomethylated. **(A)** Graphic showing the average m^{6}A/A ratios obtained by UPLC-Q-TOF-MS after enzymatic digestion of the ssRNA m^{6}A-monomethylated. The demethylation activity was evaluated in two independent experiments. **(B)** Representative UPLC-PDA-Q/TOF-MS chromatogram showing the retention times of the nucleosides adenosine "A" and N^6-methyladenosine "m^{6}A" after incubation of the m^{6}A-containing ssRNA with non-protein (m^{6}A-ssRNA free), *at*ALKBH9B, *pa*ALKBH5-*like* and *pa*ALKBH5-like H448A. The peak "G" corresponds to the nucleoside guanosine.

a positive impact on TE fitness (with an acceptable impact on host fitness). Consequently, such a functional bias suggests that these fixed domains may have been selected out of a cryptic flow of domain acquisition events, the majority of which being rejected throughout evolution. Remarkably, several of these domains have been acquired independently in different mobile genetic elements. For instance, PMD domain is also found in MULE DNA transposons (Babu et al., 2006). The G-patch has also been detected in type D retroviruses (Aravind and Koonin, 1999). Furthermore, AlkB domain was reported in several families of RNA viruses, including Alphaflexiviridae, Betaflexiviridae, and Closteroviridae (Aravind and Koonin, 2001; Moore and Meng, 2019). The virus-encoded proteins display RNA demethylation activity *in vitro* and it has been postulated that its RNA repair function could help to maintain genome integrity (van den Born et al., 2008).

Here, we described that a specific family of Phalaenopsis Gypsy-type elements has acquired an additional ORF and retained it over several rounds of replication as attested by the diversity of consensus sequences reconstructed for this family (**Figure 2A**) which probably reflects speciation into different subfamilies after the ORF was captured. This ORF contains a 2-ODD domain, conserves the cofactor-binding residues necessary for its activity (**Figure 4B**), is most closely related to known RNA demethylases from human and *Arabidopsis* (**Figure 4A**) and, most importantly, the purified protein shows RNA demethylase activity *in vitro* (**Figure 5**) and capability to bind RNA (**Figure 6**). Interestingly, the apparent constant dissociation (Kd) calculated

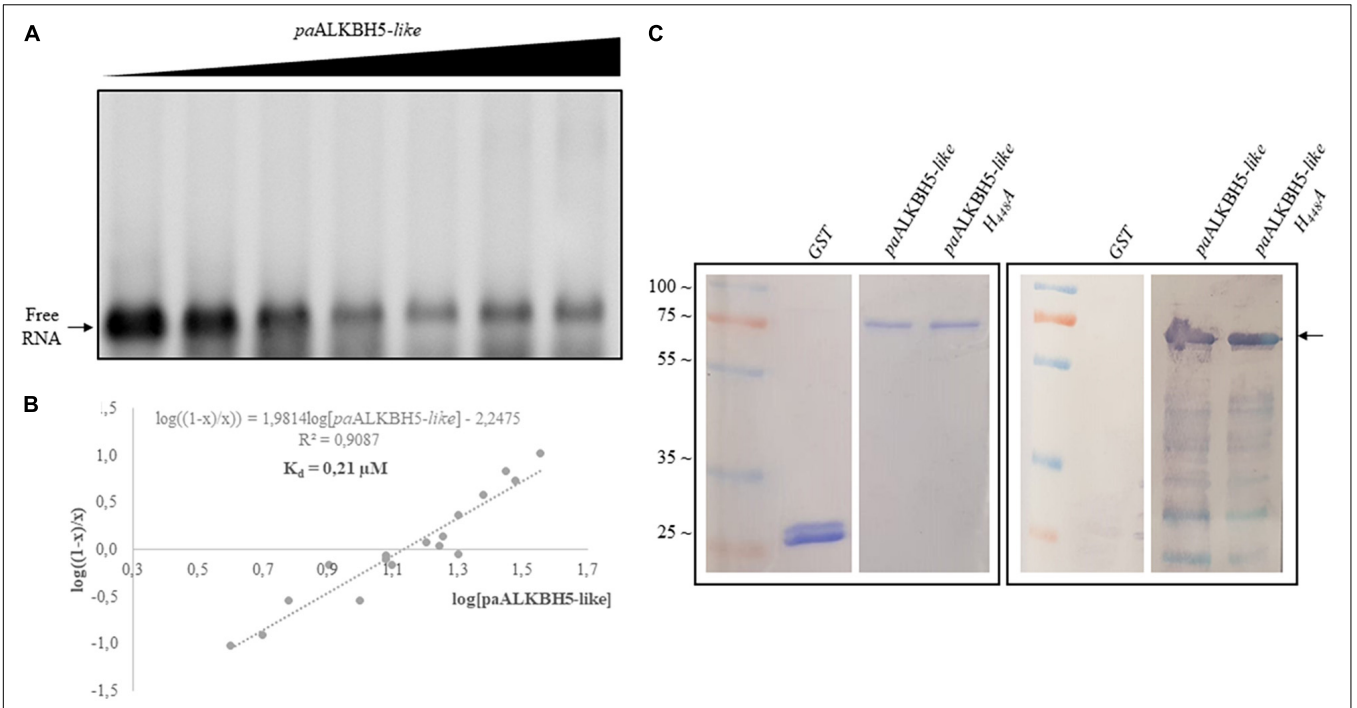


FIGURE 6 | Analysis of the RNA binding capacity of paALKBH5-like. **(A)** EMSA after incubation of 5 ng of AMV 5′UTR transcript with no protein (lane 1) or with 5, 50, 100, 125, 175, 200, and 250 ng of His:paALKBH5-like protein. **(B)** Representation of Hill transformation of RNA-His:paALKBH5-like obtained from three independent experiments. R2 coefficient, regression equation and the dissociation constant (Kd) are shown. **(C)** Northwestern blot analysis result of the incubation of 1.5 µg GST, His:paALKBH5-like or its mutant variant H448A (indicated on the top) with 50 ng of AMV RNA 4 labeled with digoxigenin. The left panel shows the Coomassie blue-stained gel, and the right panel shows the nitrocellulose membrane after incubation. Positions of full-length His:paALKBH5-like proteins are indicated with an arrow. Ladder molecular mass is indicated on the left in kiloDaltons (KDa). Lines of the markers and GST were not contiguous to those of His:paALKBH5-like or its mutant variant H448A.

for the paALKBH5-like protein (0.21 µM, **Figure 6**) was similar to that observed for atALKBH9B (0.30 µM) (Alvarado-Marchena et al., 2021) reinforcing the notion that both proteins are functionally equivalent.

Following the "no superflue" postulate regarding the genomes of functional and autonomous TE sequences, one can suppose that its demethylase activity could bring selective advantage to the TE family itself, e.g., by increasing its fitness. After transcription, the LTR retrotransposon mRNA is exported to the cytoplasm where it is translated (Sabot and Schulman, 2006). We can speculate that, upon translation, the RNA demethylase protein (ORF2), which is expressed independently of the Gag-Pol polyprotein (ORF1), could bind and be enzymatically active against the cognate TE mRNA or any cellular RNA. To bring any advantage to the TE, this could suppose either that the TE mRNA is subject to m6A methylation by the host and that it is deleterious to the TE replication cycle or/and that the demethylation of other cellular RNA by TE-encoded AlkB protein impacts the stability and translation of genes involved in different layers of TE control, as recently proposed for some AlkB-encoding plant RNA viruses (Yue et al., 2022).

m6A methylated mRNAs recruit specific reader proteins and, in mammals, it has been found that a polymethylated mRNA recruits several reader molecules causing the juxtaposition of their intrinsically disordered regions (IDRs) that activates the phenomenon of liquid-liquid phase separation. Thus, the polymethylated mRNA-eraser protein complexes accumulate into different phase-separated membraneless compartments such as stress granules or P- bodies that affect either the stability or translation of these mRNAs (Ries et al., 2019). Plant m6A readers also contain IDRs and some of them have been found to form dense assemblages *in vitro* and accumulate in granules upon stress (Arribas-Hernandez et al., 2018). Thus, the abundance of m6A sites in TE-derived mRNAs might influence their expression and retrotransposition activities.

Most recently, it has been reported in human cells that RNA transcripts of young LINE-1 retrotransposons present elevated levels of m6A modification that acts promoting their expression and retrotransposition (Xiong et al., 2021). In *Arabidopsis*, most TE transcripts exhibit a relatively high extent of m^6A modification compared to gene transcript level (Wan et al., 2015). However, this has been observed under standard conditions in which mostly fragmented forms of TE transcripts can be detected. It would be interesting to measure the level of m6A methylation in TE mRNA in the context of their replicative cycles and to address its impact on the regulation of the transposition rates. Remarkably, a recent report deposited in bioRxiv shows that the RNA of the heat-activated retroelement ONSEN contains m6A,

which suppresses the transposon mobility via its sequestration in stress granules in an m6A-dependent manner. Furthermore, the RNA demethylase AtALKBH9B directly demethylates ONSEN m6A-modified RNA located in stress granules allowing its mobilization. In fact, it has been proposed that in plants, some RNA demethylases would have diversified to preferentially target genetic elements such as retrotransposons (Fan et al., 2022).

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/ Figshare: https://figshare.com/projects/Orchid_Ogre_with_RNA_demethylase_-_Supplementary_dataset/137215.

AUTHOR CONTRIBUTIONS

LA-M and MM-P performed the experiments. VP supervised the experiments. FM performed the bioinformatic analyses and conceived the study. FM, FA, MM-P, and VP interpreted the results and wrote the manuscript. All authors analyzed and discussed the results.

FUNDING

This research was funded by grant PID2020-115571RB-I00 to FA and VP from the Spanish MCIN/AEI/10.13039/501100011033 granting agency and by financial support from INRAE Plant Biology and Breeding Division to FM. LA-M was recipient of a Predoctoral contract from the Ministerio de Ciencia, Innovación, Tecnología y Telecomunicaciones from Costa Rica (MI-CITT-PINN-CON-624-2019). UPLC-PDAQ/TOF-MS analyses were performed by the Metabolic Analysis Department of the IBMCP.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://figshare.com/projects/Orchid_Ogre_with_RNA_demethylase_-_Supplementary_dataset/137215

REFERENCES

Alvarado-Marchena, L., Marquez-Molins, J., Martinez-Perez, M., Aparicio, F., and Pallas, V. (2021). Mapping of functional subdomains in the atALKBH9B m(6)A-demethylase required for its binding to the viral RNA and to the Coat Protein of Alfalfa Mosaic Virus. *Front. Plant Sci.* 12:701683. doi: 10.3389/fpls.2021.701683

Anderson, S. J., Kramer, M. C., Gosai, S. J., Yu, X., Vandivier, L. E., Nelson, A. D. L., et al. (2018). N(6)-Methyladenosine inhibits local ribonucleolytic cleavage to stabilize mRNAs in *Arabidopsis*. *Cell Rep.* 25, 1146–1157 e3. doi: 10.1016/j.celrep.2018.10.020

Aravind, L., and Koonin, E. V. (1999). G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins. *Trends Biochem. Sci.* 24, 342–344.

Aravind, L., and Koonin, E. V. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* 2:RESEARCH0007. doi: 10.1186/gb-2001-2-3-research0007

Arkhipova, I. R., and Yushenova, I. A. (2019). Giant transposons in eukaryotes: is bigger better? *Genome Biol. Evol.* 11, 906–918. doi: 10.1093/gbe/evz041

Arribas-Hernandez, L., Bressendorff, S., Hansen, M. H., Poulsen, C., Erdmann, S., and Brodersen, P. (2018). An m(6)A-YTH module controls developmental timing and Morphogenesis in *Arabidopsis*. *Plant Cell* 30, 952–967. doi: 10.1105/tpc.17.00833

Arribas-Hernandez, L., and Brodersen, P. (2020). Occurrence and functions of m(6)A and Other covalent modifications in plant mRNA. *Plant Physiol.* 182, 79–96. doi: 10.1104/pp.19.01156

Babu, M. M., Iyer, L. M., Balaji, S., and Aravind, L. (2006). The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* 34, 6505–6520. doi: 10.1093/nar/gkl888

Bratlie, M. S., and Drablos, F. (2005). Bioinformatic mapping of AlkB homology domains in viruses. *BMC Genomics* 6:1. doi: 10.1186/1471-2164-6-1

Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741.

Chelmicki, T., Roger, E., Teissandier, A., Dura, M., Bonneville, L., Rucli, S., et al. (2021). m(6)A RNA methylation regulates the fate of endogenous retroviruses. *Nature* 591, 312–316. doi: 10.1038/s41586-020-03135-1

Corley, M., Burns, M. C., and Yeo, G. W. (2020). How RNA-binding proteins interact with RNA: molecules and mechanisms. *Mol. Cell* 78, 9–29.

Duan, H. C., Wei, L. H., Zhang, C., Wang, Y., Chen, L., Lu, Z., et al. (2017). ALKBH10B Is an RNA N(6)-methyladenosine demethylase affecting *Arabidopsis* floral transition. *Plant Cell* 29, 2995–3011. doi: 10.1105/tpc.16.00912

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18. doi: 10.1186/1471-2105-9-18

Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J., and Joshua-Tor, L. (2013). The making of a slicer: activation of human Argonaute-1. *Cell Rep.* 3, 1901–1909. doi: 10.1016/j.celrep.2013.05.033

Fan, W., Wang, L., Lei, Z., Chu, J., and Cho, J. (2022). Suppression of transposon mobilization by m6A-mediated RNA sequestration in stress granules. *bioRxiv* [Preprint] doi: 10.1101/2022.03.22.485398

Farrow, S. C., and Facchini, P. J. (2014). Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front. Plant Sci.* 5:524. doi: 10.3389/fpls.2014.00524

Feng, C., Liu, Y., Wang, G., Deng, Z., Zhang, Q., Wu, W., et al. (2014). Crystal structures of the human RNA demethylase Alkbh5 reveal basis for substrate recognition. *J. Biol. Chem.* 289, 11571–11583. doi: 10.1074/jbc.M113.546168

Gao, X., Hou, Y., Ebina, H., Levin, H. L., and Voytas, D. F. (2008). Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* 18, 359–369. doi: 10.1101/gr.7146408

Gonzales-van Horn, S. R., and Sarnow, P. (2017). Making the mark: the role of adenosine modifications in the life cycle of RNA viruses. *Cell Host Microbe* 21, 661–669. doi: 10.1016/j.chom.2017.05.008

Gorinsek, B., Gubensek, F., and Kordis, D. (2004). Evolutionary genomics of chromoviruses in eukaryotes. *Mol. Biol. Evol.* 21, 781–798.

He, Y., Li, Y., Yao, Y., Zhang, H., Wang, Y., Gao, J., et al. (2021). Overexpression of watermelon m(6)A methyltransferase ClMTB enhances drought tolerance in tobacco by mitigating oxidative stress and photosynthesis inhibition and modulating stress-responsive gene expression. *Plant Physiol. Biochem.* 168, 340–352. doi: 10.1016/j.plaphy.2021.10.007

Herr, C. Q., and Hausinger, R. P. (2018). Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases. *Trends Biochem. Sci.* 43, 517–532. doi: 10.1016/j.tibs.2018.04.002

Hu, J., Cai, J., Park, S. J., Lee, K., Li, Y., Chen, Y., et al. (2021a). N(6)-Methyladenosine mRNA methylation is important for salt stress tolerance in *Arabidopsis*. *Plant J.* 106, 1759–1775. doi: 10.1111/tpj.15270

Hu, J., Cai, J., Umme, A., Chen, Y., Xu, T., and Kang, H. (2021b). Unique features of mRNA m6A methylomes during expansion of tomato (*Solanum lycopersicum*) fruits. *Plant Physiol.* 3:kiab509. doi: 10.1093/plphys/kiab509

Hu, J., Manduzio, S., and Kang, H. (2019). Epitranscriptomic RNA methylation in plant development and abiotic stress responses. *Front. Plant Sci.* 10:500. doi: 10.3389/fpls.2019.00500

Ikeda, Y., Pelissier, T., Bourguet, P., Becker, C., Pouch-Pelissier, M. N., Pogorelcnik, R., et al. (2017). *Arabidopsis* proteins with a transposon-related domain act in gene silencing. *Nat. Commun.* 8:15122. doi: 10.1038/ncomms15122

Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343.

Kazazian, H. H. Jr. (2004). Mobile elements: drivers of genome evolution. *Science* 303, 1626–1632.

Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., et al. (2009). Functional analyses of LONELY GUY cytokinin-activating enzymes reveal the importance of the direct activation pathway in *Arabidopsis*. *Plant Cell* 21, 3152–3169. doi: 10.1105/tpc.109.068676

Kurowski, M. A., Bhagwat, A. S., Papaj, G., and Bujnicki, J. M. (2003). Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* 4:48. doi: 10.1186/1471-2164-4-48

Kyrpides, N. C., Woese, C. R., and Ouzounis, C. A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* 21, 425–426. doi: 10.1016/s0968-0004(96)30036-4

Laten, H. M., Majumdar, A., and Gaucher, E. A. (1998). SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6897–6902. doi: 10.1073/pnas.95.12.6897

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518, 560–564. doi: 10.1038/nature14234

Llorens, C., Futami, R., Covelli, L., Dominguez-Escriba, L., Viu, J. M., Tamarit, D., et al. (2011). The gypsy database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* 39, D70–D74.

Llorens, C., Munoz-Pomer, A., Bernad, L., Botella, H., and Moya, A. (2009). Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol. Direct* 4:41. doi: 10.1186/1745-6150-4-41

Lu, L., Zhu, C., Xia, B., and Yi, C. (2014). Oxidative demethylation of DNA and RNA mediated by non-heme iron-dependent dioxygenases. *Chem. Asian J.* 9, 2018–2029. doi: 10.1002/asia.201402148

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2010). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229.

Marcos, J. F., Vilar, M., Perez-Paya, E., and Pallas, V. (1999). In vivo detection, RNA-binding properties and characterization of the RNA-binding domain of the p7 putative movement protein from carnation mottle carmovirus (CarMV). *Virology* 255, 354–365. doi: 10.1006/viro.1998.9596

Marin, I., and Llorens, C. (2000). Ty3/Gypsy retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. *Mol. Biol. Evol.* 17, 1040–1049. doi: 10.1093/oxfordjournals.molbev.a026385

Martínez-Pérez, M., Aparicio, F., Lopez-Gresa, M. P., Belles, J. M., Sanchez-Navarro, J. A., and Pallas, V. (2017). Arabidopsis m(6)A demethylase activity modulates viral infection of a plant virus and the m(6)A abundance in its genomic RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10755–10760. doi: 10.1073/pnas.1703139114

Martínez-Pérez, M., Gomez-Mena, C., Alvarado-Marchena, L., Nadi, R., Micol, J. L., Pallas, V., et al. (2021). The m(6)A RNA demethylase ALKBH9B plays a critical role for vascular movement of alfalfa mosaic virus in *Arabidopsis*. *Front. Microbiol.* 12:745576. doi: 10.3389/fmicb.2021.745576

Matzke, M. A., Kanno, T., and Matzke, A. J. (2015). RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annu. Rev. Plant Biol.* 66, 243–267.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534.

Moore, C., and Meng, B. (2019). Prediction of the molecular boundary and functionality of novel viral AlkB domains using homology modelling and principal component analysis. *J. Gen. Virol.* 100, 691–703. doi: 10.1099/jgv.0.001237

Nayar, S. (2020). Exploring the role of a cytokinin-activating enzyme LONELY GUY in unicellular microalga *Chlorella* variabilis. *Front. Plant. Sci.* 11:611871. doi: 10.3389/fpls.2020.611871

Neumann, P., Novák, P., Hoštáková, N., and Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA* 10:1. doi: 10.1186/s13100-018-0144-1

Neumann, P., Pozarkova, D., and Macas, J. (2003). Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. *Plant Mol. Biol.* 53, 399–410. doi: 10.1023/b:plan.0000006945.77043.ce

Nicolau, M., Picault, N., Descombin, J., Jami-Alahmadi, Y., Feng, S., Bucher, E., et al. (2020). The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*. *PLoS Genet.* 16:e1008324. doi: 10.1371/journal.pgen.1008324

Ottoz, D. S. M., and Berchowitz, L. E. (2020). The role of disorder in RNA binding affinity and specificity. *Open Biol.* 10:200328.

Pallas, V., Sanchez-Navarro, J. A., and Diez, J. (1999). In vitro evidence for RNA binding properties of the coat protein of prunus necrotic ringspot ilarvirus and their comparison to related and unrelated viruses. *Arch. Virol.* 144, 797–803. doi: 10.1007/s007050050545

Ries, R. J., Zaccara, S., Klein, P., Olarerin-George, A., Namkoong, S., Pickering, B. F., et al. (2019). m(6)A enhances the phase separation potential of mRNA. *Nature* 571, 424–428.

Ruzicka, K., Zhang, M., Campilho, A., Bodi, Z., Kashif, M., Saleh, M., et al. (2017). Identification of factors required for m(6) A mRNA methylation in *Arabidopsis* reveals a role for the conserved E3 ubiquitin ligase HAKAI. *New Phytol.* 215, 157–172. doi: 10.1111/nph.14586

Sabot, F., and Schulman, A. H. (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity (Edinb)* 97, 381–388. doi: 10.1038/sj.hdy.6800903

Saint-Leandre, B., Nguyen, S. C., and Levine, M. T. (2019). Diversification and collapse of a telomere elongation mechanism. *Genome Res.* 29, 920–931. doi: 10.1101/gr.245001.118

Shen, L., Liang, Z., Gu, X., Chen, Y., Teo, Z. W., Hou, X., et al. (2016). N(6)-methyladenosine RNA modification regulates shoot stem cell fate in *Arabidopsis*. *Dev. Cell* 38, 186–200. doi: 10.1016/j.devcel.2016.06.008

Steinbauerova, V., Neumann, P., Novak, P., and Macas, J. (2011). A widespread occurrence of extra open reading frames in plant Ty3/gypsy retrotransposons. *Genetica* 139, 1543–1555. doi: 10.1007/s10709-012-9654-9

Steinegger, M., and Soding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* 35, 1026–1028. doi: 10.1038/nbt.3988

Ustyantsev, K., Novikova, O., Blinov, A., and Smyshlyaev, G. (2015). Convergent evolution of ribonuclease h in LTR retrotransposons and retroviruses. *Mol. Biol. Evol.* 32, 1197–1207. doi: 10.1093/molbev/msv008

van den Born, E., Omelchenko, M. V., Bekkelund, A., Leihne, V., Koonin, E. V., Dolja, V. V., et al. (2008). Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acids Res.* 36, 5451–5461.

Vicient, C. M., and Casacuberta, J. M. (2020). Additional ORFs in Plant LTR-retrotransposons. *Front. Plant Sci.* 11:555. doi: 10.3389/fpls.2020.00555

Wan, Y., Tang, K., Zhang, D., Xie, S., Zhu, X., Wang, Z., et al. (2015). Transcriptome-wide high-throughput deep m(6)A-seq reveals unique differential m(6)A methylation patterns between three organs in *Arabidopsis thaliana*. *Genome Biol.* 16:272. doi: 10.1186/s13059-015-0839-2

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505, 117–120.

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N(6)-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161, 1388–1399.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982.

Williams, G. D., Gokhale, N. S., and Horner, S. M. (2019). Regulation of viral infection by the RNA modification N6-methyladenosine. *Annu. Rev. Virol.* 6, 235–253.

Wright, D. A., and Voytas, D. F. (1998). Potential retroviruses in plants: tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. *Genetics* 149, 703–715. doi: 10.1093/genetics/149.2.703

Xiong, F., Wang, R., Lee, J. H., Li, S., Chen, S. F., Liao, Z., et al. (2021). RNA m(6)A modification orchestrates a LINE-1-host interaction that facilitates retrotransposition and contributes to long gene vulnerability. *Cell Res.* 31, 861–885. doi: 10.1038/s41422-021-00515-8

Xu, C., Liu, K., Tempel, W., Demetriades, M., Aik, W., Schofield, C. J., et al. (2014). Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single-stranded N6-methyladenosine RNA demethylation. *J. Biol. Chem.* 289, 17299–17311. doi: 10.1074/jbc.M114.550350

Yu, Q., Liu, S., Yu, L., Xiao, Y., Zhang, S., Wang, X., et al. (2021). RNA demethylation increases the yield and biomass of rice and potato plants in field trials. *Nat. Biotechnol.* 39, 1581–1588. doi: 10.1038/s41587-021-00982-9

Yue, H., Nie, X., Yan, Z., and Weining, S. (2019). N6-methyladenosine regulatory machinery in plants: composition, function and evolution. *Plant Biotechnol. J.* 17, 1194–1208. doi: 10.1111/pbi.13149

Yue, J., Wei, Y., and Zhao, M. (2022). The reversible methylation of m6A is involved in plant virus infection. *Biology (Basel)* 11:271.

Zaccara, S., and Jaffrey, S. R. (2020). A unified model for the function of YTHDF proteins in regulating m(6)A-modified mRNA. *Cell* 181:e1518. doi: 10.1016/j.cell.2020.05.012

Zheng, H., Sun, X., Li, J., Song, Y., Song, J., Wang, F., et al. (2021). Analysis of N(6)-methyladenosine reveals a new important mechanism regulating the salt tolerance of sweet sorghum. *Plant Sci.* 304:110801. doi: 10.1016/j.plantsci.2020.110801

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C. M., Li, C. J., et al. (2013). ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* 49, 18–29. doi: 10.1016/j.molcel.2012.10.015

Zhong, S., Li, H., Bodi, Z., Button, J., Vespa, L., Herzog, M., et al. (2008). MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* 20, 1278–1288. doi: 10.1105/tpc.108.058883

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.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 July 2022
DOI 10.3389/fpls.2022.969010

OPEN ACCESS

EDITED BY
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

REVIEWED BY
Yavar Vafaee,
University of Kurdistan,
Iran
Parviz Heidari,
Shahrood University of Technology, Iran

*CORRESPONDENCE
Gen-Fa Zhu
genfazhu@163.com
Feng-Xi Yang
fengxi_wei@sina.com

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

RECEIVED 14 June 2022
ACCEPTED 07 July 2022
PUBLISHED 28 July 2022

CITATION
Wei Y-L, Jin J-P, Liang D, Gao J, Li J, Xie Q, Lu C-Q, Yang F-X and Zhu G-F (2022) Genome-wide identification of *Cymbidium sinense* WRKY gene family and the importance of its Group III members in response to abiotic stress. *Front. Plant Sci.* 13:969010. doi: 10.3389/fpls.2022.969010

Genome-wide identification of *Cymbidium sinense* WRKY gene family and the importance of its Group III members in response to abiotic stress

Yong-Lu Wei, Jian-Peng Jin, Di Liang, Jie Gao, Jie Li, Qi Xie, Chu-Qiao Lu, Feng-Xi Yang* and Gen-Fa Zhu*

Guangdong Key Laboratory of Ornamental Plant Germplasm Innovation and Utilization, Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China

Transcription factors (TFs) of the WRKY family play pivotal roles in defense responses and secondary metabolism of plants. Although WRKY TFs are well documented in numerous plant species, no study has performed a genome-wide investigation of the WRKY gene family in *Cymbidium sinense*. In the present work, we found 64 *C. sinense* WRKY (CsWRKY) TFs, and they were further divided into eight subgroups. Chromosomal distribution of *CsWRKYs* revealed that the majority of these genes were localized on 16 chromosomes, especially on Chromosome 2. Syntenic analysis implied that 13 (20.31%) genes were derived from segmental duplication events, and 17 orthologous gene pairs were identified between *Arabidopsis thaliana WRKY* (*AtWRKY*) and *CsWRKY* genes. Moreover, 55 of the 64 *CsWRKYs* were detectable in different plant tissues in response to exposure to plant hormones. Among them, Group III members were strongly induced in response to various hormone treatments, indicating their potential essential roles in hormone signaling. We subsequently analyzed the function of *CsWRKY18* in Group III. The *CsWRKY18* was localized in the nucleus. The constitutive expression of *CsWRKY18* in Arabidopsis led to enhanced sensitivity to ABA-mediated seed germination and root growth and elevated plant tolerance to abiotic stress within the ABA-dependent pathway. Overall, our study represented the first genome-wide characterization and functional analysis of WRKY TFs in *C. sinense*, which could provide useful clues about the evolution and functional description of *CsWRKY* genes.

KEYWORDS

***Cymbidium sinense*, genome-wide, WRKY, expression pattern, abiotic stress**

Introduction

As one of the most prominent flowering plant families, Orchidaceae include 801 genera and more than 30,000 specie (Brown, 2005). In the orchid family, Cymbidium is widely fostered in East Asia, such as China, Japan, and Korea, and Southeast Asia (Liu et al., 2006). Cymbidium has become very prevalent in China, and it is highly desirable in traditional

flower markets because of its beauty, fragrant flowers, elegant and upright leaves. However, functional genomics investigations and gene identification of valuable horticultural traits are significantly restricted because of the polyploid genomes and long juvenile phases of the genus Cymbidium. Therefore, it is essential for discovery of functional genes in Cymbidium research. In recent years, next-generation sequencing (NGS) technologies have rapidly developed, offering formidable approaches for high-throughput sequence determination (Leitch et al., 2009; Liu et al., 2012). Furthermore, as RNA-Seq decrease in cost, promoting the identification of new genes by obtaining massive amounts of sequence data with enormous depth and coverage. As a result, a large number of critical regulators related to important agronomic traits and environmental adaptation have been identified in orchid species, including Apostasia (Zhang et al., 2017, 2021a), Cymbidium (Ai et al., 2021; Yang et al., 2021), Dendrobium (Niu et al., 2021; Zhang et al., 2021b), Gastrodia (Xu et al., 2021) Phalaenopsis (Cai et al., 2015; Chao et al., 2018), and Vanilla (Hasing et al., 2020).

The transcription factors (TFs) are critical in controlling plant development and stress response. WRKY, one of the largest families of higher plant TFs, plays essential roles in pathogen defense, abiotic cues, phytohormone signaling, and management of plant development and secondary metabolism. The WRKY domain (WD) is the essential characteristic of the WRKY protein, which is ~60 amino acids in length and consists of the highly conserved signature WRKYGQK, followed by a C2H2-or C2HC-type of zinc-finger motif. The high binding affinity of WRKY TFs to the consensus W-box cis-elements requires both heptapeptide sequence and zinc-finger motif (Maeo et al., 2001). Given these features of the WD, its family members are categorized into three groups (Eulgem et al., 2000). The WRKY Group II can be further divided into five subgroups (IIa, IIb, IIc, IId, and IIe) according to their primary amino acid sequences (Chen et al., 2017). In addition, WRKY preferentially binds to a markedly conservative DNA motif named the W box (T/CTGACC/T; Rushton et al., 2010). Since the first WRKY TF (SPF1) is cloned from sweet potato (*Ipomoea batatas*; Ishiguro and Nakamura, 1994), an increasing number of WRKY family genes are identified in several species, including Arabidopsis (75; Wu et al., 2005), barley (103; Kan et al., 2021), flax (102; Yuan et al., 2021), ginseng (137; Di et al., 2021), Gossypium (109; Ding et al., 2015), grape (59; Wang et al., 2015), Musa (147; Goel et al., 2016), maize (125; Hu et al., 2021), Taxus (61; Zhang et al., 2018), rice (109; Okay et al., 2014), walnut (103; Hao et al., 2021), and wheat (124; Ye et al., 2021).

In the present study, we analyzed the genome sequences of *Cymbidium sinense*, a very famous traditional orchid plant in China and Southeast Asia, and identified 64 *CsWRKY* proteins. Computational analysis was conducted to assess their physicochemical properties, including conserved domain, phylogenetic relationship, motif composition, functional annotation, and protein–protein interaction (PPI). In addition, the expression profiling of *CsWRKY* genes in various plant tissues in response to hormone treatments abscisic acid (ABA), gibberellic acid, salicylic acid, and methyl jasmonate (MeJA) indicated that most *CsWRKY* genes were responsive to various plant hormones, and Group III genes appeared to be more actively expressed to cope with the stress response. Collectively, our current findings provided valuable insights into the understanding of WRKY TFs in orchid plants and identified many candidate regulators involved in multiple hormone signaling.

Abbreviations: ABA, Abscisic acid; qRT-PCR, Quantitative RT-PCR; WGCNA, Weighted gene co-expression network analysis; WD, WRKY domain; GO, Gene ontology; PPI, Protein–protein interaction; ET, Ethylene; JA, Jasmonic acid; HMM, Hidden Markov Model; NJ, Neighbor-joining; Ks, Synonymous substitution rate; Ka, Non-synonymous substitution rate.

Results

Identification and analysis of WRKY genes in *Cymbidium sinense*

A total of 64 WRKY TFs were found through the *C. sinense* genomic database (Supplementary Table S2). Each WRKY gene was consistently named *CsWRKY1-CsWRKY64* based on its chromosomal location. The characteristics were analyzed, including the genome location, length of the open reading frame (ORF), and basic information of their encoded proteins, including length, molecular weight (MW), and isoelectric point (pI). Supplementary Table S2 shows that the length of CsWRKY proteins ranged from 67 (*CsWRKY38*) to 676 (*CsWRKY61*) amino acids, and their average length was 317 residues. The pI ranged from 4.84 (*CsWRKY46*) to 10.00 (*CsWRKY62*), and the MW ranged from 5,929.61 Da (*CsWRKY12*) to 7,4025.70 Da (*CsWRKY61*).

Classification of *Cymbidium sinense* proteins

To explore the phylogenetic relationship among *CsWRKY* proteins, 275 conserved WDs, including 114 *OsWRKY* proteins from rice (*O. sativa* japonica) and 85 *AtWRKY* proteins from Arabidopsis, were extracted to construct the evolutionary tree using the maximum likelihood (ML) method. As shown in the evolutionary tree (Figure 1), the 64 *CsWRKY* proteins were categorized into three main groups (I, II, and III), and Group II proteins were further sorted into five subgroups (IIa, IIb, IIc, IId, and IIe). The *CsWRKY* TFs had two standard motifs, including a WD and a zinc-finger-like domain. The former, typically the WRKYGQK sequence, could combine with the W box cis-element to trigger the expressions of their downstream genes. Besides the WRKYGQK sequence, three variants, WKKYGQK (*CsWRKY47*) in Group IIa, WRKYGKK (*CsWRKY08*, *19*, and *22*) in Group IIb, and WRKYGRK (*CsWRKY49*, and 62) in Group IIc were also revealed, respectively. The latter, the zinc-finger-like domain, also had two types, namely C2H2 and C2HC.

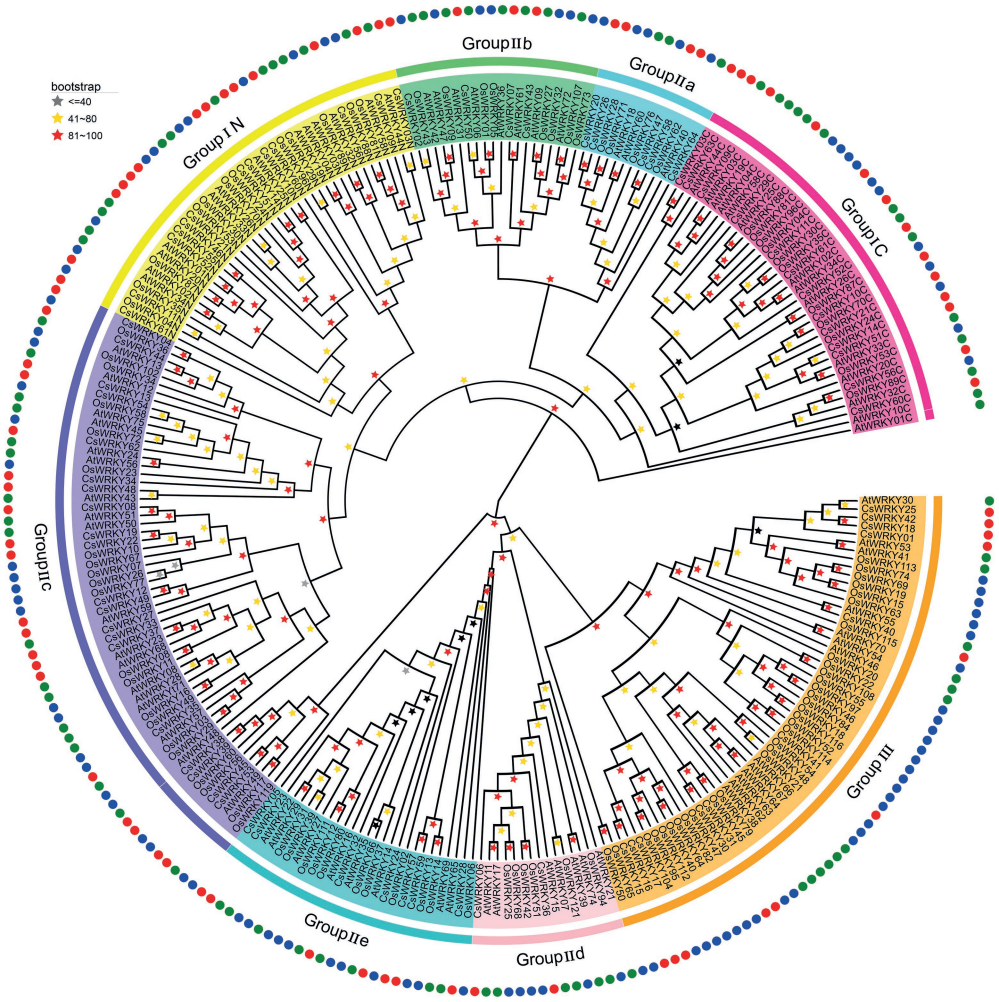


FIGURE 1
Phylogenetic tree of WRKY genes in *Cymbidium sinense* (Cs), *Arabidopsis thaliana* (At), and *Oryza sativa* (Os). An individual group (I–III) of ancestral relationships were presented in different colors.

It is worth noting that Group I was composed of 15 *CsWRKY* proteins, 11 of which contained two WDs. The remaining *CsWRKY63* had a single intact WD in its C terminus, and the other three members (*CsWRKY03*, *46*, and *55*) had a single WD at the N terminus (Supplementary Table S2; Supplementary Figure S1), implying that they most likely experienced domain loss or acquisition events in the evolutionary process (Ross et al., 2007). In addition, the zinc-finger motifs of the *CsWRKYs* in Group I belonged to the C2H2 type with a C-X4-C-X22–23-H-X1-H motif (Supplementary Table S2). There were 38 members in Group II, while 75 contained the motif of C-X4–5-6-C-X23–25-29-H-X1-H, and three members, *CsWRKY01*, *38*, and *44*, lacked a typical zinc-finger-like motif. According to their phylogenetic relationship, all 38 Group II members were further categorized into five subgroups as follows: Group IIa (five proteins), Group IIb (four proteins), Group IIc (20 proteins), Group IId (two proteins), and Group IIe (seven proteins). The zinc-finger motifs of Group III members (11) belonged to the C2HC type, with the C-X7-C-X23–24-26-H-X1-C motif (Supplementary Table S2), except for CsWRKY17 and 41 only containing fragmentary WRKY structure.

Analyses of chromosomal location, gene duplication, and genome synteny

A total of 64 candidate *CsWRKY* genes were mapped to 16 of the 20 chromosomes of *C. sinense* with an irregular arrangement (Figure 2). Chromosome 2 harbored the most significant number of *CsWRKYs* (10 genes), followed by chromosome 8 (eight genes). With only one gene, the least number of *CsWRKYs* was found on Chromosome 10. Three chromosomes contained the members from all three groups, 10 chromosomes contained the members from two groups, and the other three only had the members from one group.

To further assess the origin and evolution of the *CsWRKY* gene family, we constructed two comparative syntenic maps between

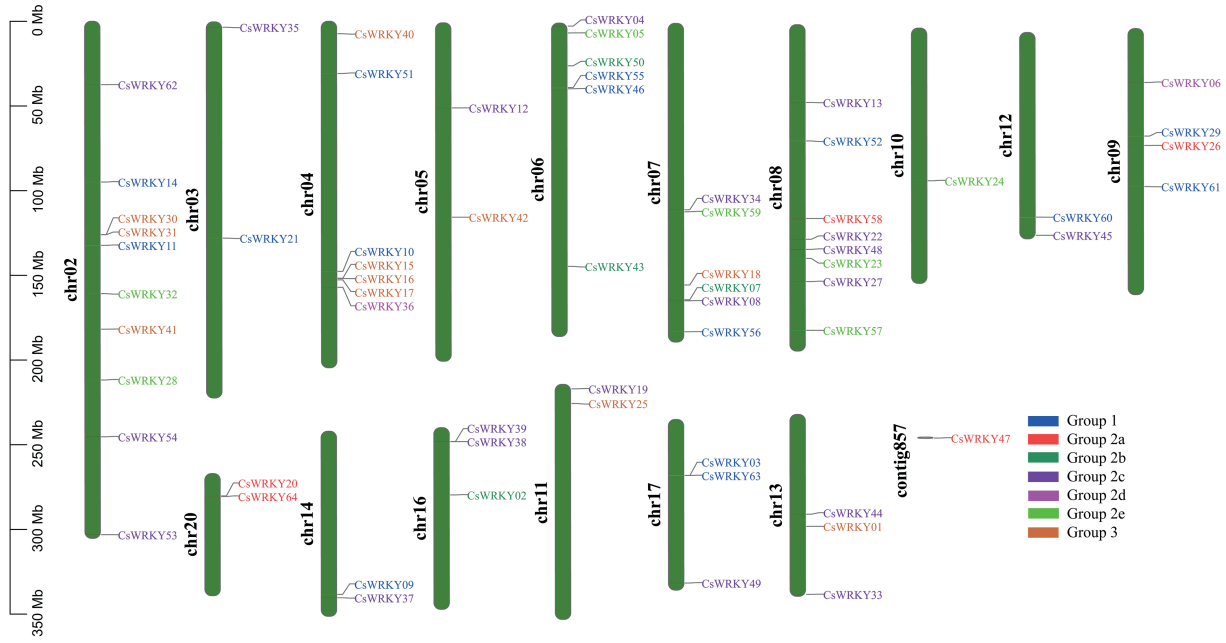


FIGURE 2
The chromosomal distribution and positioning of *CsWRKYs* across all seven chromosomes of *C. sinense*. A total of 16 chromosomes and one contig with varying lengths are shown in Mb (million base pair) scale on the left, where individual chromosomes (bars) are labeled with respective *CsWRKY* genes. The subfamily members are presented in various colors.

C. sinense and *A. thaliana*, *O. sativa*, *V. vinifera*, *M. cuminata*, or *Z. mays* at the genome-wide level. Supplementary Figure S2 shows that we finally identified *27*, *28*, and *29* orthologous gene pairs between *C. sinense* and *O. sativae*, *V. vinifera*, or *M. cuminata*, respectively. However, only 12 and 19 orthologous gene pairs between *C. sinense* and *Z. mays* or *A. thaliana*, were identified, respectively. More details about these orthologous gene pairs are given in Supplementary Table S5. There were far more orthologous genes between *C. sinense* and *O. sativae* than those between *C. sinense* and *A. thaliana*, which probably resulted from the nearer phylogenetic relationship between *C. sinense* and *O. sativae*.

To assess the selection pressure on various duplicated WRKY genes, we calculated the Ka and Ks substitution rates and the Ka/Ks ratios for each repeat *CsWRKY* gene pair. Typically, a Ka/Ks ratio of 1, >1, or <1 indicates neutral selection, adaptive evolution with positive selection, and negative or purifying selection, respectively (Nekrutenko et al., 2002). The Ka/Ks ratio for all seven segmentally duplicated gene pairs was <1 (Supplementary Table S4; Supplementary Figure S3), showing the high conservation of *CsWRKY* genes during evolution.

Gene structure and conserved motif analysis of *CsWRKY* genes

To gain more in-depth insight into the development of the WRKY gene family in *C. sinense*, we mapped the genetic structure of each *CsWRKY* gene. Based on the full-length *CsWRKY* proteins (Figure 3), we constructed a phylogenetic tree to clarify the gene structure and conserved motifs better. The number of introns in *CsWRKY* genes ranged from 0 to 10, similar to rice from 0 to 8 (Xie et al., 2005). Most *CsWRKY* genes harbored one to four introns, with 23 members containing two introns, 13 containing three introns, 10 containing one intron, and nine containing four introns. The other two *CsWRKY* genes contained five introns, and three *CsWRKY* genes contained seven, eight, and 10 introns. Additionally, the remaining four *CsWRKY* genes contained no intron. Figure 3 reveals that most *CsWRKY* genes within the same group or subgroup had a similar gene organization, showing the functional similarity among members.

Moreover, we identified 15 conserved motifs within proteins by the MEME program. MEME motif study showed that each *CsWRKY* protein had its specific conserved motifs (Supplementary Table S3). Figure 3 depicts that the WRKY family members with similar motif structures could be categorized into the same group. Almost all *CsWRKYs* had the conserved heptapeptides WRKYGQK (Motif 1 or Motif 4) and harbored at least one motif. Besides, Motif 2 was composed of the C2H2 motif, while Motif 12 consisted of the C2HC motif. Furthermore, the conserved motifs were explicitly present in various groups. For instance, Group IIa members harbored four conserved motifs (Motifs 1, 2, 3, and 6), Group IIb members contained five conserved motifs (Motifs 1, 2, 5, 6, and 15), and 14 Group IIc members contained four conserved motifs (Motifs 1, 2, 3, and 5). Clearly, a few motifs existed in one or more groups and subgroups. For example, Motifs 4, 8, 7, and 12 existed in Group I members, and Motif 10 existed only in Group III,

FIGURE 3
Phylogenetic relationships, gene structure, and architecture of conserved protein motifs in WRKY genes from *C. sinense*.

while Motif 15 was mainly present in Group IIb. This result suggested that these groups were endowed various functions during evolution. However, even if these members belong to the same clade, the functions of family members vary greatly (Yang et al., 2019). These WRKY proteins may have different functions though classified in the same group. Despite CsWRKY28 belonged to group IIe, it had the similar conserved motifs with the group IIc.

The cis-elements in the promoters of *Cymbidium sinense* WRKY genes

Cis-elements in promoter region are important for gene expression (Rombauts et al., 1999). To verify the potential function of *CsWRKY* genes in response to abiotic stress, we extracted the 2,000-bp promoter sequences of the *CsWRKY* genes and analyzed them for cis-elements using the PlantCARE database. Supplementary Figure S4 depicts that 20 types of cis-acting elements related to stresses and phytohormone responses were identified in the promoters of *CsWRKY* genes, including four defense-and stress-responsive elements (W box and TC-rich repeats) and seven hormone-related elements (ABRE, AuxRE, As-1, TGA-element, CGTCA-motif, and TGACG-motif). Supplementary Figure S4 shows that MYC and MYB were the most numerous elements in the promoter regions of all 64 *CsWRKYs*, with 61 genes containing these two elements. Box 4 and G box were two types of light-response elements, ARE was involved in anaerobic induction, and ABRE (ABA-responsive elements), CGTCA-motif, and TGACG-motif (MeJA-responsive elements) also appeared frequently in promoter regions of *CsWRKY* genes, which were found in 59, 49, 49, 47, 42, and 42 promoters, respectively. Many W-boxes were identified in 37 *CsWRKY* gene promoters, indicating that these genes were modulated by other WRKY TFs or themselves. LTR and MBS elements that respond to low temperature and drought stresses were identified in the promoters of 17 and 28 *CsWRKY* genes, respectively. Moreover, 25 promoters showed a TC-rich repeats element (cis-acting element participated in defense and stress response), and only 15 promoters had a GARE-motif (gibberellin-responsive element). Thus, the cis-element analysis showed that the expressions of *CsWRKY* genes in *C. sinense* might be related to various environmental factors.

Interaction analysis of specific *CsWRKY* proteins

A PPI network of *CsWRKY* proteins was established using STRING 10.5 software based on the Arabidopsis association model. Our results showed that there were close interaction networks among 43 *CsWRKYs* (Supplementary Figure S5). Among them, 14 proteins belonged to Group I, 22 proteins belonged to Group II, and seven proteins belonged to Group III. The extensive WRKY-WRKY interactions manifested a mechanism for functional cooperation and antagonism among WRKY proteins for dynamic regulation of target genes in *C. sinense*. In addition, *CsWRKY20*, *CsWRKY26*, *CsWRKY47*, *CsWRKY58*, *and CsWRKY64* belonging to Group IIa, *CsWRKY10*,

CsWRKY11, *CsWRKY14*, *CsWRKY21*, and *CsWRKY51* belonging to Group I, and *CsWRKY01*, *CsWRKY25*, *CsWRKY42* and *CsWRKY15* belonging to Group III were also involved in a more robust interaction network with other proteins. For instance, MAP kinase 4 (MPK4), functioning as a regulator of pathogen defense responses (Rasmussen et al., 2012; Bazin et al., 2020), and SIB1 are negative regulators of ABA-mediated leaf senescence (Zhang et al., 2022a), indicating that these genes played a kernel role in the biotic and abiotic stress response.

Analysis of expression profiles of *Cymbidium sinense* WRKY genes in different tissues

Based on RNA-seq and expression profiling analysis, 55 *CsWRKYs* were detectable in different plant tissues (Figure 4; Supplementary Table S6). However, the other nine transcripts (*CsWRKY10*, *12*, *15*, *17*, *38*, *41*, *47*, *59*, and *64*) were not detected, showing that these genes might be pseudogenes or have particular temporal and spatial expression patterns. Most of the 55 *CsWRKY* transcripts investigated were expressed in all tissues with low expression levels, suggesting that these TFs worked with other proteins synergistically or interactively during plant growth and development. Moreover, many genes had preferential expression in various tissues. For example, 19 *CsWRKY* transcripts (*CsWRKY21*, *55*, *56*, *60*, and *61* from Group I, CsWRKY*07*, *43*, and *50* from Group IIb, *CsWRKY19*, *22*, *33*, *34*, *48*, and *54* from Group IIc, *CsWRKY25* from Group IId, *CsWRKY05*, *24*, and *57* from Group IIe, and *CsWRKY42* from Group III) exhibited high expression levels in root particularly. By contrast, 2, 10, 3, and 2 exhibited the highest transcript abundances in the stem, leaf, flower, and fruit, respectively. These genes might play special roles in specific tissues. Additionally, several gene pairs with a close relationship, such as *CsWRKY03/63* and *CsWRKY19/33*, indicated similar expression profiles, suggesting that the functionality of these genes was redundant.

Expression profiles of *Cymbidium sinense* WRKY genes with hormone treatment

WRKY are widely involved in plant abiotic stress regulation (Jiang et al., 2017). To further certify the hormone response of *CsWRKY* genes, quantitative real-time PCR (qRT-PCR) was accomplished for 21 randomly chosen *CsWRKY* genes, including four Group I members, 11 Group II members, and six Group III members. We first optimized the treatment conditions, and the seedling leaves were sampled after exposure to 100 μmol GA, ABA, SA, IAA, JA, or ACC for 2 h.

Figure 5 shows that the expressions of *CsWRKY01*, *CsWRKY18*, *CsWRKY20*, *CsWRKY25*, *CsWRKY30*, *CsWRKY31*, *CsWRKY37*, and *CsWRKY42* were significantly up-regulated by over 5-fold under the GA treatment (Figures 5A,B). Under the ABA treatment, all of *CsWRKYs* were significantly induced, except that *CsWRKY45* was significantly down-regulated (Figure 5). Under the SA treatment, *CsWRKY20*, *CsWRKY25*, *CsWRKY30*, and *CsWRKY31* were significantly induced by over 15-fold (Figures 5B,C). The expression of *CsWRKY20* was up-regulated by over 5-fold under the IAA treatment, which was induced by over 6-fold under the JA treatment and over 125-fold under the ACC treatment (Figure 5B). More interestingly, Group III genes sensitive to responses to various hormones played important roles in responses to abiotic stress in *C. sinense*.

Subcellular localization of *CsWRKY18*

Since subcellular location information can provide some clues for protein function research, we used the online software Wolf PSORT to predict the subcellular locations of *CsWRKY* proteins in this study (Supplementary Table S7). Almost all CsWRKY proteins were mainly located in the nucleus, except for seven genes located in various organelles, such as cytoplasm, chloroplast, mitochondria, and peroxisome. *CsWRKY18* protein was chosen for subcellular localization verification. The full-length cDNA without the termination codon was fused in-frame to the 5′-end of the GFP gene under the control of the CaMV 35S promoter (Figure 6A). The recombined vector pAN580-*CsWRKY18*-GFP was transfected into leaf base protoplasts of Cymbidium. Confocal microscopy revealed that the fusion protein *CsWRKY18*-GFP was detected specifically in the nucleus, while GFP control displayed ubiquitous distribution in the whole cell. Furthermore, DAPI (4′,6-diamidino-2-phenylindole) staining showed that CsWRKY was clearly localized to the nucleus (Figure 6B).

Overexpression of the *CsWRKY18* gene increases ABA sensitivity

Figure 7A shows that under normal conditions, no significant difference in germination between WT plants and three transgenic lines was found. However, when 5 μM exogenous ABA was given, the transgenic *Arabidopsis* lines showed dramatically lower cotyledon greening rates than WT plants. Moreover, during the post-germinative growth stage, 4-day-old seedlings grown on 0.5× MS medium were transferred to vertical agar plates containing 1 μM or 5 μM ABA. We found that the root growths of WT and transgenic lines were both markedly suppressed. However, the root lengths of transgenic plants were remarkably shorter compared with WT plants (Figure. 7C).

Overexpression of *CsWRKY18* enhances plant tolerance to abiotic stress

We further compared the root morphology of the WT and transgenic plants grown on 1/2 MS medium containing 0 mM, 100 mM, 150 mM, and 200 mM NaCl (Figure 8). Three transgenic

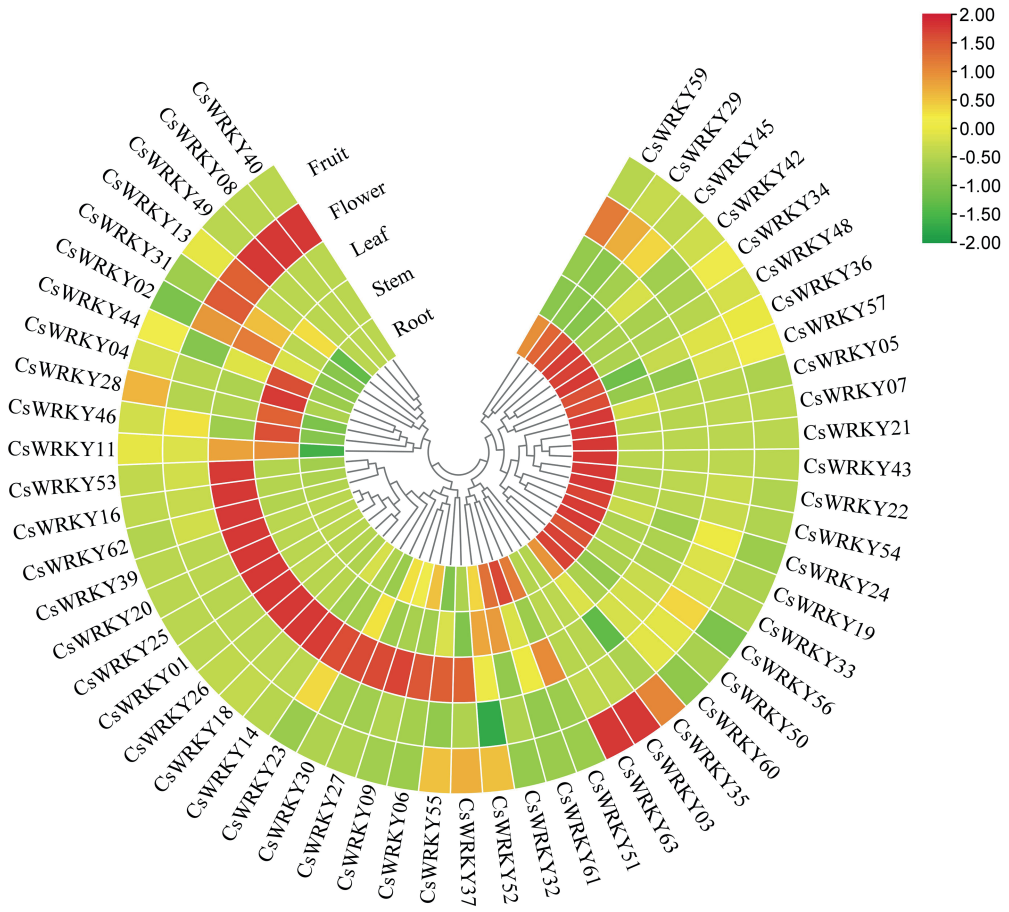


FIGURE 4
Heat map representation and hierarchical clustering of the *CsWRKY* gene expression profiles in five tissues.

lines exhibited longer roots compared with WT plants growing on all culture medium (Figure 8). In parallel, leaves of 4-week-old seedlings were treated with 200 mg/L PEG6000. After 12 h, dehydration and shrinkage of the WT leaves were observed. In contrast, the leaves of transgenic plants remained fresh and upright (Supplementary Figure S6).

To explore the mechanism of *CsWRKY18* in response to salt and osmotic stress, Tu, a well-known ABA biosynthesis inhibitor, was applied. When 0.75 mM Tu was supplemented to a medium containing 150 mM NaCl buffer, the salt and osmotic tolerance phenotypes in the transgenic lines were obviously weakened (Supplementary Figure S7). These results suggested that the *CsWRKY18* transgenic plants were tolerant to salt and osmotic stress, indicating a potential role of *CsWRKY18* in the ABA pathway.

Materials and methods

Plant materials and treatment

In the present study, the plants of *C. sinense*' Dharma' were grown in a greenhouse under controlled conditions (26°C/23°C, day/night, 16 h/8 h light/dark photoperiod with relative humidity 80%) at Environmental Horticulture Research Institute, Guangdong Academy of Agricultural Sciences (China). For tissue/organ-specific expression analysis, various plant parts, such as roots, leaves, stems, and flowers, were harvested from a 2-year-old plant. For expression in fruit tissues, 1/2/3/cm, mature green fruits, breaker fruits, and 10-day breaker fruits were harvested. For phytohormone treatment, 2-year-old plants in flowerpots were used, and each group ($n = 3$) was sprayed with 100 μM ABA, gibberellin (GA3), auxin (IAA), and MeJA. The samples were collected at 0, 0.5, 1, 2, 3, and 4 h after exposure from nine individual plants. For different organ samples, roots, stems, and leaves were collected from three plants grown for 45 days. All materials were immediately frozen in liquid nitrogen and preserved at −80°C for further analysis of WRKY genes. Considering the sensitive response of Group III to hormone signals, we randomly selected *CsWRKY18* for further functional analysis. We generated transgenic *A. thaliana* plants overexpressing the *CsWRKY18* gene under the CaMV 35S promoter, according to the general method (Clough and Bent, 1998; Feng et al., 2021) and we established 23 homozygous transgenic lines using hygromycin selection and PCR. Three T2-independent lines, L2, L11, and L13, were chosen

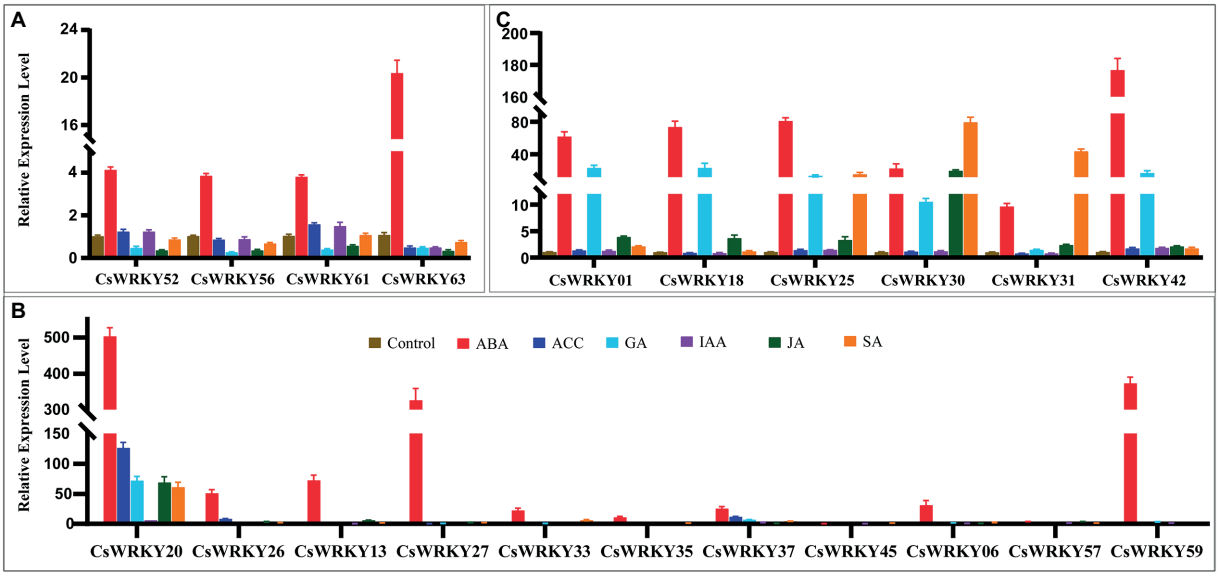


FIGURE 5
Expression patterns of 21 selected *CsWRKY* genes in *C. sinense* leaves under ABA, ACC, GA, IAA, JA, and SA stresses (**A**. Group I, **B**. Group II, and **C**. Group III). qRT-PCR data were normalized using the Cymbidium actin gene. X-axes represent various treatments (CK, normal condition). Different genes and y-axes are scales of relative expression level. Error bars result from three biological replicates.

for further phenotypic analysis according to the PCR analysis (Figure 7B). Seeds from transgenic and WT plants were incubated for germination on 1/2 MS medium supplemented with 1 μM ABA.

Identification of WRKY genes from *Cymbidium sinense*

The genome sequences of *C. sinense*, which were sampled in Guangzhou, Guangdong Province, China, were established by our lab (Yang et al., 2021), and the HMM file (PF03106) of the WD was obtained from the Pfam database.[1] To comprehensively identify the *CsWRKYs*, HMMER 3.0[2] software was adopted to search against the *C. sinense* protein sequences with default parameters (*E*-value cut-off <1E-5). The incorrect and redundant predicted sequences were manually removed, and then all putative *CsWRKY* genes were further verified using NCBI's Conserved Domain Database.[3] The MW and PI were evaluated using ExPASy-ProtParam online software.[4] Meanwhile, the protein sequences of *A. thaliana* and *O. sativa* were downloaded from TAIR[5] and Rice Genome Annotation Project,[6] respectively.

Phylogenetic analysis of *Cymbidium sinense* WRKY TFs

Multiple sequence alignment of the WDs and full-length proteins was conducted by CLUSTAL (gap opening penalty: 10 and gap extension penalty: 0.2; Larkin et al., 2007). The phylogenetic tree was reconstructed from 1,000 ultrafast bootstrap ML tree replicates using IQ-TREE v1.6.12 (Nguyen et al., 2015) with best-fit model selection (PMB + I + G4) by ModelFinder (Kalyaanamoorthy et al., 2017). The ML phylogenetic tree was visualized using Evolview v3 (Subramanian et al., 2019).

Characterization of the structure and motif of *CsWRKYs*

The available information of exons and introns was retrieved from the *C. sinense* genome sequences and then visualized by the TBtools v1.098725 (Chen et al., 2020) with coding sequences and genomic sequences. The MW and pI of the putative WRKY proteins were calculated by the ExPASy proteomics server.[7] The motifs of each deduced *CsWRKY* protein were analyzed by MEME suite software[8] (Bailey et al., 2009) with parameters as follows: the maximum number of motifs, 10. The upstream 2-Kb sequences of WRKY genes were

1 http://pfam.xfam.org/ (Accessed December 4, 2021).
2 http://hmmer.janelia.org/
3 https://www.ncbi.nlm.nih.gov/Structure/cdd/ (Accessed December 3, 2021).
4 http://web.expasy.org/protparam/ (Accessed January 3, 2022).
5 https://www.arabidopsis.org/
6 http://rice.plantbiology.msu.edu/ (Accessed November 11, 2021).
7 https://web.expasy.org/protparam/https://web.expasy.org/protparam/ (Accessed January 3, 2022).
8 http://meme-suite.org/tools/meme (Accessed January 12, 2022).

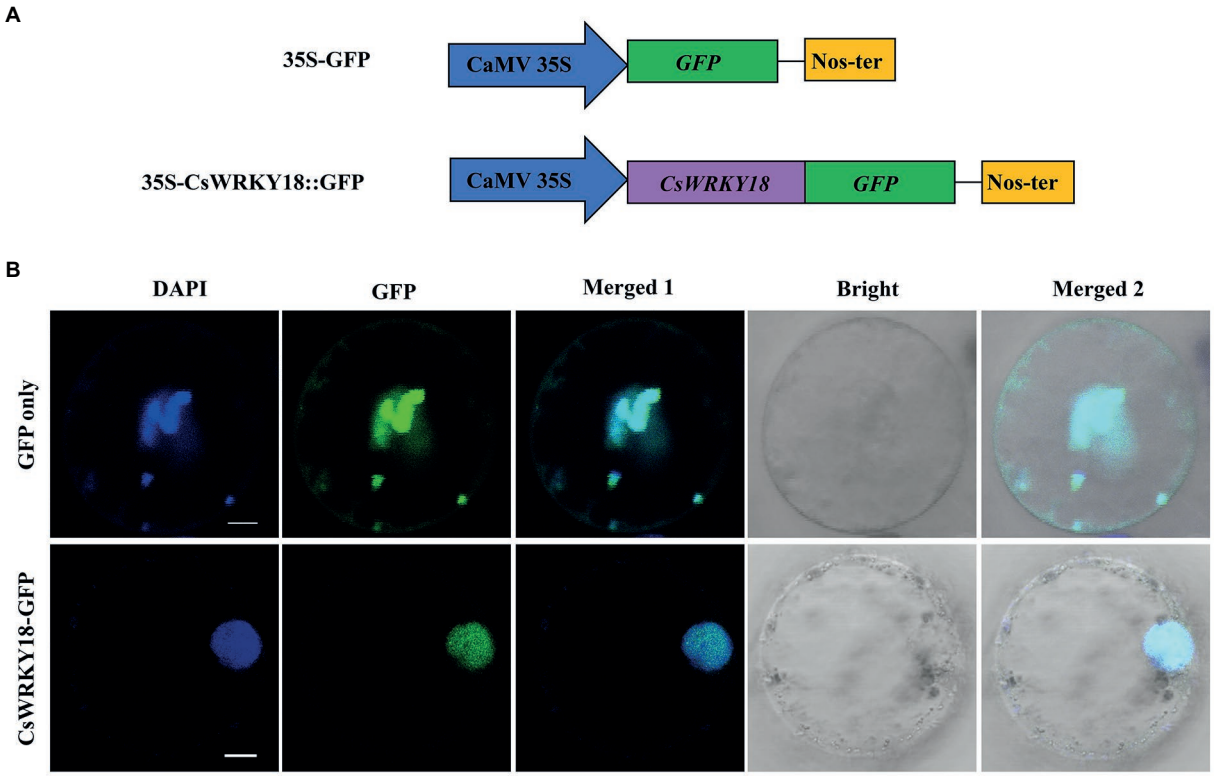


FIGURE 6
Subcellular localization of Group III *CsWRKY18* protein. **(A)** Schematic illustration of the constructs. **(B)** Subcellular location of free GFP and *CsWRKY18*-GFP protein in *C. sinense* leaves.

extracted to determine the cis-elements by using the PlantCARE database.[9]

Comparative genome synteny analysis

The alignment was performed by LASTZ using CDS sequences of *C. sinense* and five representative species (*A. thaliana*, *Oryza sativa*, *Vitis vinifera*, *Musa acuminate*, and *Zea mays*). The syntenic block map was established by MCscan with cscore = 0.7. Each orthologous gene pair was then further investigated with PAL2NAL[10] (Suyama et al., 2006) to determine the Ks (synonymous substitution rate) and Ka (non-synonymous substitution rate).

Transcriptomic data sets to analyze the expression patterns of WRKYs

To evaluate gene expression profiles of *CsWRKYs*, the expression patterns of *CsWRKY* genes in five different tissues were identified, such as roots, stems, leaves, flowers, and fruits. The transcriptome data of all tissues were came from our previous report (Yang et al., 2021). The expression values were calculated by log2 (FPKM) and displayed as a heat map generated using TBtools.

Subcellular localization of *CsWRKY* proteins

The subcellular localization was predicted using the online software WOLF PSORT.[11]

RNA extraction, cDNA synthesis, and qRT-PCR analysis

The above leaf samples were ground into a fine powder. Total RNA of these samples was extracted using Trizol Reagent (Invitrogen), and then 1 μg RNA was used to synthesize cDNA in a 20-μl reaction system according to the

9 http://bioinformatics.psb.ugent.be/webtools/plantcare/html/ (Accessed February 2, 2022).

10 http://www.bork.embl.de/pal2nal/

11 https://wolfpsort.hgc.jp/

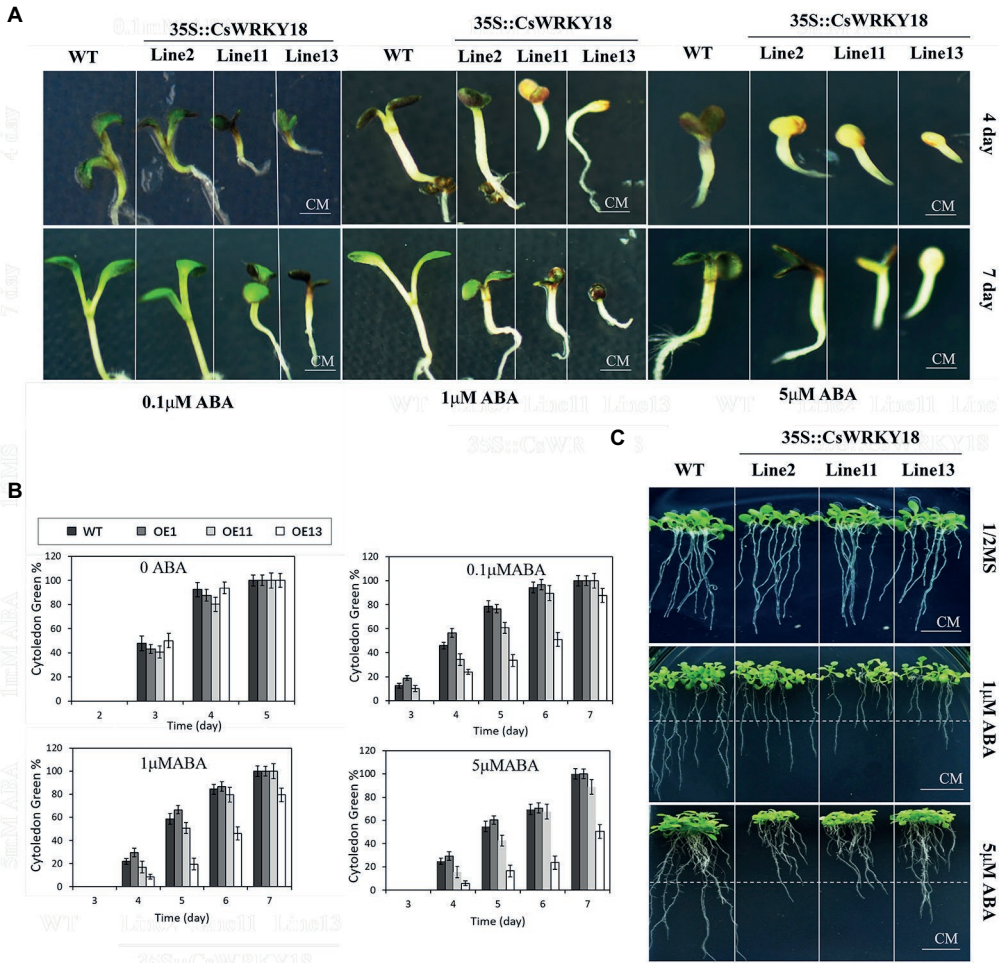


FIGURE 7
35S::*CsWRKY18* seedlings development under ABA treatment. **(A)** Seedling development of wild-type and 35S::*CsWRKY18* treated with 0.1, 1, and 5μM ABA. **(B)** Cytoledon green ratio of seedlings under ABA treatment. **(C)** Root elongation of wild-type and 35S::*CsWRKY18*.

instructions in PrimeScript™ RT reagent Kit with gDNA Eraser (TAKARA). qRT-PCR was carried out on a Bio-Rad iCycler Real-Time PCR Detection System (United States) by TaKaRa SYBR Premix Ex Taq™ (Tli RNaseH Plus) with three biological replications. Briefly, the reaction was conducted in a 20-μl reaction system containing 10 μl SYBR Premix (2×), 1 μl cDNA, 1 μl sense and anti-sense primer (10 μM), and 7 μl ddH_2O. After an initial denaturation step at 95°C for 1 min, the amplifications were carried out with 40 cycles at a melting temperature of 95°C for 10 s, an annealing temperature of 56°C for 30 s, and an extension temperature of 72°C for 30 s. Primer pairs were designed using Primer Premier 5.0, and the NCBI blast program was used to identify the specificity of all primers (Supplementary Table S1). The β-actin gene (Mol013347) in *C.sinense* was used as an internal reference, and the relative expression levels were calculated by the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001).

CsWRKY PPI analysis

A functional protein association network was constructed in the STRING program based on the Arabidopsis association model with the confidence parameter of 0.15 and the number of interactions of 5.

Discussion

Genes of the WRKY superfamily, one of the largest groups of TFs in plants, are also present in other eukaryotic lineages, such as flagellated protozoan (*Giardia lamblia*) and soil-living amoeba-like slime mold (*Dictyostelium discoideum*; Zhang and Wang, 2005; Phukan et al., 2016). WRKY family plays a vital functional Progress in gene identification and functional evaluation of WRKY family in plant germination to senescence or seed production in the whole plant's life cycle. WRKY TFs

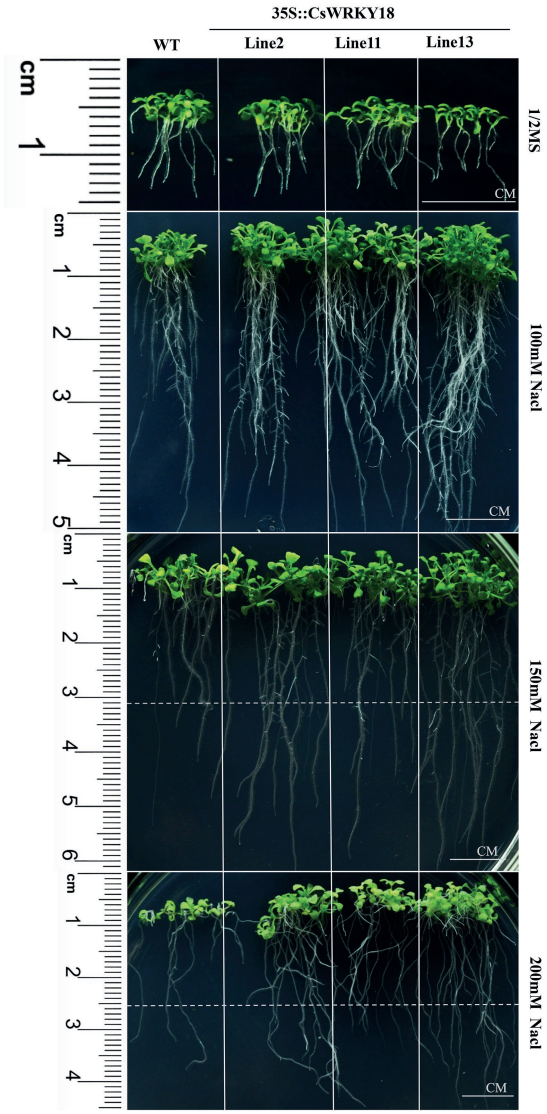


FIGURE 8
Root elongation of wild-type and 35S::*CsWRKY18* under NaCl treatment in 100, 150, and 200mM.

can enhance tolerance against heat, freezing, drought, salt, cadmium stress, UV radiation, and disease-causing organisms in different crop species, including Arabidopsis, Chrysanthemum, cucumber, peanut, pepper, rice, tobacco, tomato, and wheat (Dang et al., 2013; Liang et al., 2017; Chakraborty et al., 2018; Gao et al., 2018; Han et al., 2018, 2019; Hong et al., 2018; Khan et al., 2018; Kiranmai et al., 2018; Liu et al., 2018; Luan et al., 2019; Hussain et al., 2021).

However, compared with other crops, knowledge of the WRKY gene family has remained relatively scarce in the orchid field due to the relatively backward release of a complete and high-quality genome. In the present study, we identified 64 WRKY genes from the *C. sinense* genome. All *CsWRKY* were also classified into seven groups (groups I, IIa, IIb, IIc, IId, IIe, and III). Possibly since the green alga *Chlamydomonas reinhardtii* contains only one Group I WRKY gene (Zhang and Wang, 2005), Group I WRKY genes are considered the ancestors of other WRKY genes. In *C. sinense*, as many monocotyledonous species occur domain loss events (Ross et al., 2007; Wei et al., 2012; Brand et al., 2013), the 15 members of Group I contained 27 WDs. Subgroup II genes exist early in the evolution of plants (Hori et al., 2014). Interestingly, *CsWRKY47* in Group IIa had unique sequences in the WD (WKKYGQK) except for *CsWRKY03* and *CsWRKY63* in Group I. The WD sequence variation may influence the normal interactions and binding specificities with downstream target genes (van Verk et al., 2008; Zhou et al., 2008). Recently, a great quantity of Group I and Group II functions have been identified. For example, *CitWRKY28* (Group I) and *CitNAC029* can promote the accumulation of cuticular wax in Arabidopsis leaves (Yang et al., 2022). In *Dioscorea composita*, the expression of *DcWRKY3* (Group I) is strongly affected by salt stress (Yu et al., 2022). The transcription level of *CsWRKY25* (Group I) is up-regulated in *Penicillium digitatum*-infected citrus peels (Wang et al., 2022b). *PcWRKY11* (Group II WRKY from *Polygonum cuspidatum*) significantly increases the tolerance to salt stress in transgenic *A. thaliana* (Wang et al., 2022a).

Compared with Group I and II WRKY TFs, Group III TFs alter C2H2 to C2HC zinc-finger motif C-X7-C-X26-HTC. It has been considered the most adaptable and most advanced in monocot evolution (Eulgem et al., 2000; Zhang and Wang, 2005). In addition, Group III is the youngest in the WRKY family (Song and Gao, 2014; Wu et al., 2016). Therefore, in numerous species, Group III is the key to determining the number of WRKY genes. Nevertheless, Group III WRKY genes are considered the most vital group in gene family evolution and seemingly play an important role in the adaptation and evolution of plants (Zhang and Wang, 2005; Goel et al., 2016; Shan et al., 2021; Khan et al., 2022). In *C. sinense*, 11 *CsWRKY* gene members belonged to Group III, and such a number is fewer than the corresponding number in most other plants, such as 13 in Arabidopsis (Kalde et al., 2003), 28 in rice (Wang et al., 2015), 35 in maize (Hu et al., 2021), 62 in *Dendrobium catenatum* (Zhang et al., 2022b), 57 in *Phalaenopsis*, and 43 in *Apostasia* (unpublished). These findings suggested the functional conservation of WRKY genes during orchid evolution.

We analyzed the function of *CsWRKY18* and indicated its role in the ABA pathway and abiotic stress. Different abiotic stresses can induce *AtWRKY30*. Meanwhile, the overexpression of *AtWRKY30* greatly enhances the resistance of Arabidopsis in response to salt stress (Scarpeci et al., 2013) and improves heat and drought stress tolerance in wheat (El-Esawi et al., 2019). Moreover, *AtWRKY30* expression is induced by oxidative stress treatment, fungal elicitor, mosaic virus, JA, auxin, SA, and ABA (Jiang et al., 2017; Zou et al., 2019). Evolutionary processes including duplication events and chromosomal set changes (polyploidy)

could extend the members of a gene family in plants (Faraji et al., 2021; Heidari et al., 2021). Besides, it seems that WRKY gene family has been subjected to more evolutionary pressures and extended. However, evolutionary events could increase the members of gene family, but point mutations in coding site regions and upstream/downstream site of new members can affect the function of new members and cause diversity of members expression (Abdullah et al., 2021; Musavizadeh et al., 2021). Tandem duplication event as a 200 kb chromosome region containing two or more genes (Holub, 2001). All the 13 *CsWRKY* gene pairs among 64 WRKY genes were identified as segmental duplication, but no tandem duplication, indicating that tandem duplication events might have not anticipated in the amplification of *CsWRKY* gene family. More functions of *CsWRKYs* should be further studied.

Furthermore, six *CsWRKYs* were subdivided into one subclass IIIa, and five *CsWRKYs* were classified into another subclass IIIb. *OsWRKY45* (Group IIIa) is an essential positive player in the resistance against the rice blast fungus *M. oryzae* (Shimono et al., 2007). Various abiotic stresses can induce *AtWRKY30* (Group IIIb). Meanwhile, the overexpression of *AtWRKY30* significantly improves the resistance of Arabidopsis in response to salt stress (Scarpeci et al., 2013). These years, WRKY-c311842 TFs play key roles in berberine accumulation in Coptis chinensis (Liu et al., 2022). As an essential plant hormone in plants, ABA plays a vital role in development, physiological processes, and stress responses, and the ABA signaling pathway is an important component of the stress regulation network (Li et al., 2017, 2018; Sun et al., 2018). Several WRKY TFs have been reported to be positive regulators of ABA-mediated stomatal closure, while some are negative regulators of seed germination and can also indirectly control flowering (Rushton et al., 2012). The ectopic expressions of *TaWRKY75*-A and *ZmWRKY79* in Arabidopsis improve the survival rate under drought stress by regulating ABA biosynthesis (Gulzar et al., 2021; Ye et al., 2021). *OsWRKY50* enhances salt stress tolerance *via* an ABA-independent pathway (Huang et al., 2021). In order to explore the *CsWRKY* Group III gene family, we elucidated the function of *CsWRKY18* in drought stress through regulating ABA biosynthesis.

In summary, we identified 64 WRKY genes in *C. sinense* genome and analyzed their expression patterns in response to GA, ABA, SA, IAA, JA, and ACC treatments and their expression profiles in the leaves of *C. sinense*. The ectopic expression of *CsWRKY18* in Arabidopsis improved the survival rate under drought stress, suggesting that *CsWRKY18* was a regulator in ABA biosynthesis.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: NCBI, PRJNA743748.

Author contributions

G-FZ and F-XY designed the experiments and edited the manuscript. Y-LW, J-PJ, and JG executed the experiments and assembled the figures. C-QL, DL, JL, and QX conducted the qRT-PCR and PPI. F-XY, J-PJ, and Y-LW wrote the paper with inputs from other authors. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Natural Science Foundation of Guangdong province (2017A030312004), grants from the National Key R&D Program (2018YFD1000404 and 2019YFD1001003), the National Natural Science Foundation of China (31902065), the Science and Technology Project of Guangdong Province (2019B030316033), Guangzhou Science and Technology Project (201707010307 and 202002030016), Innovation Team of Modern Agriculture Industry Technology System in Guangdong Province (2021KJ121), and Guangdong Academy of Agricultural Sciences Discipline Team Construction Project (202127TD and BZ202006).

Acknowledgments

We thank the engineers of Nextomics Biosciences Co., Ltd. and PubBio Co., Ltd. for technical assistance in data processing and bioinformatics analysis and the Yuandong Chinese Orchid Corporation for plant cultivation. We are grateful to Da Luo of Sun Yat-sen University and Di-Qiu Yu of Yunnan University for their help during the preparation of the 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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.969010/full#supplementary-material

References

Abdullah, Faraji, S., Mehmood, F., Malik, H. M. T., Ahmed, I., Heidari, P., et al. (2021). The GASA gene family in cacao (*Theobroma cacao*, Malvaceae): genome wide identification and expression analysis. *Agronomy* 11:1425. doi: 10.3390/agronomy11071425

Ai, Y., Li, Z., Sun, W. H., Chen, J., Zhang, D. Y., Ma, L., et al. (2021). The Cymbidium genome reveals the evolution of unique morphological traits. *Hortic. Res.* 8:264. doi: 10.1038/s41438-021-00709-6

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

Bazin, J., Mariappan, K., Jiang, Y., Blein, T., Voelz, R., Crespi, M., et al. (2020). Role of MPK4 in pathogen-associated molecular pattern-triggered alternative splicing in Arabidopsis. *PLoS Pathog.* 16:e1008401. doi: 10.1371/journal.ppat.1008401

Brand, L. H., Fischer, N. M., Harter, K., Kohlbacher, O., and Wanke, D. (2013). Elucidating the evolutionary conserved DNA-binding specificities of WRKY transcription factors by molecular dynamics and in vitro binding assays. *Nucleic Acids Res.* 41, 9764–9778. doi: 10.1093/nar/gkt732

Brown, P. M. (2005). *Wild Orchids of Florida: With References to the Atlantic and Gulf Coastal Plains*. Gainesville: University Press of Florida.

Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W. C., Liu, K. W., et al. (2015). The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* 47, 65–72. doi: 10.1038/ng.3149

Chakraborty, J., Ghosh, P., Sen, S., and Das, S. (2018). Epigenetic and transcriptional control of chickpea WRKY40 promoter activity under Fusarium stress and its heterologous expression in Arabidopsis leads to enhanced resistance against bacterial pathogen. *Plant Sci.* 276, 250–267. doi: 10.1016/j.plantsci.2018.07.014

Chao, Y. T., Chen, W. C., Chen, C. Y., Ho, H. Y., Yeh, C. H., Kuo, Y. T., et al. (2018). Chromosome-level assembly, genetic and physical mapping of *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnol. J.* 16, 2027–2041. doi: 10.1111/pbi.12936

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

Chen, F., Hu, Y., Vannozzi, A., Wu, K. C., Cai, H. Y., Qin, Y., et al. (2017). The WRKY transcription factor family in model plants and crops. *Crit. Rev. Plant Sci.* 36, 311–335. doi: 10.1080/07352689.2018.1441103

Clough, S. J., and Bent, A. F. (1998). Floral dip:a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343

Dang, F. F., Wang, Y. N., Yu, L., Eulgem, T., Lai, Y., Liu, Z. Q., et al. (2013). CaWRKY40, a WRKY protein of pepper, plays an important role in the regulation of tolerance to heat stress and resistance to *Ralstonia solanacearum* infection. *Plant Cell Environ.* 36, 757–774. doi: 10.1111/pce.12011

Di, P., Wang, P., Yan, M., Han, P., Huang, X. Y., Yin, L., et al. (2021). Genome-wide characterization and analysis of WRKY transcription factors in Panax ginseng. *BMC Genomics* 22:834. doi: 10.1186/s12864-021-08145-5

Ding, M., Chen, J., Jiang, Y., Lin, L., Cao, Y., Wang, M., et al. (2015). Genome-wide investigation and transcriptome analysis of the WRKY gene family in Gossypium. *Mol. Gen. Genomics.* 290, 151–171. doi: 10.1007/s00438-014-0904-7

El-Esawi, M. A., Al-Ghamdi, A. A., Ali, H. M., and Ahmad, M. (2019). Overexpression of AtWRKY30 transcription factor enhances heat and drought stress tolerance in wheat (*Triticum aestivum* L.). *Genes* 10:13. doi: 10.3390/genes10020163

Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199–206. doi: 10.1016/s1360-1385(00)01600-9

Faraji, S., Heidari, P., Amouei, H., Filiz, E., Abdullah, and Poczai, P. (2021). Investigation and computational analysis of the Sulfotransferase (SOT) gene family in potato (*Solanum tuberosum*): insights into sulfur adjustment for proper development and stimuli responses. *Plan. Theory* 10:2597. doi: 10.3390/plants10122597

Feng, C., He, C., Wang, Y., Xu, H., Xu, K., Zhao, Y., et al. (2021). Genome-wide identification of soybean shaker K+ channel gene family and functional characterization of GmAKT1 in transgenic *Arabidopsis thaliana* under salt and drought stress. *J. Plant Physiol.* 266:153529. doi: 10.1016/j.jplph.2021.153529

Gao, H. M., Wang, Y. F., Xu, P., and Zhang, Z. B. (2018). Overexpression of a WRKY transcription factor TaWRKY2 enhances drought stress tolerance in transgenic wheat. *Front. Plant Sci.* 9:997. doi: 10.3389/fpls.2018.00997

Goel, R., Pandey, A., Trivedi, P. K., and Asif, M. H. (2016). Genome-wide analysis of the Musa WRKY gene family: evolution and differential expression during development and stress. *Front. Plant Sci.* 7:299. doi: 10.3389/fpls.2016.00299

Gulzar, F., Fu, J. Y., Zhu, C. Y., Yan, J., Li, X. L., Meraj, T. A., et al. (2021). Maize WRKY Transcription factor ZmWRKY79 positively regulates drought tolerance through elevating ABA biosynthesis. *Int. J. Mol. Sci.* 22:19. doi: 10.3390/ijms221810080

Han, D. G., Ding, H. B., Chai, L. J., Liu, W., Zhang, Z. Y., Hou, Y. J., et al. (2018). Isolation and characterization of MbWRKY1, a WRKY transcription factor gene from *Malus baccata* (L.) Borkh involved in drought tolerance. *Can. J. Plant Sci.* 98, 1023–1034. doi: 10.1139/cjps-2017-0355

Han, Y. Y., Fan, T. T., Zhu, X. Y., Wu, X., Ouyang, J., Jiang, L., et al. (2019). WRKY12 represses GSH1 expression to negatively regulate cadmium tolerance in Arabidopsis. *Plant Mol. Biol.* 99, 149–159. doi: 10.1007/s11103-018-0809-7

Hao, F., Yang, G., Zhou, H. J., Yao, J. J., Liu, D. R. L., Zhao, P., et al. (2021). Genome-wide identification and transcriptional expression profiles of transcription factor WRKY in common walnut (*Juglans regia* L.). *Genes* 12:19. doi: 10.3390/genes12091444

Hasing, T., Tang, H. B., Brym, M., Khazi, F., Huang, T. F., and Chambers, A. H. (2020). A phased *Vanilla planifolia* genome enables genetic improvement of flavour and production. *Nature Food* 1, 811–819. doi: 10.1038/s43016-020-00197-2

Heidari, P., Abdullah, Faraji, S., and Poczai, P. (2021). Magnesium transporter gene family: genome-wide identification and characterization in *Theobroma cacao*, Corchorus capsularis, and *Gossypium hirsutum* of family Malvaceae. *Agronomy* 11:1651. doi: 10.3390/agronomy11081651

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

Hong, Y. H., Cui, J., Liu, Z., and Luan, Y. S. (2018). SpWRKY6 acts as a positive regulator during tomato resistance to *Phytophthora infestans* infection. *Biochem. Biophys. Res. Commun.* 506, 787–792. doi: 10.1016/j.bbrc.2018.10.155

Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat. Commun.* 5:3978. doi: 10.1038/ncomms4978

Hu, W., Ren, Q., Chen, Y., Xu, G., and Qian, Y. (2021). Genome-wide identification and analysis of WRKY gene family in maize provide insights into regulatory network in response to abiotic stresses. *BMC Plant Biol.* 21:427. doi: 10.1186/s12870-021-03206-z

Huang, S. Z., Hu, L. J., Zhang, S. H., Zhang, M. X., Jiang, W. Z., Wu, T., et al. (2021). Rice OsWRKY50 mediates ABA-dependent seed germination and seedling growth, and ABA-independent salt stress tolerance. *Int. J. Mol. Sci.* 22:12. doi: 10.3390/ijms22168625

Hussain, A., Khan, M. I., Albaqami, M., Mahpara, S., Noorka, I. R., Ahmed, M. A. A., et al. (2021). CaWRKY30 positively regulates pepper immunity by targeting CaWRKY40 against *Ralstonia solanacearum* inoculation through modulating defense-related genes. *Int. J. Mol. Sci.* 22:18. doi: 10.3390/ijms222112091

Ishiguro, S., and Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5′ upstream regions of genes coding for sporamin and beta-amylase from sweet potato. *Mol. Gen. Genet.* 244, 563–571. doi: 10.1007/bf00282746

Jiang, J. J., Ma, S. H., Ye, N. H., Jiang, M., Cao, J. S., and Zhang, J. H. (2017). WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* 59, 86–101. doi: 10.1111/jipb.12513

Kalde, M., Barth, M., Somssich, I. E., and Lippok, B. (2003). Members of the Arabidopsis WRKY group III transcription factors are part of different plant defense signaling pathways. *Mol. Plant-Microbe Interact.* 16, 295–305. doi: 10.1094/MPMI.2003.16.4.295

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

Kan, J. H., Gao, G. Q., He, Q., Gao, Q., Jiang, C. C., Ahmar, S., et al. (2021). Genome-wide characterization of WRKY transcription factors revealed gene duplication and diversification in populations of wild to domesticated barley. *Int. J. Mol. Sci.* 22:17. doi: 10.3390/ijms22105354

Khan, M. A., Kang, D. R., Wu, Y. F., Wang, Y., Ai, P. H., and Wang, Z. C. (2022). Characterization of WRKY gene family in whole-genome and exploration of flowering improvement genes in *Chrysanthemum lavandulifolium*. *Front. Plant Sci.* 13:861193. doi: 10.3389/fpls.2022.861193

Khan, M. I., Zhang, Y. W., Liu, Z. Q., Hu, J., Liu, C. L., Yang, S., et al. (2018). CaWRKY40b in pepper acts as a negative regulator in response to *Ralstonia solanacearum* by directly modulating defense genes including CaWRKY40. *Int. J. Mol. Sci.* 19:17. doi: 10.3390/ijms19051403

Kiranmai, K., Rao, G. L., Pandurangaiah, M., Nareshkumar, A., Reddy, V. A., Lokesh, U., et al. (2018). A novel WRKY transcription factor, MuWRKY3 (*Macrotyloma uniflorum* lam. Verdc.) enhances drought stress tolerance in

transgenic groundnut (*Arachis hypogaea* L.) plants. *Front. Plant Sci.* 9:346. doi: 10.3389/fpls.2018.00346

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Leitch, I. J., Kahandawala, I., Suda, J., Hanson, L., Ingrouille, M. J., Chase, M. W., et al. (2009). Genome size diversity in orchids: consequences and evolution. *Ann. Bot.* 104, 469–481. doi: 10.1093/aob/mcp003

Li, W., De Ollas, C., and Dodd, I. C. (2018). Long-distance ABA transport can mediate distal tissue responses by affecting local ABA concentrations. *J. Integr. Plant Biol.* 60, 16–33. doi: 10.1111/jipb.12605

Li, K., Yang, F., Miao, Y., and Song, C. P. (2017). Abscisic acid signaling is involved in regulating the mitogen-activated protein kinase cascade module, AIK1-MKK5-MPK6. *Plant Signal. Behav.* 12:e1321188. doi: 10.1080/15592324.2017.1321188

Liang, Q. Y., Wu, Y. H., Wang, K., Bai, Z. Y., Liu, Q. L., Pan, Y. Z., et al. (2017). Chrysanthemum WRKY gene DgWRKY5 enhances tolerance to salt stress in transgenic chrysanthemum. *Sci. Rep.* 7:4799. doi: 10.1038/s41598-017-05170-x

Liu, Z., Chen, S., Chen, S., and Chen, S. (2006). *The Genus CYMBIDIUM in China*. The Genus Cymbidium in China: Science Press.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* 2012:251364. doi: 10.1155/2012/251364

Liu, Q., Li, X., Yan, S. J., Yu, T., Yang, J. Y., Dong, J. F., et al. (2018). OsWRKY67 positively regulates blast and bacteria blight resistance by direct activation of PR genes in rice. *BMC Plant Biol.* 18:257. doi: 10.1186/s12870-018-1479-y

Liu, X. M., Tan, J. P., Cheng, S. Y., Chen, Z. X., Ye, J. B., Zheng, J. R., et al. (2022). Comparative transcriptome analysis provides novel insights into the molecular mechanism of berberine biosynthesis in Coptis chinensis. *Sci. Hortic.* 291:110585. doi: 10.1016/j.scienta.2021.110585

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Luan, Q. Q., Chen, C. H., Liu, M. Y., Li, Q., Wang, L. N., and Ren, Z. (2019). CsWRKY50 mediates defense responses to Pseudoperonospora cubensis infection in *Cucumis sativus*. *Plant Sci.* 279, 59–69. doi: 10.1016/j.plantsci.2018.11.002

Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A., and Nakamura, K. (2001). Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. *Biosci. Biotechnol. Biochem.* 65, 2428–2436. doi: 10.1271/bbb.65.2428

Musavizadeh, Z., Najafi-Zarrini, H., Kazemitabar, S. K., Hashemi, S. H., Faraji, S., Barcaccia, G., et al. (2021). Genome-wide analysis of potassium channel genes in rice: expression of the OsAKT and OsKAT genes under salt stress. *Genes* 12:784. doi: 10.3390/genes12050784

Nekrutenko, A., Makova, K. D., and Li, W. H. (2002). The K(A)/K(S) ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. *Genome Res.* 12, 198–202. doi: 10.1101/gr.200901

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

Niu, Z. T., Zhu, F., Fan, Y. J., Li, C., Zhang, B. H., Zhu, S. Y., et al. (2021). The chromosome-level reference genome assembly for *Dendrobium officinale* and its utility of functional genomics research and molecular breeding study. *Acta Pharm. Sin. B* 11, 2080–2092. doi: 10.1016/j.apsb.2021.01.019

Okay, S., Derelli, E., and Unver, T. (2014). Transcriptome-wide identification of bread wheat WRKY transcription factors in response to drought stress. *Mol. Gen. Genomics.* 289, 765–781. doi: 10.1007/s00438-014-0849-x

Phukan, U. J., Jeena, G. S., and Shukla, R. K. (2016). WRKY transcription factors: molecular regulation and stress responses in plants. *Front. Plant Sci.* 7:760. doi: 10.3389/fpls.2016.00760

Rasmussen, M. W., Roux, M., Petersen, M., and Mundy, J. (2012). MAP kinase cascades in Arabidopsis innate immunity. *Front. Plant Sci.* 3:169. doi: 10.3389/fpls.2012.00169

Rombauts, S., Déhais, P., Van Montagu, M., and Rouzé, P. (1999). PlantCARE, a plant cis-acting regulatory element database. *Nucleic Acids Res.* 27, 295–296. doi: 10.1093/nar/27.1.295

Ross, C. A., Liu, Y., and Shen, Q. J. (2007). The WRKY gene family in Rice (*Oryza sativa*). *J. Integr. Plant Biol.* 49, 827–842. doi: 10.1111/j.1744-7909.2007.00504.x

Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). WRKY transcription factors. *Trends Plant Sci.* 15, 247–258. doi: 10.1016/j.tplants.2010.02.006

Rushton, D. L., Tripathi, P., Rabara, R. C., Lin, J., Ringler, P., Boken, A. K., et al. (2012). WRKY transcription factors: key components in abscisic acid signalling. *Plant Biotechnol. J.* 10, 2–11. doi: 10.1111/j.1467-7652.2011.00634.x

Scarpeci, T. E., Zanor, M. I., Mueller-Roeber, B., and Valle, E. M. (2013). Overexpression of AtWRKY30 enhances abiotic stress tolerance during early growth stages in *Arabidopsis thaliana*. *Plant Mol. Biol.* 83, 265–277. doi: 10.1007/s11103-013-0090-8

Shan, N., Xiang, Z. J., Sun, J. Y., Zhu, Q. L., Xiao, Y., Wang, P. T., et al. (2021). Genome-wide analysis of valine-glutamine motif-containing proteins related to abiotic stress response in cucumber (*Cucumis sativus* L.). *BMC Plant Biol.* 21:492. doi: 10.1186/s12870-021-03242-9

Shimono, M., Sugano, S., Nakayama, A., Jiang, C. J., Ono, K., Toki, S., et al. (2007). Rice WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell* 19, 2064–2076. doi: 10.1105/tpc.106.046250

Song, Y., and Gao, J. (2014). Genome-wide analysis of WRKY gene family in Arabidopsis lyrata and comparison with Arabidopsis thaliana and *Populus trichocarpa*. *Chin. Sci. Bull.* 59, 754–765. doi: 10.1007/s11434-013-0057-9

Subramanian, B., Gao, S., Lercher, M. J., Hu, S., and Chen, W. H. (2019). Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 47, W270–W275. doi: 10.1093/nar/gkz357

Sun, L. R., Wang, Y. B., He, S. B., and Hao, F. S. (2018). Mechanisms for Abscisic acid inhibition of primary root growth. *Plant Signal. Behav.* 13:e1500069. doi: 10.1080/15592324.2018.1500069

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315

van Verk, M. C., Pappaioannou, D., Neeleman, L., Bol, J. F., and Linthorst, H. J. (2008). A novel WRKY transcription factor is required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. *Plant Physiol.* 146, 1983–1995. doi: 10.1104/pp.107.112789

Wang, Y., Feng, L., Zhu, Y., Li, Y., Yan, H., and Xiang, Y. (2015). Comparative genomic analysis of the WRKY III gene family in populus, grape, arabidopsis and rice. *Biol. Direct* 10:48. doi: 10.1186/s13062-015-0076-3

Wang, W., Li, T., Chen, Q., Yao, S., Deng, L., and Zeng, K. (2022b). CsWRKY25 improves resistance of Citrus fruit to *Penicillium digitatum via* modulating reactive oxygen species production. *Front. Plant Sci.* 12:818198. doi: 10.3389/fpls.2021.818198

Wang, G., Wang, X., Ma, H., Fan, H., Lin, F., Chen, J., et al. (2022a). PcWRKY11, an II-d WRKY transcription factor from *Polygonum cuspidatum*, enhances salt tolerance in transgenic *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 23:4357. doi: 10.3390/ijms23084357

Wei, K. F., Chen, J., Chen, Y. F., Wu, L. J., and Xie, D. X. (2012). Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Res.* 19, 153–164. doi: 10.1093/dnares/dsr048

Wu, J., Chen, J. B., Wang, L. F., and Wang, S. M. (2016). Genome-wide investigation of WRKY transcription factors involved in terminal drought stress response in common bean. *Front. Plant Sci.* 8:380. doi: 10.3389/fpls.2017.00380

Wu, K. L., Guo, Z. J., Wang, H. H., and Li, J. (2005). The WRKY family of transcription factors in rice and Arabidopsis and their origins. *DNA Res.* 12, 9–26. doi: 10.1093/dnares/12.1.9

Xie, Z., Zhang, Z. L., Zou, X. L., Huang, J., Ruas, P., Thompson, D., et al. (2005). Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol.* 137, 176–189. doi: 10.1104/pp.104.054312

Xu, Y. X., Lei, Y. T., Su, Z. X., Zhao, M., Zhang, J. X., Shen, G. J., et al. (2021). A chromosome-scale *Gastrodia elata* genome and large-scale comparative genomic analysis indicate convergent evolution by gene loss in mycoheterotrophic and parasitic plants. *Plant J.* 108, 1609–1623. doi: 10.1111/tpj.15528

Yang, F. X., Gao, J., Wei, Y. L., Ren, R., Zhang, G. Q., Lu, C. Q., et al. (2021). The genome of *Cymbidium sinense* revealed the evolution of orchid traits. *Plant Biotechnol. J.* 19, 2501–2516. doi: 10.1111/pbi.13676

Yang, K., Li, Y., Wang, S., Xu, X., Sun, H., Zhao, H., et al. (2019). Genome-wide identification and expression analysis of the MYB transcription factor in moso bamboo (*Phyllostachys edulis*). *PeerJ* 6:e6242. doi: 10.7717/peerj.6242

Yang, H., Zhu, Z., Zhang, M., Li, X., Xu, R., Zhu, F., et al. (2022). CitWRKY28 and CitNAC029 promote the synthesis of cuticular wax by activating CitKCS gene expression in citrus fruit. *Plant Cell Rep.* 41, 905–920. doi: 10.1007/s00299-021-02826-x

Ye, H., Qiao, L., Guo, H., Guo, L., Ren, F., Bai, J., et al. (2021). Genome-wide identification of wheat WRKY gene family reveals That TaWRKY75-A is referred to drought and salt resistances. *Front. Plant Sci.* 12:663118. doi: 10.3389/fpls.2021.663118

Yu, S. J., Lan, X., Zhou, J. C., Gao, K. X., Zhong, C. M., and Xie, J. (2022). *Dioscorea composita* WRKY3 positively regulates salt-stress tolerance in transgenic *Arabidopsis thaliana*. *J. Plant Physiol.* 269:153592. doi: 10.1016/j.jplph.2021.153592

Yuan, H. M., Guo, W. D., Zhao, L. J., 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* 22:375. doi: 10.1186/s12864-021-07697-w

Zhang, M., Chen, Y., Nie, L., Jin, X., Liao, W., Zhao, S., et al. (2018). Transcriptome-wide identification and screening of WRKY factors involved in the regulation of taxol biosynthesis in Taxus chinensis. *Sci. Rep.* 8:5197. doi: 10.1038/s41598-018-23558-1

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

Zhang, Y., and Wang, L. (2005). The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol. Biol.* 5:1. doi: 10.1186/1471-2148-5-1

Zhang, T. T., Xu, Y., Ding, Y. D., Yu, W. A., Wang, J., Lai, H. G., et al. (2022b). Identification and expression analysis of WRKY gene family in response to abiotic stress in *Dendrobium catenatum*. *Front. Genet.* 13:800019. doi: 10.3389/fgene.2022.800019

Zhang, H., Zhang, L., Ji, Y., Jing, Y., Li, L., Chen, Y., et al. (2022a). Arabidopsis SIGMA FACTOR BINDING PROTEIN1 (SIB1) and SIB2 inhibit WRKY75 function in abscisic acid-mediated leaf senescence and seed germination. *J. Exp. Bot.* 73, 182–196. doi: 10.1093/jxb/erab391

Zhang, W. X., Zhang, G. Q., Zeng, P., Zhang, Y. Q., Hu, H., Liu, Z. J., et al. (2021a). Genome sequence of Apostasia ramifera provides insights into the adaptive evolution in orchids. *BMC Genomics* 22:536. doi: 10.1186/s12864-021-07852-3

Zhang, Y., Zhang, G. Q., Zhang, D., Liu, X. D., Xu, X. Y., Sun, W. H., et al. (2021b). Chromosome-scale assembly of the Dendrobium chrysotoxum genome enhances the understanding of orchid evolution. *Hortic Res* 8:183. doi: 10.1038/s41438-021-00621-z

Zhou, Q. Y., Tian, A. G., Zou, H. F., Xie, Z. M., Lei, G., Huang, J., et al. (2008). Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnol. J.* 6, 486–503. doi: 10.1111/j.1467-7652.2008.00336.x

Zou, L. J., Yang, F., Ma, Y. H., Wu, Q. G., Yi, K. X., and Zhang, D. W. (2019). Transcription factor WRKY30 mediates resistance to cucumber mosaic virus in Arabidopsis. *Biochem. Biophys. Res. Commun.* 517, 118–124. doi: 10.1016/j.bbrc.2019.07.030

frontiers | Frontiers in Plant Science

TYPE Perspective
PUBLISHED 05 October 2022
DOI 10.3389/fpls.2022.976283

OPEN ACCESS

EDITED BY
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

REVIEWED BY
Surendra Sarsaiya,
Zunyi Medical University,
China
Pandiyan Muthuramalingam,
Gyeongsang National University,
South Korea
Ali Raza,
Fujian Agriculture and Forestry University,
China

*CORRESPONDENCE
Darren C. J. Wong
darren.wong@anu.edu.au
wongdcj@gmail.com

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

RECEIVED 23 June 2022
ACCEPTED 24 August 2022
PUBLISHED 05 October 2022

CITATION
Wong DCJ, Perkins J and Peakall R (2022) Conserved pigment pathways underpin the dark insectiform floral structures of sexually deceptive *Chiloglottis* (Orchidaceae). *Front. Plant Sci.* 13:976283. doi: 10.3389/fpls.2022.976283

Conserved pigment pathways underpin the dark insectiform floral structures of sexually deceptive *Chiloglottis* (Orchidaceae)

Darren C. J. Wong*, James Perkins and Rod Peakall

Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia

Sexually deceptive plants achieve pollination by enticing specific male insects as pollinators using a combination of olfactory, visual, and morphological mimicry. The sexually deceptive orchid genus *Chiloglottis* is comprised of some 30 species with predominantly dull green-red flowers except for the dark insectiform calli/callus structure from the labellum lamina. This unique structure mimics the female of the pollinator and potentially enhances the visibility of the mimic. However, the chemical and genetic basis for the color of these structures remains poorly understood across the genus. The goal of this study was to investigate the flower color biochemistry and patterns of gene expression across the anthocyanin and flavonol glycoside biosynthetic pathway within the calli structures across the three distinct clades of *Chiloglottis* (Formicifera, Reflexa, and Valida) using chemical and transcriptome analysis. Our phylogenomic analysis confirmed the close sister relationship between the Reflexa/Formicifera clades and reaffirms the basal position of the Valida clade. Additionally, the biochemical basis of the dark calli/callus structures is conserved across the genus. Nonetheless, the proportion of methoxylated anthocyanin and flavonol glycoside derivatives and the mean gene expression levels appear to differentiate the Reflexa and Formicifera clades from the Valida clade. In future studies, it will be of interest to tease apart the role of phylogeny, environment, pollinators, and other factors as potential drivers of the observed biochemistry and gene expression differences. It will also be important to characterize the function of candidate genes such as *DFR*, *LDOX,* and *FLS* in this fascinating case of flower color mimicry.

KEYWORDS

***Chiloglottis*, anthocyanin, sexual deception, transcriptome, phylogenomics, Orchidaceae, flavonol glycoside, pollination**

Introduction

Sexually deceptive plants achieve pollination by enticing specific male insects as pollinators using a combination of olfactory, visual, and morphological mimicry of female insects. Olfactory mimicry is the best-studied trait, but visual mimicry is perhaps the most striking aspect of sexually deceptive flowers to humans. Most sexually deceptive flowers are

predominantly dull red and green, aside from one or several prominent dark insect-like (insectiform) floral structures on the labellum. These structures, which can appear as stalked calli (de Jager and Peakall, 2016), fine "hairs" (Peakall, 1990; Vereecken et al., 2012; Vignolini et al., 2012; Cohen et al., 2021), and wart-like bumps (Gaskett and Herberstein, 2010; Gaskett et al., 2017), are thought to be visual and tactile cues that enhance the insect mimicry.

The sexually deceptive orchid genus *Chiloglottis* contains approximately 30 species occurring predominantly in eastern Australia and Tasmania, but with one species extending to New Zealand and sub-Antarctic islands (Jones, 2021). *Chiloglottis* species typically have dark maroon to black calli/callus which contrast with their otherwise green or dull red labellum lamina (Figure 1). To the human eye, the calli/callus resemble an insect perched on the flower. These structures serve dual roles: (1) as the primary source of chiloglottones, a class of 2,5-dialkylcyclohexan-1,3-dione natural products that are sexually attractive to the thynnine wasp pollinators of the group (Schiestl et al., 2003; Franke et al., 2009; Peakall et al., 2010; Wong et al., 2017, 2018, 2019), and (2) as a visual and tactile mimic of the female (de Jager and Peakall, 2016). Other floral traits have also been shown to be relevant to pollinator attraction and pollination. These include the amount (Schiestl, 2004) and composition of the semiochemicals used for pollinator attraction (Peakall et al., 2010), flower height (Peakall and Handel, 1993), and labellum size (de Jager and Peakall, 2016, 2019).

Our study builds on this large body of prior work, and in particular on recent work on the chemical composition and genetic regulation of the temporal and spatial color patterns underpinning the visual mimicry in *Chiloglottis trapeziformis* (Wong et al., 2022). Targeted metabolite profiling in this species revealed that the localized distribution of abundant cyanidin-based anthocyanins and diverse flavonol glycoside co-pigments give the black callus structure its distinct dark color. Developmental stage and tissue-specific differential expression of genes encoding dihydroflavonol reductase, leucoanthocyanidin dioxygenase, flavonol synthase, and flavonoid O-methyltransferase that controls specific branch point into respective anthocyanin and flavonol glycoside biosynthesis (and their methylated derivatives) may also assist in maintaining the color from the earliest bud stage through to the mature flower. These findings highlight the multi-gene and regulatory complexity of the floral color adaptations involved in securing pollination by sexual deception (Wong et al., 2022).

In this perspective article, phylogenetically-informed sampling across the three distinct clades of *Chiloglottis,* each with characteristic phenological and morphological traits, were performed to assess whether there are any phylogenetic differences in flower color biochemistry and patterns of gene expression within the calli. We first leveraged a new phylogenomic approach to confirm if phylogenetic relationships within and among the major clades continue to hold with many additional loci. Next, targeted metabolite profiling and transcriptomes analysis of representative *Chiloglottis* were performed to determine whether floral color pathways that are largely responsible for the distinct coloration of "insectiform" calli is conserved (see Supplementary Figure S1 for a general framework). Our results confirmed that the biochemistry and biosynthetic machinery is broadly conserved across three major clades of the genus. However, they also revealed some phylogenetic differences. We conclude with a discussion and suggestions for further research to aid in teasing apart the role of phylogeny, environment, pollinators, and other factors that may drive these observed biochemistry and gene expression shifts.

Phylogenomic relationships of *Chiloglottis*

Molecular phylogenetic studies of *Chiloglottis* based on limited nuclear and plastid sequences (Mant et al., 2002, 2005; Peakall et al., 2010; Miller and Clements, 2014) have consistently indicated that there are three major clades: Reflexa (REF), Formicifera (FOR) and Valida (VAL). Here we performed a targeted sequence assembly based on a customized multitiered sequence capture strategy (see Supplementary Methods) of 14 *Chiloglottis* species (18 samples obtained in this study, see Supplementary Table S1 for details) and 17 samples from an earlier study encompassing three *Chiloglottis* (3 samples) and 12 non-*Chiloglottis* (14 samples) species belonging to the Drakaeinae with representatives from *Arthrochilus*, *Caleana*, *Drakaea*, *Paracaleana*, and *Spiculaea* (Peakall et al., 2021).

Although our analysis encompasses some 5.8 million bp of sequence and spans both exonic and non-exonic (e.g., potential homologous untranslated regions, introns, and other off-target nuclear, chloroplast, and mitochondrial DNA) regions (Supplementary Figure S2), the shortcut coalescent ASTRAL phylogeny was broadly congruent with earlier studies (Figure 1A). Notably, the close sister relationship between the REF/FOR clades, and the basal position of the VAL clade was reconfirmed (Peakall et al., 2010). Very short branches and high levels of gene tree discordance continue to be observed within the clades (Figure 1A; Supplementary Figure S3 for details). These results point to widespread incomplete lineage sorting and rapid radiation within clades, a characteristic also noted as a feature of other sexually deceptive orchid genera (Bateman et al., 2018, 2021; Peakall et al., 2021; Wong and Peakall, 2022).

Distinguishable labellum and insectiform calli traits across the *Chiloglottis* clades

Drawing on a recent flora of Australian orchids (Jones, 2021), we have mapped the flowering time, flower size, and labellum size traits of our study species across the phylogeny (Supplementary Figures S4, S5). Generally, members of each clade share a characteristic flowering time: REF in late summer to autumn (December–May), FOR in late winter to spring (August–November), and VAL in late winter to summer (range

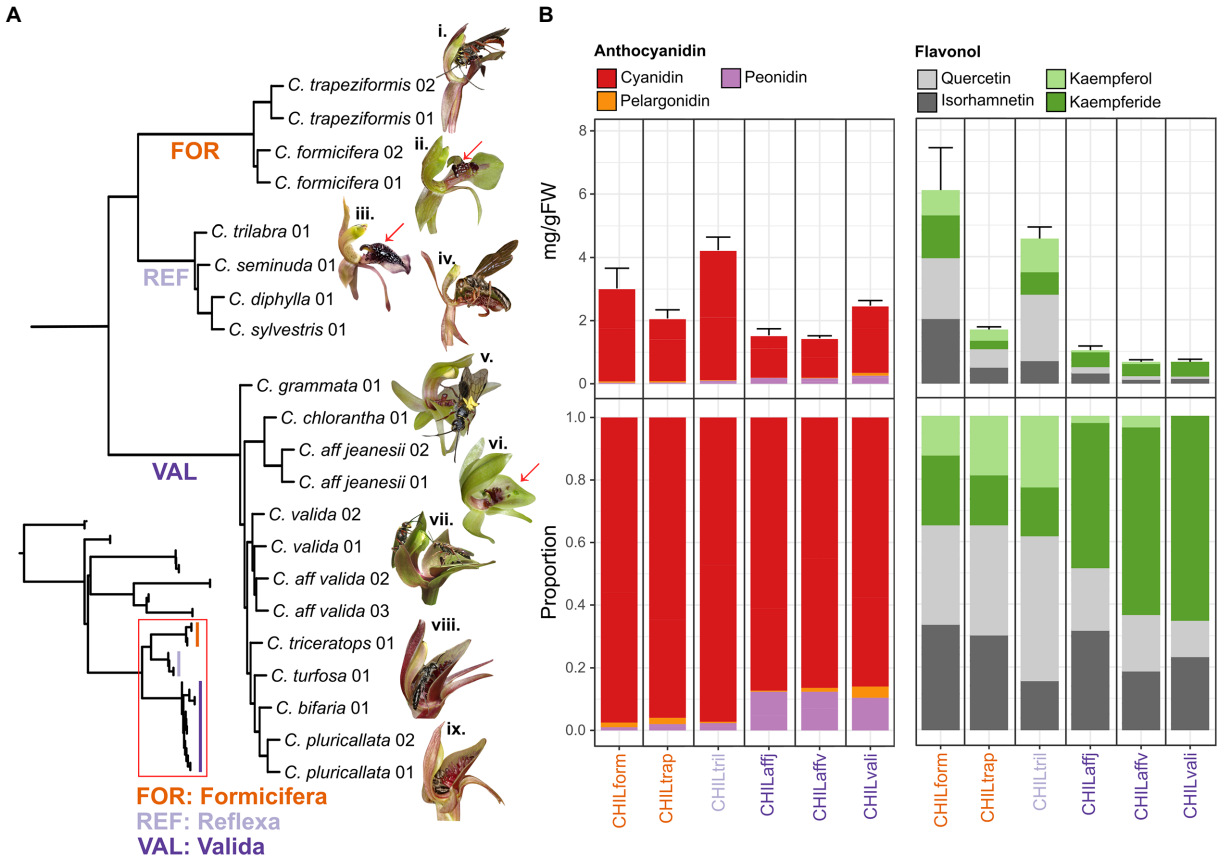


FIGURE 1
Phylogenetic patterns of anthocyanin and flavonol glycosides in a cross-section of Chiloglottis orchids. **(A)** The full (inset, top left) and close up (red boxed) ASTRAL species tree phylogeny of *Chiloglottis* (red boxed) with and without representatives from *Arthrochilus*, *Caleana*, *Drakaea*, *Paracaleana*, and *Spiculaea* genera as outgroup is illustrated, respectively. Representative flowers from the FOR (**i.** *C. trapeziformis*, **ii.** *C. formicifera*), REF (**iii.** *C. trilabra*, **iv.** *C. diphylla*), and VAL (**v.** *C. chlorantha*, **vi.** *C. aff. jeanesii*, **vii.** *C. aff. valida*, **viii.** *C. turfosa*, **ix.** *C. bifaria*) clades are shown. Red arrows point to the dark, three-dimensional insectiform calli/callus structure on the labellum lamina of representative FOR, REF, and VAL clades. Floral traits in each clade typically range as follows—FOR: flowers 6–8 mm wide × 12–16 mm long; labellum 6–7 mm wide × 8–10 mm long; length to width ratio 1.3–1.4 ×, REF: flowers 4–9 mm wide × 10–34 mm long; labellum 5–8 mm wide × 9–11 mm long; length to width ratio 1.5–1.9 ×, and VAL: flowers 15–30 mm wide × 18–30 mm long; labellum 9–14 mm wide × 10–17 mm long; length to width ratio close to 1. **(B)** Bar graphs showing total anthocyanins and flavonol glycoside co-pigment content grouped by aglycone (average±s.e.) and their corresponding proportion in the calli/callus of naturally opened flowers. See Supplementary Figure S3 for full ASTRAL species tree phylogeny and Supplementary Figures S4, S5 for detailed flowering time, flower size, and labellum size traits.

August–February). The sister REF/FOR clades share a common floral morphology with small single flowers [REF: 5–7 (*w*) × 17–21 (*l*) mm, FOR: 6–8 (*w*) × 12–16 (*l*) mm average] borne on taller peduncles (40–120 mm high). All flowers have a slim labellum for the proximal first third before expanding to a broadly triangular and spathulate (diamond) shape toward the tips, with their other petals reflexed against the ovary. The labella bear clusters of black calli that appear insectiform to human eyes. These highly clustered calli occupy 30%–80% of the total labellum area. By contrast, the single larger flowers (25–30 × 13–23 mm average) of the VAL clade are held closer to the ground (15–75 mm high), while the flowers have a larger and wider chordate-shaped labellum, with larger petals that are either held outward or incurved upwards. Furthermore, while the flowers also bear dark calli, these tend to cover a smaller proportion (10–40%) of the total labellum area, and generally appear much less insectiform to human eyes (see also Figure 1 legend for floral sizes by clade).

Floral color chemistry of *Chiloglottis* calli

Targeted metabolite profiling was performed to determine whether the floral color chemistry is conserved across the phylogeny. For each of the six *Chiloglottis* species surveyed, the calli and/or callus structures were first dissected from labellum lamina for subsequent anthocyanin/flavonol extraction and analysis by UHPLC–MS/MS (Supplementary Table S2) and LC-DAD-MS (Supplementary Figure S6). Five anthocyanins and 16 flavonol glycosides were detected with total anthocyanin content (expressed as mg g^{-1} FW) ranging between 1.42–4.21 in the calli (Figure 1B). The composition of anthocyanidins was broadly consistent across

the three clades with cyanidin-based anthocyanins as the dominant component (85%–99%) followed by peonidin- (1%–12%) and pelargonidin- (0%–4%) based anthocyanins as minor constituents. However, the proportion of peonidin-based anthocyanins were consistently higher in the VAL clade (10%–12%) compared to members of the FOR and REF (1%–3%) clades.

Total flavonol glycoside content (expressed as mg g^{-1} FW per flower) ranged between 0.64–6.07 in the calli (Figure 1B), with *C. formicifera* and *C. trilabra* showing markedly higher total flavonol glycoside content in their calli/callus compared to *C. trapeziformis*, *C. aff. valida*, *C. aff. jeanesii*, and *C. valida*. Four main types of flavonol glycosides were found, with the proportions of quercetin (Q)-, isorhamnetin- (IR), kaempferol- (K), and kaempferide- (Kde) based flavonol glycosides ranging between 11%–46%, 16%–44%, 0%–23%, and 16%–65%, respectively. Interestingly, Q and its methylated derivative, IR together often made up 62%–65% of total flavonol glycoside content in FOR and REF groups, whereas Kde alone make up 46%–65% of total flavonol glycoside across the VAL group. While variable across species, in general, the ratios of total anthocyanins to flavonol glycosides in the calli, were higher in VAL group members (1.5–3.8 to 1) than in FOR and REF group members (0.9–1.2 to 1).

Collectively, these findings indicate a phylogenetic pattern of compositional differences in anthocyanin, and especially flavonol glycoside pigment chemistry of the calli between VAL and REF/FOR flowers with higher proportions of methoxylated derivatives (peonidin, kaempferide, and isorhamnetin) appears to be a characteristic of the VAL clade. Here, we observed that the chemical basis of these black "insectiform" structures also shares surprising parallels observed in the European alpine food-rewarding orchid, *Gymnadenia rhellicani* (Kellenberger et al., 2019) where its high anthocyanin content, composed of cyanidin 3-O-glucoside and cyanidin 3-O-(6″-malonyl-glucoside) in the petal and labellum (similar composition observed across *Chiloglottis*), provides some flower morphs its black color. Beyond orchids, uniformly black flowers (Markham et al., 2004; Thill et al., 2012) or dark markings/spots in petals (Vignolini et al., 2015; van der Kooi and Stavenga, 2019) and unusual black 3D floral structures (Thomas et al., 2009) is also due to an oversaturation of anthocyanins, often in the presence of co-pigments.

Pigment biosynthetic pathways of *Chiloglottis* calli

Floral transcriptomes of species belonging to FOR (*C. trapeziformis*), REF (*C. seminuda*), and VAL (*C. turfosa*, *C. aff. valida*, and *C. valida*) clades were also investigated. The same sets of candidate genes encoding shared flavonoid and dedicated anthocyanin and flavonol glycoside pathway enzymes in *C. trapeziformis* were also identified in the representative REF and VAL species (Figure 2; Supplementary Table S3). Additionally, genes involved in the modification (e.g., glycosylation, methylation, and acylation) of anthocyanidins and flavonols at one or more positions were also identified. Generally, the diversity of anthocyanins and flavonol glycosides detected in the callus is in accord with the expression of candidate genes downstream of F3'H (two cyanidin and two peonidin anthocyanins, five Q and five IR flavonol glycosides) and those that do not depend on F3'H activity such as one pelargonidin anthocyanin, three K, and three Kde flavonol glycosides (Figure 1; Supplementary Figure S6; Supplementary Table S2).

The patterns of anthocyanin and flavonol glycoside-related pathway gene expression were also investigated. Mean levels of gene expression were compared across two groups: the three species representing the VAL clade (*C. turfosa*, *C. aff. valida,* and *C. valida*), and the pooled single representatives each of the REF (*C. seminuda*) and FOR (*C. trapeziformis*) group. The outcomes of this analysis are summarized in Figure 2, which maps the patterns of gene expression onto the anthocyanin/flavonol glycoside-related pathway. The mean levels of gene expression were frequently significantly higher in the VAL group than the REF/FOR group. For example, this pattern was found at the shared flavonoid (*CHS*, *CHI*, and *F3H*), dedicated anthocyanin (*DFR* and *LDOX*), and flavonol glycoside (*FLS*) biosynthesis genes. Significantly higher mean expression levels were also detected for some downstream modification genes (e.g., *FMT, MAT*, *5′GT*, and *7′GT*; Figure 2; Supplementary Figure S7) in the VAL group. These findings indicate a strong phylogenetic pattern of gene expression difference in the anthocyanin and flavonol glycoside pathways of the calli at the open flower stage when the calli were dissected in this study.

Environmental or pollinator-driven differences in biochemistry and gene expression?

Chiloglottis orchids are embedded within a well-resolved subtribe (Diurideae: Drakaeinae) in which all genera are sexually deceptive barring a few self-pollinating cases (Weston et al., 2014; Peakall et al., 2021). Because of this, floral adaptations associated with insect mimicry have likely been finely tuned by a long evolutionary process. Many studies have shown that differences in seasons, day length, light quality, and temperature additively affect the flavonoid composition of plant tissues (Jaakola and Hohtola, 2010). For example, shading or light exclusion significantly increased peonidin anthocyanins compared to exposed skins of berries (i.e., grapevine and bilberries) while cyanidin anthocyanins were generally unaffected (Downey et al., 2004; Zoratti et al., 2014). Some co-occurring *Chiloglottis* species flower at different times of the year (Jones, 2021), meaning that environmental conditions are likely to vary even among sympatric species (Supplementary Figure S4). Indeed, there are distinctly different flowering times for each of the three clades, with FOR being mainly spring flowering, REF being mainly autumn flowering, and VAL being mainly spring and summer flowering (Jones, 2021). Consequently, it is possible that seasonal differences affect the developmental conditions of the flowers and may partially account for the anthocyanin/flavonol glycoside compositional differences observed between the clades.

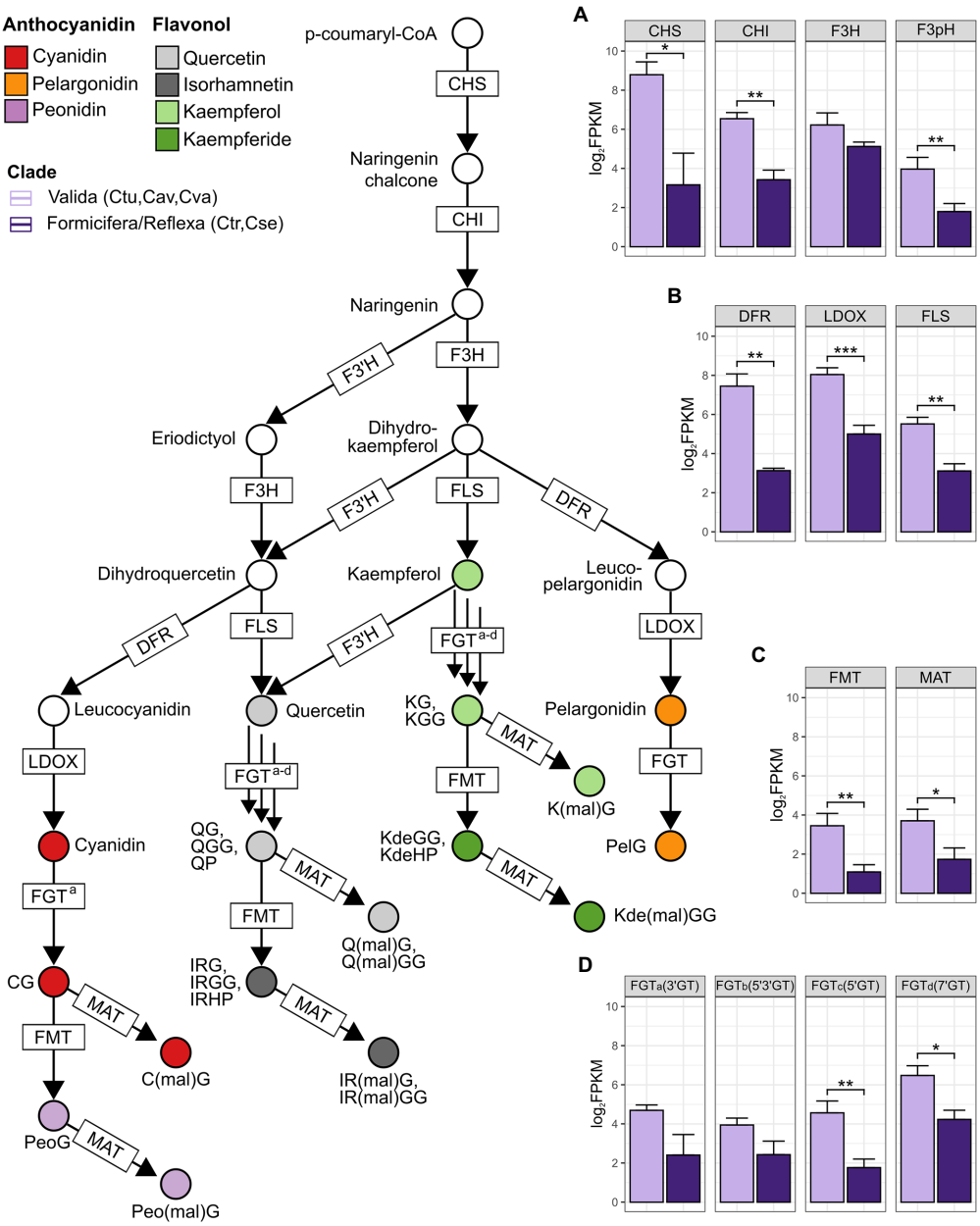


FIGURE 2
Phylogenetic patterns of anthocyanin/flavonol glycoside-related pathway expression of mature flower calli in a cross-section of *Chiloglottis* orchids. Boxplots from **(A–D)** depict the general patterns of genes in respective pathway steps across the three clades—REF (*Chiloglottis seminuda*), FOR (*Chiloglottis trapeziformis*), and VAL (*Chiloglottis turfosa*, *Chiloglottis aff. valida*, *Chiloglottis valida*). Asterisks indicate significant differences in the total gene expression between clade comparisons at $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***) based on ANOVA and Tukey's HSD test. See Supplementary Table S3 for full gene names and predicted function of corresponding gene symbols.

Interestingly, strong experimental evidence for pollinator-mediated selection on the floral traits that distinguish the VAL and FOR/REF clades is indicated by studies of one pair of *Chiloglottis* species. *C. trapeziformis* (FOR) and *C. valida* (VAL) use the same semiochemical, chiloglottone 1, to attract their own phylogenetically distinct wasp pollinators, *Neozeleboria cryptoides* and *N. monticola*, respectively. Key floral trait differences between the two orchids are strongly correlated with behavioral differences in approach, landing position, and orientation during attempted copulation by the two pollinators (de Jager and Peakall, 2016). However, in this study, the largest anthocyanin/flavonol glucoside compositional differences were found between these two species. Therefore, it seems unlikely that seasonal differences and distinct floral morphological features can fully explain the observed differences. Could these composition differences instead reflect pollinator preference? Although direct experiments on wasp

pollinator color preferences are lacking, spectral reflectance measurements provide some clues. When projected into hymenopteran vision space, the measurements indicate the calli/callus colors are very similar. In both orchids these structures are predicted to be perceived as achromatic to the male pollinator(s), matching the achromaticity of the females (de Jager and Peakall, 2016). Thus, it also seems unlikely that pollinator color preference could have driven the anthocyanin/flavonol glycoside composition differences observed between the VAL and REF/FOR clades.

Conclusion and future directions

Selection for flower color has been predicted to be stronger for plants with specialized pollinator interactions compared to those with multiple pollinators involving many diverse taxonomic groups (Trunschke et al., 2021). As expected the biochemical and biosynthetic basis of the dark calli/callus structures of the labellum is conserved across the genus (Wong et al., 2022). However, subtle shifts in anthocyanin and flavonol glucoside biochemistry such as a higher proportion of methoxylated derivatives, and differences in gene expression were evident between the VAL clade and the sister REF/FOR clades. In order to better understand the basis for these observed expression differences, a first step is to confirm that peak anthocyanin/flavonol glucoside and corresponding gene expression levels are consistently shifted to earlier developmental stages in REF/FOR clades compared to VAL clades. One promising candidate pair of study species to first test this hypothesis would be *C. trapeziformis* and *C. valida,* where there is occasional co-flowering and pollinator sharing.

Further investigation into the tissue-specific flower pigmentation patterns (e.g., calli/callus vs. labellum lamina) across the clade also warrants further investigation. In *C. trapeziformis*, increased accumulation of anthocyanins in the callus relative to the labellum lamina at flowering (Wong et al., 2022) results in strong within-flower chromatic/achromatic contrast, potentially aiding detectability by pollinators (de Jager and Peakall, 2016). However, for many other species from the FOR/REF clades, the black calli occupy nearly the entire labellum lamina (Supplementary Figure S5) leaving little to no room for within-flower contrast. The sexually deceptive *Drakaea livida* shares this characteristic, and in this case, the entire labellum is contrasted with the background substrate (Gaskett et al., 2017). Similarly, the dominance of the calli in FOR and REF clade species may promote contrast of the whole labellum with the background rather than within-flower contrast which is common in the VAL clade.

Innovative field experiments will be required to tease apart the importance of the color from the structure of the calli. Working with rare flower color variants may prove particularly informative for determining the importance of color to pollination success. Alternatively, working with 3D printed flowers with varying calli/callus colors at varying flower developmental stages is another promising avenue, and has been effective in *Dracula* orchids (Policha et al., 2016). In all cases, it will remain essential to hold the chemistry of pollinator attraction constant. Here the use of synthetic chiloglottones can be effectively used, as already demonstrated by the multiple studies that have investigated the role of chemistry in *Chiloglottis* pollinator attraction (e.g., Schiestl et al., 2003; Peakall et al., 2010). In future studies, it will also be of interest to tease apart the key signals (developmental and/or environmental), the function of candidate genes (e.g., *DFR*, *LDOX*, *FLS*, and *FMT*), and to more fully evaluate the role of pollinators in this fascinating case of flower color mimicry.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: NCBI PRJNA661963, PRJNA486025, and PRJNA390683.

Author contributions

DW and JP performed the experiments. RP and DW secured the funding, designed the study, and coordinated the experiments and data analysis. DW wrote the article with assistance from RP and JP. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Australian Research Council Projects (DE190100249 to DW and DP1094453 and DP150102762 to RP) and an Australian Government Research Training Program to JP.

Acknowledgments

We thank the staff at the ANU RSB/RSC Joint Mass Spectrometry Facility for assistance with LC–MS operation and Ryan Philips for fruitful discussion on the manuscript. The photographs of *Chiloglottis* orchids were taken by RP.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those

of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.976283/full#supplementary-material

References

Bateman, R. M., Rudall, P. J., Murphy, A. R. M., Cowan, R. S., Devey, D. S., and Peréz-Escobar, O. A. (2021). Whole plastomes are not enough: Phylogenomic and morphometric exploration at multiple demographic levels of the bee orchid clade *Ophrys* sect Sphegodes. *J. Exp. Bot.* 72, 654–681. doi: 10.1093/jxb/eraa467

Bateman, R. M., Sramkó, G., and Paun, O. (2018). Integrating restriction site-associated DNA sequencing (RAD-seq) with morphological cladistic analysis clarifies evolutionary relationships among major species groups of bee orchids. *Ann. Bot.* 121, 85–105. doi: 10.1093/aob/mcx129

Cohen, C., Liltved, W. R., Colville, J. F., Shuttleworth, A., Weissflog, J., Svatoš, A., et al. (2021). Sexual deception of a beetle pollinator through floral mimicry. *Curr. Biol.* 31, 1962–1969.e6. doi: 10.1016/j.cub.2021.03.037

de Jager, M. L., and Peakall, R. (2016). Does morphology matter? An explicit assessment of floral morphology in sexual deception. *Funct. Ecol.* 30, 537–546. doi: 10.1111/1365-2435.12517

de Jager, M. L., and Peakall, R. (2019). Experimental examination of pollinator-mediated selection in a sexually deceptive orchid. *Ann. Bot.* 123, 347–354. doi: 10.1093/aob/mcy083

Downey, M. O., Harvey, J. S., and Robinson, S. P. (2004). The effect of bunch shading on berry development and flavonoid accumulation in shiraz grapes. *Aust. J. Grape Wine Res.* 10, 55–73. doi: 10.1111/j.1755-0238.2004.tb00008.x

Franke, S., Ibarra, F., Schulz, C. M., Twele, R., Poldy, J., Barrow, R. A., et al. (2009). The discovery of 2,5-dialkylcyclohexan-1,3-diones as a new class of natural products. *Proc. Natl. Acad. Sci.* 106, 8877–8882. doi: 10.1073/pnas.0900646106

Gaskett, A. C., Endler, J. A., and Phillips, R. D. (2017). Convergent evolution of sexual deception via chromatic and achromatic contrast rather than colour mimicry. *Evol. Ecol.* 31, 205–227. doi: 10.1007/s10682-016-9863-2

Gaskett, A. C., and Herberstein, M. E. (2010). Colour mimicry and sexual deception by tongue orchids (*Cryptostylis*). *Naturwissenschaften* 97, 97–102. doi: 10.1007/s00114-009-0611-0

Jaakola, L., and Hohtola, A. (2010). Effect of latitude on flavonoid biosynthesis in plants. *Plant Cell Environ.* 33, 1239–1247. doi: 10.1111/j.1365-3040.2010.02154.x

Jones, D. L. (2021). *A complete guide to native orchids of Australia. 3rd Edn.* New South Wales, Australia: New Holland Publishers.

Kellenberger, R. T., Byers, K. J. R. P., De Brito Francisco, R. M., Staedler, Y. M., LaFountain, A. M., Schönenberger, J., et al. (2019). Emergence of a floral colour polymorphism by pollinator-mediated overdominance. *Nat. Commun.* 10, 63–11. doi: 10.1038/s41467-018-07936-x

Mant, J., Peakall, R., and Weston, P. H. (2005). Specific pollinator attraction and the diversification of sexually deceptive *Chiloglottis* (Orchidaceae). *Plant Syst. Evol.* 253, 185–200. doi: 10.1007/s00606-005-0308-6

Mant, J. G., Schiestl, F. P., Peakall, R., and Weston, P. H. (2002). A phylogenetic study of pollinator conservatism among sexually deceptive orchids. *Evolution* 56, 888–898. doi: 10.1111/j.0014-3820.2002.tb01402.x

Markham, K. R., Bloor, S. J., Nicholson, R., Rivera, R., Shemluck, M., Kevan, P. G., et al. (2004). Black flower coloration in wild *Lisianthius nigrescens*: its chemistry and ecological consequences. *Zeitschrift fur Naturforsch Sect C J. Biosci.* 59, 625–630. doi: 10.1515/znc-2004-9-1003

Miller, J. T., and Clements, M. A. (2014). Molecular phylogenetic analyses of Drakaeinae: Diurideae (Orchidaceae) based on DNA sequences of the internal transcribed spacer region. *Aust. Syst. Bot.* 27, 3–22. doi: 10.1071/SB13036

Peakall, R. (1990). Responses of male *Zaspilothynnus trilobatus* turner wasps to females and the sexually deceptive orchid it pollinates. *Funct. Ecol.* 4, 159–167. doi: 10.2307/2389335

Peakall, R., Ebert, D., Poldy, J., Barrow, R. A., Francke, W., Bower, C. C., et al. (2010). Pollinator specificity, floral odour chemistry and the phylogeny of Australian sexually deceptive *Chiloglottis* orchids: implications for pollinator-driven speciation. *New Phytol.* 188, 437–450. doi: 10.1111/j.1469-8137.2010.03308.x

Peakall, R., and Handel, S. N. (1993). Pollinators discriminate among floral heights of a sexually deceptive orchid: implications for selection. *Evolution* 47, 1681–1687. doi: 10.1111/j.1558-5646.1993.tb01260.x

Peakall, R., Wong, D. C. J., Phillips, R. D., Ruibal, M., Eyles, R., Rodriguez-Delgado, C., et al. (2021). A multitiered sequence capture strategy spanning broad evolutionary scales: application for phylogenetic and phylogeographic studies of orchids. *Mol. Ecol. Resour.* 21, 1118–1140. doi: 10.1111/1755-0998.13327

Policha, T., Davis, A., Barnadas, M., Dentinger, B. T. M., Raguso, R. A., and Roy, B. A. (2016). Disentangling visual and olfactory signals in mushroom-mimicking *Dracula* orchids using realistic three-dimensional printed flowers. *New Phytol.* 210, 1058–1071. doi: 10.1111/nph.13855

Schiestl, F. P. (2004). Floral evolution and pollinator mate choice in a sexually deceptive orchid. *J. Evol. Biol.* 17, 67–75. doi: 10.1046/j.1420-9101.2003.00650.x

Schiestl, F. P., Peakall, R., Mant, J. G., Ibarra, F., Schulz, C., Franke, S., et al. (2003). The chemistry of sexual deception in an orchid-wasp pollination system. *Science* 302, 437–438. doi: 10.1126/science.1087835

Thill, J., Miosic, S., Ahmed, R., Schlangen, K., Muster, G., Stich, K., et al. (2012). "Le Rouge et le noir": a decline in flavone formation correlates with the rare color of black dahlia (Dahlia variabilis hort.) flowers. *BMC Plant Biol.* 12:225. doi: 10.1186/1471-2229-12-225

Thomas, M. M., Rudall, P. J., Ellis, A. G., Savolainen, V., and Glover, B. J. (2009). Development of a complex floral trait: the pollinator-attracting petal spots of the beetle daisy, *Gorteria diffusa* (asteraceae). *Am. J. Bot.* 96, 2184–2196. doi: 10.3732/ajb.0900079

Trunschke, J., Lunau, K., Pyke, G. H., Ren, Z. X., and Wang, H. (2021). Flower color evolution and the evidence of pollinator-mediated selection. *Front. Plant Sci.* 12:617851. doi: 10.3389/fpls.2021.617851

van der Kooi, C. J., and Stavenga, D. G. (2019). Vividly coloured poppy flowers due to dense pigmentation and strong scattering in thin petals. *J. Comp. Physiol. A Neuroethol. Sensory Neural Behav. Physiol.* 205, 363–372. doi: 10.1007/s00359-018-01313-1

Vereecken, N. J., Wilson, C. A., Hotling, S., Schulz, S., Banketov, S. A., and Mardulyn, P. (2012). Pre-adaptations and the evolution of pollination by sexual deception: Cope's rule of specialization revisited. *Proc. R. Soc. B Biol. Sci.* 279, 4786–4794. doi: 10.1098/rspb.2012.1804

Vignolini, S., Davey, M. P., Bateman, R. M., Rudall, P. J., Moyroud, E., Tratt, J., et al. (2012). The mirror crack'd: both pigment and structure contribute to the glossy blue appearance of the mirror orchid, *Ophrys speculum*. *New Phytol.* 196, 1038–1047. doi: 10.1111/j.1469-8137.2012.04356.x

Vignolini, S., Moyroud, E., Hingant, T., Banks, H., Rudall, P. J., Steiner, U., et al. (2015). The flower of *Hibiscus trionum* is both visibly and measurably iridescent. *New Phytol.* 205, 97–101. doi: 10.1111/nph.12958

Weston, P. H., Perkins, A. J., Indsto, J. O., and Clements, M. A. (2014). "Phylogeny of Orchidaceae tribe Diurideae and its implications for the evolution of pollination systems," in *Darwin's orchids: then and now*. eds. R. Edens-Meier, P. Bernhardt (University of Chicago), 91–154.

Wong, D. C. J., Amarasinghe, R., Falara, V., Pichersky, E., and Peakall, R. (2019). Duplication and selection in β-ketoacyl-ACP synthase gene lineages in the sexually deceptive *Chiloglottis* (Orchidaceace). *Ann. Bot.* 123, 1053–1066. doi: 10.1093/aob/mcz013

Wong, D. C. J., Amarasinghe, R., Pichersky, E., and Peakall, R. (2018). Evidence for the involvement of fatty acid biosynthesis and degradation in the formation of insect sex pheromone-mimicking chiloglottones in sexually deceptive *Chiloglottis* orchids. *Front. Plant Sci.* 9:839. doi: 10.3389/fpls.2018.00839

Wong, D. C. J., Amarasinghe, R., Rodriguez-Delgado, C., Eyles, R., Pichersky, E., and Peakall, R. (2017). Tissue-specific floral transcriptome analysis of the sexually deceptive orchid *Chiloglottis trapeziformis* provides insights into the biosynthesis and regulation of its unique UV-B dependent floral volatile, chiloglottone 1. *Front. Plant Sci.* 8:1260. doi: 10.3389/fpls.2017.01260

Wong, D. C. J., and Peakall, R. (2022). Orchid phylotranscriptomics: the prospects of repurposing multi-tissue transcriptomes for phylogenetic analysis and beyond. *Front. Plant Sci.* 13:910362. doi: 10.3389/fpls.2022.910362

Wong, D. C. J., Perkins, J., and Peakall, R. (2022). Anthocyanin and flavonol glycoside metabolic pathways underpin floral color mimicry and contrast in a sexually deceptive orchid. *Front. Plant Sci.* 13:860997. doi: 10.3389/fpls.2022.860997

Zoratti, L., Sarala, M., Carvalho, E., Karppinen, K., Martens, S., Giongo, L., et al. (2014). Monochromatic light increases anthocyanin content during fruit development in bilberry. *BMC Plant Biol.* 14, 377–310. doi: 10.1186/s12870-014-0377-1

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 October 2022
DOI 10.3389/fpls.2022.1028245

OPEN ACCESS

EDITED BY
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

REVIEWED BY
Surendra Sarsaiya,
Zunyi Medical University, China
Pandiyan Muthuramalingam,
Gyeongsang National University, South Korea
Ali Raza,
Fujian Agriculture and Forestry University, China

*CORRESPONDENCE
Chunmei He
hechunmei2012@scbg.ac.cn

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

RECEIVED 25 August 2022
ACCEPTED 20 September 2022
PUBLISHED 06 October 2022

CITATION
Zhang M, Liu N, Teixeira da Silva JA, Liu X, Deng R, Yao Y, Duan J and He C (2022) Physiological and transcriptomic analysis uncovers salinity stress mechanisms in a facultative crassulacean acid metabolism plant *Dendrobium officinale*.
Front. Plant Sci. 13:1028245.
doi: 10.3389/fpls.2022.1028245

Physiological and transcriptomic analysis uncovers salinity stress mechanisms in a facultative crassulacean acid metabolism plant *Dendrobium officinale*

Mingze Zhang[1,2], Nan Liu[3], Jaime A. Teixeira da Silva[4], Xuncheng Liu[2], Rufang Deng[5], Yuxian Yao[1], Jun Duan[2,6] and Chunmei He[2]*

[1]The Department of Life Science and Agriculture, Qiannan Normal University for Nationalities, Duyun, China, [2]Key Laboratory of South China Agricultural Plant Molecular Analysis and Genetic Improvement, Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [3]Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, [4]Independent Researcher, Kagawa-ken, Japan, [5]Opening Public Laboratory, Chinese Academy of Sciences, Guangzhou, China, [6]Center of Economic Botany, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou, China

Dendrobium officinale is a precious medicinal Chinese herb that employs facultative crassulacean acid metabolism (CAM) and has a high degree of abiotic stress tolerance, but the molecular mechanism underlying the response of this orchid to abiotic stresses is poorly understood. In this study, we analyzed the root microstructure of *D. officinale* plantlets and verified the presence of chloroplasts by transmission electron microscopy. To obtain a more comprehensive overview of the molecular mechanism underlying their tolerance to abiotic stress, we performed whole-transcriptome sequencing of the roots of 10-month-old plantlets exposed to salt (NaCl) treatment in a time-course experiment (0, 4 and 12 h). The total of 7376 differentially expressed genes that were identified were grouped into three clusters ($P < 0.05$). Metabolic pathway analysis revealed that the expression of genes related to hormone (such as auxins, cytokinins, abscisic acid, ethylene and jasmonic acid) biosynthesis and response, as well as the expression of genes related to photosynthesis, amino acid and flavonoid metabolism, and the SOS pathway, were either up- or down-regulated after salt treatment. Additionally, we identified an up-regulated WRKY transcription factor, *DoWRKY69*, whose ectopic expression in Arabidopsis promoted seed germination under salt tress. Collectively, our findings provide a greater understanding of the salt stress response mechanisms in the roots of a facultative CAM plant. A number of candidate genes that were discovered may help plants to cope with salt stress when introduced *via* genetic engineering.

KEYWORDS

***Dendrobium officinale*, differentially expressed genes, metabolic adjustment, pigment biosynthesis, plant hormones signaling**

Introduction

Plants may face adverse environmental conditions, such as extreme temperatures, drought, and salinity are major abiotic stressors, that hamper their metabolism, delay growth and development, reduce productivity, or even cause plant death (Krasensky and Jonak, 2012; Raza et al., 2022a). High salinity strongly and negatively influences the productivity of agricultural crops by impacting seed germination and plant vegetative growth, and stimulates osmotic and oxidative stress, as well as ion toxicity (Deinlein et al., 2014).

However, under salt stress, plants can decrease ion toxicity and scavenge reactive oxygen species (ROS) by regulating ion homeostasis, induce antioxidant defense systems, and synthesize a variety of plant hormones and osmoprotectants (Raza et al., 2022a). Salinity can impact plants rapidly by inducing osmotic stress, or slowly by inducing ionic stress (Munns and Tester, 2008). In addition, salinity can disrupt the integrity of the chloroplast envelope, leading to the disorganization of grana, resulting in a reduced photosynthetic rate that can be partly attributed to the reduction of photosynthetic pigments (Kwon et al., 2019). To allow plants to cope with salt stress, signal transduction pathways such as phytohormone-mediated and Ca^{2+} signaling pathways are activated (Knight et al., 1997; Zhu, 2002; Kaleem et al., 2018). Under high salinity stress, plants modify their levels of endogenous hormones such as gibberellins (GAs), abscisic acid (ABA) and jasmonic acid (JA), thereby activating the genes involved in the metabolism of those hormones to cope with this stress (Kaleem et al., 2018). GA is involved in salt stress response pathways by negatively regulating a class of DELLA proteins (Verma et al., 2016). ABA promotes stomatal closure to reduce water loss by rapidly altering ion fluxes in guard cells under osmotic stress (Munemasa et al., 2015). Most studies have illustrated the role of JA in response to biotic stresses, while evidence of the involvement of JA in salinity stress has only emerged over the last decade (Kazan, 2015; Wasternack, 2015; Abouelsaad and Renault, 2018). In addition, auxins and cytokinins mediate stress-adaptation responses in plants (Bielach et al., 2017).

Some studies have demonstrated that hormone-mediated signal transduction pathways trigger transcription factors (TFs) such as members of the bZIP, NAC, ERF, WRKY, MYB, ZF-HD, and bHLH families to amplify signals by inducing or repressing functional genes, thereby initiating protective mechanisms that allow plants to cope with salinity stress (Deinlein et al., 2014; Kaleem et al., 2018; Shalmani et al., 2019; Mansour and Hassan, 2022). For example, the dehydration-responsive element binding/C-repeat-binding factor (DREB/CBF), which is a member of the AP2/ERF TF subfamily, can modulate drought-, high temperature-, salt-, and cold-responsive gene expression (Dubouzet et al., 2003). WRKY TFs are involved in the response of plants to salinity stress. In *Arabidopsis thaliana*, *AtWRKY46* plays a role in regulating stomatal opening, with mutant *wrky46* plants showing sensitivity to drought and salt stress (Ding et al., 2015). The interaction between GmWRKY27 and GmMYB174 proteins from *Glycine max* regulates *GmNAC29* expression, improving salt and drought tolerance (Wang et al., 2015b). *CgWRKY57*, which was identified in an orchid (*Cymbidium goeringii*), sensed ABA signals and might play a role in stress response (Liu et al., 2021). Overexpression of *WRKY* genes from *Dendrobium nobile* enhanced salt and stress tolerance in transgenic tobacco (Xu et al., 2014). Compatible solutes, such as proline, glycine betaine, sugars, mannitol and polyols, may accumulate in the cytoplasm, allowing water potential to be adjusted and cell turgor to be maintained under salinity stress (Aziz and Larher, 1995; Krasensky and Jonak, 2012; Moharramnejad et al., 2015). In *Medicago sativa*, most *MsZF-HDs* responded to salt stress (He et al., 2022). In *A. thaliana*, MYC2, which is a bHLH TF, negatively regulated proline biosynthesis by repressing *P5CS1* expression in response to salt stress (Verma et al., 2020). Moreover, an R2R3-MYB TF-encoding gene *IbMYB308* from sweet potato (*Ipomoea batatas*) improve the salt tolerance of transgenic tobacco (Wang et al., 2022).

The Orchidaceae is the most diverse flowering plant family and includes approximately 880 genera and 28,000 species around the world (Givnish et al., 2015). *Dendrobium*, which is the second largest genus in the Orchidaceae, has adapted to diverse natural habitats, ranging from high altitudes to lowland tropical forests, and even to the dry climate of the Australian desert, implying its strong adaptability to adverse environmental conditions. *Dendrobium* plants are known to survive under high salinity (15 dS m^{-1}) (Abdullakasim et al., 2018), indicating that they display high endurance to salinity stress. *Dendrobium officinale* is a precious Chinese medicinal herbal orchid with various bioactivities (Teixeira da Silva and Ng, 2017). A previous study by our group showed that *D. officinale* could adapt to high concentrations of salt stress, and in doing so, accumulated bioactive metabolites (Zhang et al., 2021).

Although a tremendous amount of research has been conducted on salt stress in model plants and crops, known strategies for coping with salt stress in orchids is not commensurate with their recalcitrant nature. Understanding the mechanisms by which orchid plants cope with the challenge of salt stress will pave a way for planning future initiatives for abiotic (salt) stress engineering in orchids to protect orchid plants in the wild. In this study, we investigated changes to the transcriptome under short-term (within 12 h) salinity stress in the roots of *D. officinale* plantlets. We identified differentially expressed genes (DEGs) and focused on phytohormone signaling, the SOS pathway, photosynthesis and metabolic adjustments such as changes in amino acid content and flavonoid metabolism. We also identified differentially expressed TFs and characterized the role of a single *WRKY* gene in seed germination under salinity stress. Our results provide new insight into the biochemical and

molecular regulation of salt stress tolerance mechanisms of orchid plants that enhance their stress tolerance.

Materials and methods

Plant growth conditions and treatments

D. officinale plantlets that were geminated from seeds within the same capsule, were grown *in vitro* on half-strength Murashige and Skoog (MS) (Murashige and Skoog, 1962) medium containing 2% sucrose and 0.6% agar (pH 5.4) in a growth chamber (25 ± 1°C, 40 μmol m^{-2} s^{-1}, 12-h photoperiod). Uniform plantlets 8-9 cm in height were transferred to half-strength MS containing 2% sucrose and 0.6% agar, and supplemented with 250 mM NaCl (pH 5.4) and kept at 25 ± 1°C, under 40 μmol m^{-2} s^{-1}, in a 12-h photoperiod. Root samples were collected at three time points [0 (control), 4 and 12 h], frozen in liquid nitrogen and used to isolate RNA. Three independent biological replicates were used for each time point and six plantlets were used for each replicate. To analyze chlorophyll (Chl) and carotenoids, plantlets were subjected to salinity treatment or not (the control) for two weeks, and roots were harvested and used to detect these pigments. To analyze total flavonoids and free amino acids, roots were collected after two weeks from control and salt-treated plants and freeze dried. Root samples were ground to a fine powder by a RETSCH MM400 Mixer Mill (Retsch Technology, Haan, Germany). Three independent biological replicates were utilized for each sample.

Histological analysis of *D. officinale* plantlet roots

The roots of plantlets 8-9 cm in height were collected and fixed in fixation buffer [2.5% glutaraldehyde and 2% paraformaldehyde in 0.1 M sodium phosphate buffer (pH 7.2)]. To facilitate penetration of the fixative, samples were vacuum infiltrated for at least 30 min. Root samples were rinsed in wash buffer (1% sodium phosphate) six times (30 min each time) after fixation, then fixed in a 0.1 M sodium cacodylate buffer (pH 7.2) containing 1% osmium tetroxide (OsO_4) for 4 h. An ethanol concentration gradient (30, 50, 75, 85, 95, 100%, v/v) was used to dehydrate samples. After dehydration, samples were treated in different ratios of acetone and Epon812 (3:1, 1:1 and 1:3, v/v) for 30 min. Finally, samples were embedded in absolute Epon812 overnight and placed at 60°C for two days. Transverse sections of root samples were cut into slices of 1 μm thickness with an LKB-11800 ultramicrotome and stained by periodic acid Schiff (PAS) (Tütüncü Konyar et al., 2013). Root samples were cut into 50-70 nm thick slices with a LeicaUC6 ultramicrotome for observations using a JEOL JEM 1010 (JEOL Ltd., Tokyo, Japan) transmission electron microscope.

Transcriptomic analysis

Total RNA from root samples at 0 (control), 4 and 12 h after salinity stress was isolated using Column Plant RNAout2.0 (Tiandz Inc., Beijing, China). Biological triplicates for each time point were used for RNA sequence (RNA-seq) analysis in this study. The mRNA of each sample was purified from total RNA using oligo $d(T)_{25}$ magnetic beads (New England BioLabs Inc., Ipswich, MA, USA). A library of the isolated mRNA was prepared with the NEBNext® Ultra™ RNA Library Prep Kit (New England Biolabs Inc.) then subjected to paired-end sequencing with the Illumina Novaseq 6000 Sequencing System at Biomarker Technologies Inc. (Beijing, China). The raw reads produced from sequencing were processed through in-house perl scripts (Biomarker Technologies Inc.) to remove reads containing an adapter, ploy-N or poor-quality reads (the quality Q ≤ 30 accounted for more than 50% of all reads). The remaining reads were clean reads, which were mapped with the *D. officinale* version 2 genome generated by Zhang et al. (2017) using TopHat version 2.0.8 (Kim et al., 2013a). About 90% of clean reads were mapped to the *D. officinale* genome. We performed differential expression analysis of RNA-seq data by comparing the expression of genes in treatments (4 h and 12 h) and the control (0 h) by using DESeq2 (Love et al., 2014). Genes with a fold change (treatment/control) ≥ 2.0 and a false discovery rate (FDR) ≤ 0.01 were regarded as up-regulated genes while genes with a fold change (treatment/control) ≤ 0.5 and an FDR ≤ 0.01 were regarded as down-regulated genes. All the up- and down-regulated genes were defined as DEGs. The clean reads generated in this study were submitted to the Sequence Read Archive (SRA) database of the National Center for Biotechnology (NCBI) under the following accession numbers: SRS8480001, SRS8480000 and SRS8480010 for 0 h; SRS8480011, SRS8480012 and SRS8480013 for 4 h; SRS8480016, SRS8480004 and SRS8480002 for 12 h.

Functional annotation of genes

Seven public gene functional annotation databases, NCBI non-redundant protein sequences database (Nr, http://www.ncbi.nlm.nih.gov), Protein family (Pfam, http://pfam.xfam.org/) (Finn et al., 2013), Clusters of Orthologous Groups of proteins (COG, http://www.ncbi.nlm.nih.gov/COG) (Galperin et al., 2021), a manually annotated and reviewed protein sequence database (Swiss-Prot, https://www.expasy.org/) (The UniProt Consortium, 2017), Gene Ontology (GO, http://geneontology.org/) (Ashburner et al., 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.kegg.jp/) (Kanehisa et al., 2004)

and evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG, http://eggnog.embl.de/) (Huerta-Cepas et al., 2019), were used to annotate gene functions. We identified genes based on functional gene annotation. Information about the metabolic pathway of genes was obtained from KEGG annotation. Based on GO annotation, the number of DEGs assigned to each GO term was calculated and visualized in a diagram. To cluster the DEGs, the Short Time-series Expression Miner (STEM) method in STEM software was used (Ernst and Bar-Joseph, 2006). The maximum number of model profiles was 20, and the maximum unit change in model profiles between time points was 1. KOBAS software (Xie et al., 2011) was used to test the statistical enrichment of DEGs in pathways.

Reverse transcription PCR and quantitative real-time RT-PCR

Eight-day-old seedlings of wild-type Arabidopsis (Col-0) and transgenic lines were grown in half-strength MS medium supplemented with 1.5% sucrose, 0.8% agar (pH 5.7) and placed under a 16-h photoperiod (100 μmol m^{-2} s^{-1}) at 22°C. Total RNA was isolated using Column Plant RNAout2.0 (Tiandz Inc.). RNA samples were used for first-strand cDNA synthesis with the GoScript Reverse Transcription System (Promega, Madison, WI, USA). The cDNA of each sample (400 ng μL^{-1}) was used as a template for PCR amplification. The genes were amplified by 30 cycles of 98°C for 10 sec, 55°C for 30 sec and 72°C for 30 sec. PCR products (5 mL) of each sample were surveyed by 1% agarose gel electrophoresis and photographed with a gel imaging system (GenoSens1880, Shanghai Qinxiang Scientific Instrument Co. Ltd., Shanghai, China). Quantitative real-time PCR (qRT-PCR) was performed with the Unique Aptamer™ qPCR SYBR® Green Master Mix (Beijing Novogene Technology Co. Ltd.) in a LightCycler 480 system (Roche, Basel, Switzerland). Primer sets are listed in Supplementary Table 1.

Determination of chlorophyll and total carotenoids in roots

Fresh root samples were homogenized in a mortar with a pestle using silica sand and 80% acetone as the extracting solvent. The homogenized mixture was placed at 4°C for 1 h in the dark before centrifuging at 10,000 rpm for 15 min at 4°C. The supernatant was collected and used immediately to detect absorption at 663.2, 646.8 and 470 nm with a UV-6000 spectrophotometer (Shanghai Metash, Shanghai, China). Chl *a* was calculated as $12.25A_{663.2} - 2.79A_{646.8}$, Chl *b* was calculated as $21.5A_{646.8} - 5.1A_{663.2}$, and total carotenoids was calculated as (1000 A_{470} – 1.82 Chl *a* – 85.02 Chl *b*)/198 (Sumanta et al., 2014).

Amino acid quantification

Root powder (100 mg) was solubilized in 5 mL of 0.01 M HCl. After incubating for 30 min at room temperature, samples were centrifuged at 13,000 rpm for 10 min. The supernatant was mixed with absolute ethyl alcohol (20:80, v/v). The mixture was incubated for 15 min at room temperature, then centrifuged at 13,000 g for 10 min. The extract was dried by evaporation under vacuum by a rotary evaporator (Eyela N-1300V-W, Tokyo Rikakikai Co. Ltd., Tokyo, Japan). HCl (0.01 M) was added to dissolve the isolated free amino acids. After filtering through a 0.22 μm filter membrane (Corning Inc., Corning, NY, USA), amino acid profiles were quantified using an Automatic Amino Acid Analyzer (S 433-D, Sykam GmbH, Eresing, Germany). Ninhydrin-derivatized amino acids were measured at 570 nm and at 440 nm. Amino acid concentrations were reported as mg of amino acid per 100 mg of dry weight. A standard stock solution type pH (for physiological amino acid analysis) (Sykam catalog No. S000031, Sykam GmbH) was used as the standard solution.

Determination of total flavonoids

A colorimetric method described by Ren et al. (2020) was used to analyze total flavonoid content with rutin solutions (4, 8, 12, 16 and 20 μg/mL) serving as standards. Briefly, powdered samples were extracted with 50% (v/v) methanol in an ultrasonication bath (VCX600, Sonics and Materials Inc., Newtown, CT, USA) for 90 min at room temperature, then centrifuged at 12,000 rpm for 20 min. The supernatant was collected and used to measure absorbance at 360 nm with a UV-6000 spectrophotometer. The calibration standard was 50% (v/v) methanol.

Generation of transgenic plants and germination assay

The *DoWRKY69* gene was isolated and inserted into the *Nco*I site of the pCAMBIA1302 vector. The validated recombinant vector was transformed into *Agrobacterium tumefaciens* EHA105 (Shanghai Weidi Biotechnology Co. Ltd, Shanghai, China) by the freeze-thaw method (Weigel and Glazebrook, 2006) and then transformed into *A. thaliana* by the floral dip method (Clough and Bent, 1998). Seeds of both wild type (WT) and transgenic lines were surface-disinfected then seeded on half-strength MS medium containing 1.5% sucrose, 0.8% agar (pH 5.7), and different concentrations of NaCl (150, 200, and 250 mM). Medium without NaCl served as the control. Seeds were kept at 4°C in the dark for 2 d, then transferred to a 16-h photoperiod (100 μmol m^{-2} s^{-1}) at 22°C.

When the radicle emerged from the testa, a seed was considered to have germinated. About 60 seeds of each genotype were used and the entire experiment was repeated in triplicate.

Statistical analyses

Statistical analyses were performed with SigmaPlot12.5 software (Systat Software Inc., San Jose, CA, USA). The Dunnett test ($P < 0.05$) was used to indicate statistically significant differences.

Results

Anatomical traits of *D. officinale* roots

The roots of *D. officinale* plantlets were green and possessed meristematic and elongation zones (Figure 1A). The differentiation zone could not clearly or easily be distinguished from the meristematic and elongation zones because root hairs were absent (Figure 1A). To better understand the anatomical traits of *D. officinale* plantlet roots, we performed histological analysis to investigate the microstructure in transverse sections. The root consists of a velamen, cortex and stele, the largest proportion consisting of the cortex and a smaller stele (Figure 1B). The cortex is composed of an epidermis, an exodermis and a mass of cortical parenchyma tissue (Figure 1B). Starch granules (purple spots) were observed in the cortical parenchyma tissue but were absent in the stele and velamen (Figure 1B). Transmission electron microscopic observations revealed irregularly shaped starch granules that were easy to distinguish (Figures 1C, D). The stele cell contained a large (about 10 μm) nucleus and nucleolus, as well as small chloroplasts (Figure 1E). The thylakoids were stacked, similar to grana in leaves (Figure 1F). These observations demonstrate that *D. officinale* plantlet roots contain chloroplasts in their stele.

Salinity induces multiple cellular events in roots of *D. officinale* plantlets

In order to reveal global transcriptional dynamics in the roots of *D. officinale in vitro* plantlets exposed to salinity stress, we performed a time course transcriptomic analysis at 0, 4 and 12 h after exposure to 250 mM NaCl. In this study, total clean bases of each library generated from sequencing amounted to more than 6 Gb (Supplementary Table 2). The percentage of bases of each library having a quality score of 30 or higher exceeded 93% (Supplementary Table 2). We verified the expression patterns of 12 selected genes by qRT-PCR (Supplementary Figure 1). These were in agreement with the changes in fragments per kilobase per million (FPKM) values, supporting the reliability of our RNA-seq data. A total of 7376 DEGs were identified in the 0 h *vs* 4 h and 0 h *vs* 12 h comparisons. DEGs from the 0 h *vs* 4 h comparison consisted of 2282 down-regulated genes and 2844 up-regulated genes, while the number of DEGs from the 0 h *vs* 12 h comparison consisted of 2877 down-regulated genes and 3236 up-regulated genes (Figure 2A). Among all DEGs, 3863 genes showed a differential expression pattern at both time points: 2155 genes were up-regulated and 1702 genes were down-regulated (Figure 2B). We first performed a GO classification analysis of all the DEGs: 2538 and 3098 DEGs were assigned to three major GO categories (biological processes, cellular components, and molecular functions) in the 0 h *vs* 4 h and 0 h *vs* 12 h comparisons, respectively (Figure 2C). In the biological processes category, 'metabolic process' contained the greatest number of DEGs in both comparisons (Figure 2C). In the molecular function category, most DEGs were assigned to 'catalytic activity' in both comparisons (Figure 2C). To decipher the general trends of gene expression profiles, all DEGs were subjected to STEM clustering analysis (Figure 2D). These results suggest that salinity stress triggers a series of biological processes and molecular functions that mainly affect metabolism and defense in *D. officinale* roots.

Auxin, cytokinin, ABA, ethylene and JA are involved in salinity response

To analyze the relationship between plant hormones and salinity response in *D. officinale* roots, we investigated the genes related to the biosynthesis of plant hormones under salinity stress using RNA-seq data. The indole-3-acetic acid (IAA) biosynthetic pathway genes *TAA1* and *YUC2* showed a down-regulated expression pattern after salinity stress (Figure 3A). The genes involved in zeatin biosynthesis showed a similar expression pattern as the IAA biosynthetic pathway genes. For example, the expression of three *IPT* genes (encoding ADP/ATP-dependent enzymes, isopentenyltransferases) and one cytokinin *trans*-hydroxylase gene *CYP735A* was repressed after salinity stress (Figure 3B). As expected, most of the genes involved in ABA, ethylene and JA biosynthesis were induced by salinity stress (Figures 3C-E). The *NCED* genes, which encode a key enzyme (9-*cis*-epoxycarotenoid dioxygenase) for ABA biosynthesis, were induced by salinity stress. The expression of one *NCED* gene increased at 4 h but dropped to its initial level at 24 h; another *NCED* gene was detected strongly at both 4 h and 12 h (more than 85-fold increase in expression) (Figure 3C). The DEGs related to ethylene biosynthesis increased at both time points after salinity treatment (Figure 3D). In total, we found that among the 14 DEGs related to JA biosynthesis, 13 were up-regulated at least at one time point while only one gene *AOC* was down-regulated at both time points (Figure 3E).

10.3389/fpls.2022.1028245

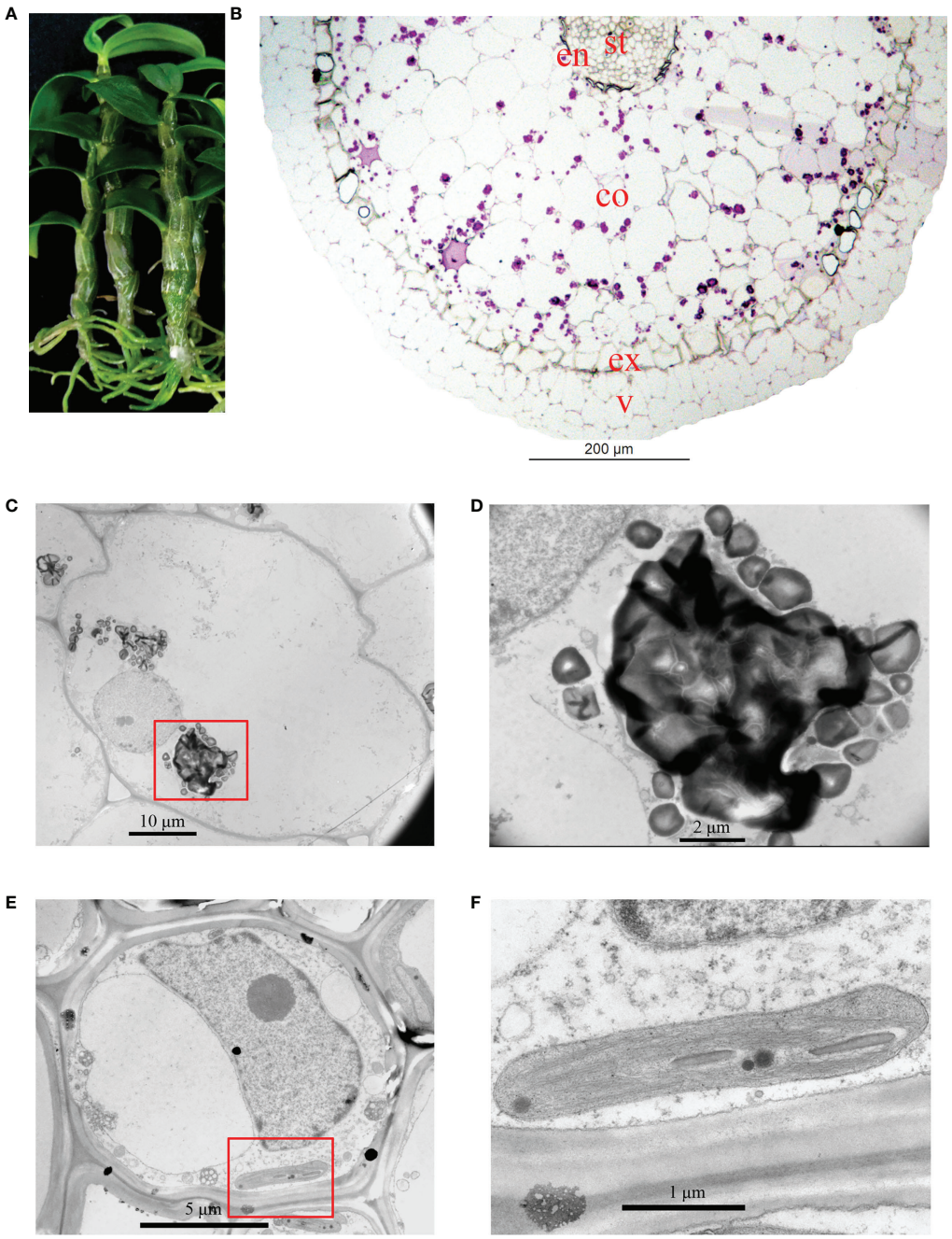


FIGURE 1
Analysis of the structure of *in vitro Dendrobium officinale* plantlet roots. **(A)** The roots of plantlets are green (chlorophyllous). **(B)** Overview of the anatomical traits of roots in a transverse section. Root tissue was stained by PAS. St, stele; en, endodermis; co, cortex; ex, exodermis; v, velamen. **(C)** Transmission electron microscopic observations of cortex cell morphology. **(D)** Enlarged view of starch granules from **(C)**. **(E)** Morphology of a stele cell. **(F)** Enlarged view of a chloroplast from **(E)**.

We then analyzed the auxin, cytokinin, ABA, ethylene and JA signal transduction pathway genes. AUX1 is a transmembrane amino acid transporter family protein that is involved in an early step of auxin signaling. Two *AUX1* homologs found in *D. officinale* were down-regulated after salinity treatment (Figure 4A). Our data showed the down-regulation of auxin responsive factors (ARFs) in response to salt stress (Figure 4A). In addition, auxin-responsive genes coding for proteins such as auxin/indole acetic acid (AUX/IAA), GH3, and small auxin-up RNA (SAUR) were differentially expressed

10.3389/fpls.2022.1028245

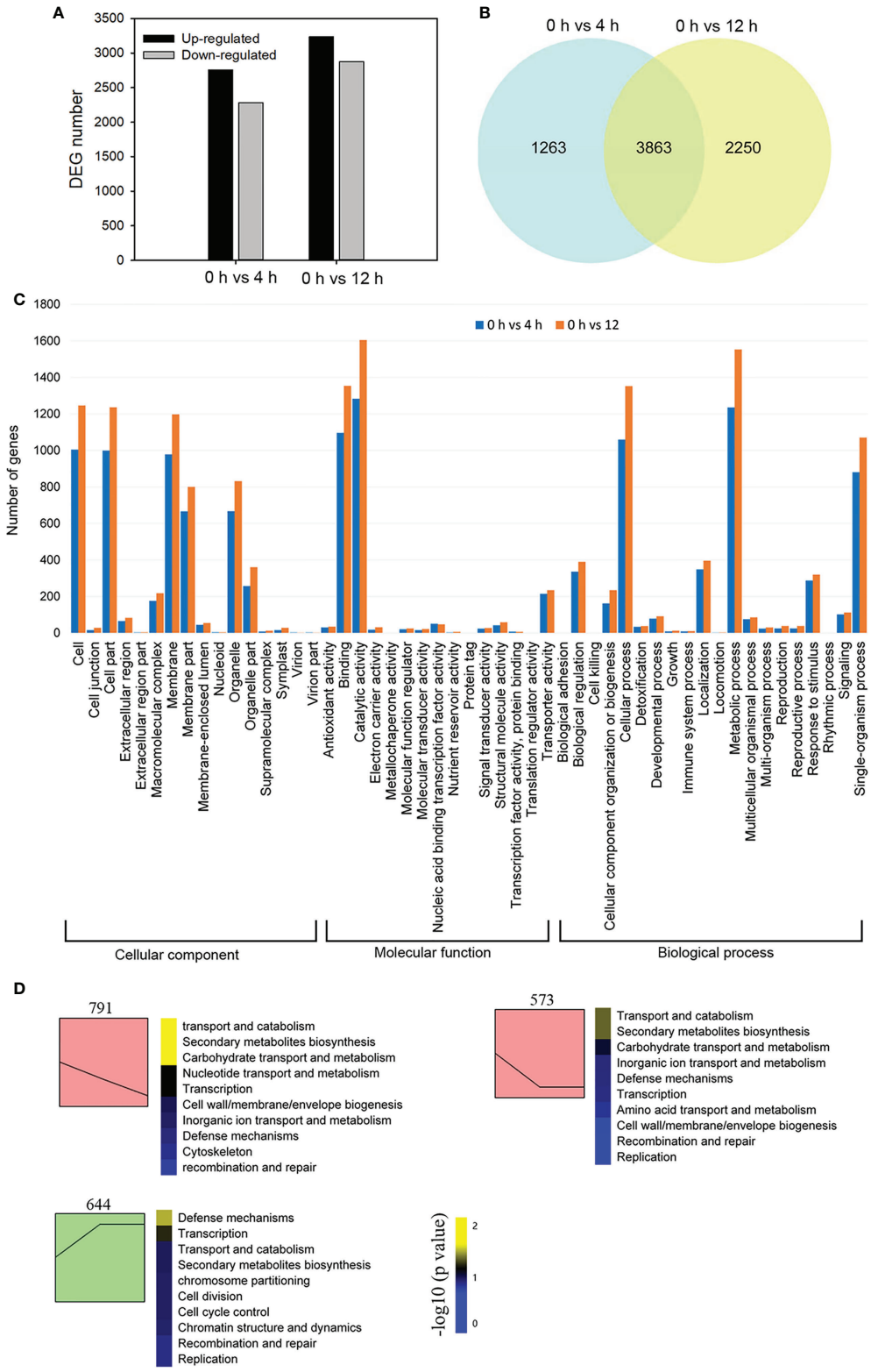


FIGURE 2
Transcriptome sequencing analysis of *Dendrobium officinale* plantlet roots after salt treatment. **(A)** Analysis of differentially expressed genes (DEGs) in response to salt tress. DEGs were generated by 0 h *vs* 4 h and 0 h *vs* 12 h comparisons. **(B)** Venn diagram showing shared DEGs from the 4 h and 12 h comparison after salt stress. **(C)** Gene ontology analysis of DEGs. **(D)** All DEGs were subjected to STEM clustering analysis.

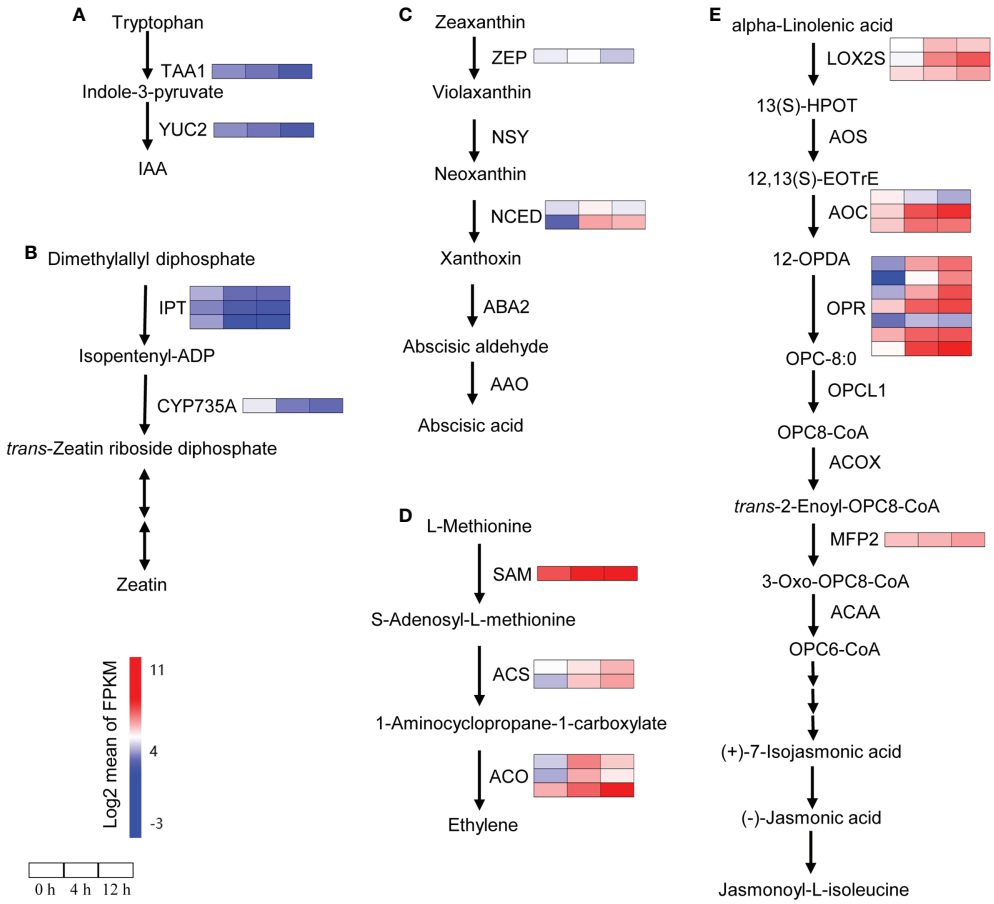


FIGURE 3
Changes in the expression of biosynthetic genes encoding phytohormones after salt stress. **(A)** Expression analysis of IAA biosynthetic genes. IAA, indole-3-acetic acid; TAA, tryptophan aminotransferase of Arabidopsis; YUC, indole-3-pyruvate monooxygenase 2. **(B)** Expression analysis of zeatin biosynthetic genes. IPT, isopentenyltransferase; CYP735A, cytokinin *trans*-hydroxylase. **(C)** Expression analysis of ABA biosynthetic genes. ABA, abscisic acid; ZEP, zeaxanthin epoxidase; NSY, neoxanthin synthase; NCED, 9-*cis*-epoxycarotenoid dioxygenase; ABA2, xanthoxin dehydrogenase; AAO, abscisic-aldehyde oxidase. **(D)** Expression analysis of ethylene biosynthetic genes. SAM, S-adenosylmethionine synthetase; ACS, 1-aminocyclopropane-1-carboxylate synthase; ACO, aminocyclopropanecarboxylate oxidase. **(E)** Expression analysis of JA biosynthetic genes. LOX2S, lipoxygenase; AOS, hydroperoxide dehydratase; AOC, allene oxide cyclase; OPR, 12-oxophytodienoic acid reductase; OPCL1, OPC-8:0 CoA ligase 1; ACX, acyl-CoA oxidase; MFP2, enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase 2; ACAA, acetyl-CoA acyltransferase. Genes that changed significantly at least one time point after salt treatment are shown. Red indicates high expression and blue indicates low expression.

after salinity stress, suggesting the importance of auxin in the salt stress response. For the cytokinin signal transduction pathway, signal transduction occurs *via* phosphotransfer between the sensor kinase and the receiver domain of the response regulator (Kieber and Schaller, 2018). In the cytokinin signal transduction pathway, *HK2/3*, which encodes a histidine kinase and *AHP*, which encodes a histidine-containing phosphotransfer factor, were down-regulated after salinity stress at both time points (Figure 4B). There are two types of response regulators (B-ARRs and A-ARRs) involved in cytokinin signaling (Kieber and Schaller, 2018). Type-B ARRs directly activate type-A ARRs in response to cytokinin (Kieber and Schaller, 2018). Four *B-ARR* genes were up-regulated and one was down-regulated after salt stress (Figure 4B). One *A-ARR* gene was repressed by salinity stress (Figure 4B). In addition, our transcriptomic analysis showed that more that 50% of ABA-, ethylene- and JA-responsive genes were up-regulated after salt stress (Figures 4C-E). These results indicate that the biosynthesis and signal transduction of auxin and cytokinin were repressed while the biosynthesis and signal transduction of ABA, ethylene and JA were activated in response to salinity stress, suggesting that auxin, cytokinin, ABA, ethylene and JA may be required for the response of *D. officinale* roots to salinity.

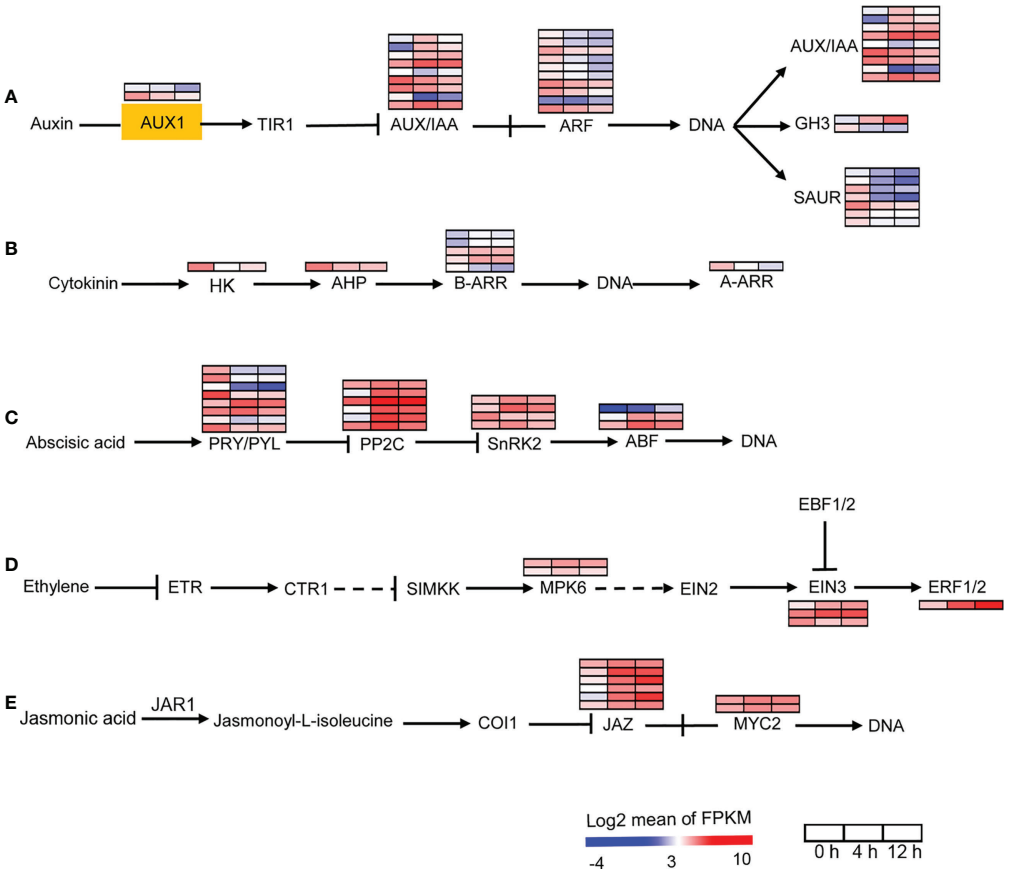


FIGURE 4
Changes in the expression of plant hormone signal transduction pathway under salt treatment. **(A)** Expression analysis of auxin signal transduction pathway genes. AUX1, auxin influx carrier 1; TIR1, transport inhibitor response 1; Aux/IAA, auxin/indole acetic acid protein; ARF, auxin response factor; GH3, Gretchen Hagen; SAUR, small auxin-up RNAs. **(B)** Expression analysis of cytokinin signal transduction pathway genes. HK, histidine kinase; AHP, histidine-containing phosphotransfer protein; A-ARR, type-A Arabidopsis response regulator; B-ARR, type-B Arabidopsis response regulator. **(C)** Expression analysis of ABA signal transduction pathway genes. PRY/PRL, abscisic acid receptor PYR/PYL family; PPC2C, protein phosphatase-2C; SnRK2, SNF1-related protein kinase 2, ABF, ABA responsive element binding factor. **(D)** Expression analysis of ethylene signal transduction pathway genes. ETR, ethylene receptor; CTR1, serine/threonine-protein kinase 1; SIMKK, mitogen-activated protein kinase kinase; MPK6, mitogen-activated protein kinase 6; EIN2/3, ethylene-insensitive protein 2/3; EBF1/2, EIN3-binding F-box protein 1/2; ERF1/2, ethylene-responsive transcription factor 1/2. **(E)** Expression analysis of JA signal transduction pathway genes. JAR1, jasmonic acid-amino synthetase; COI1, coronatine-insensitive protein 1, MYC2, transcription factor MYC2. The pathways were redraw based on 'Plant hormone signal transduction' in the KEGG database (https://www.kegg.jp/). Genes that changed significantly in at least one time point after salt treatment are shown.

The SOS pathway was activated in response to salinity

In addition to the plant hormone signaling transduction pathway, another plant signaling pathway – the calcium signaling pathway – is also involved in the salt stress response. The SOS signaling network is activated by Ca^{2+} signaling (Kaleem et al., 2018). Three genes *Salt Overly Sensitive 1-3* (*SOS1-3*) are require for salt tolerance in plants (Ji et al., 2013). SOS3, which is a calcineurin B-like protein that serves as a Ca^{2+}-binding protein, transduces the signal downstream by sensing changes in Ca^{2+} concentration in the cytoplasm (Tuteja, 2007). As expected, the *SOS3* gene was rapidly up-regulated at 4 h, but decreased at 12 h relative to 4 h (Figure 5). *SOS2* (also known as *AtCIPK24*) encodes a CBL-interacting serine/threonine-protein kinase (CIPK), which interacts with the SOS3 protein to form the SOS3–SOS2 protein kinase complex (Tuteja, 2007). Only one CIPK DEG, annotated as *CIPK24*, was found in this study, and its expression was repressed after salinity stress treatment (Figure 5). The *SOS1* gene encodes a Na^+/H^+ antiporter that results in a low concentration of cytoplasmic Na^+ ions by enabling an efflux of excess Na^+ ions across the plasma membrane. The vacuolar Na^+/H^+ exchanger 1 (NHX1) can reduce cytoplasmic Na^+ by transferring cytoplasmic

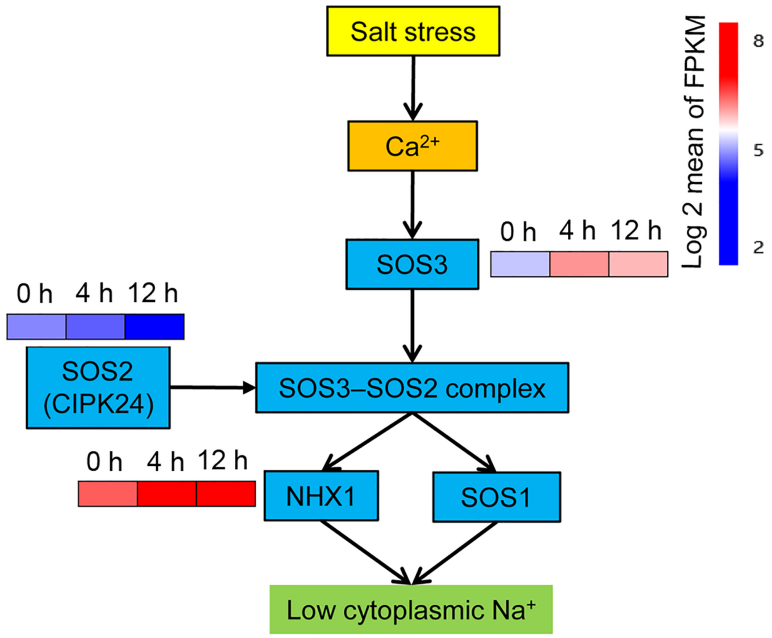


FIGURE 5
Expression analysis of SOS pathway genes. SOS, salt overly sensitive. The pathways were redrawn according to Tuteja (2007). Genes that changed significantly in at least one time point after salt treatment are shown.

Na^+ into vacuoles and maintaining osmotic balance in vacuoles (Apse et al., 1999; Bhaskaran and Savithramma, 2011). The SOS3–SOS2 protein kinase complex activates both SOS1 and NHX to trigger an Na^+ exclusion response (Tuteja, 2007). No differential expression of the *SOS1* gene was observed in this study, while the expression of one gene encoding a vacuolar Na^+/H^+ exchanger was up-regulated after salt stress, with 2.4-fold and 2.2-fold greater expression than the control at 4 h and 12 h, respectively (Figure 5).

Photosynthetic pigments, photosynthesis and carbon assimilation were impaired by salinity stress

When salt concentration increases, this results in osmotic stress and the absorption of more Na^+ and Cl^- by roots, which may negatively affect plant growth by decreasing photosynthetic efficiency (Deinlein et al., 2014). To reveal the impact of photosynthesis in *D. officinale* roots in response to salinity stress, we first analyzed photosynthesis-related biosynthetic genes and pigments. After examining the transcription profiles of Chl biosynthesis and catabolic genes, we found that all the Chl biosynthesis genes, except for one chlorophyll(ide) *b* reductase gene *NYC1* and one chlorophyllase gene *CLH2*, the expression of all other genes was down-regulated (Figure 6A). For example, the protochlorophyllide reductase gene *POR* was strongly expressed at 0 h with an average FPKM value > 750, but its expression dropped rapidly after salt stress treatment, with average FPKM values of 137 and 46 at 4 and 12 h, respectively (Figure 6A). In addition, the pheophorbide a oxygenase gene *PAO*, which encodes a key enzyme in the catabolism of Chl, was considerably up-regulated (with an average FPKM of about 90 at 0 h in contrast to an average FPKM value of > 500 at both 4 and 12 h) (Figure 6A). These results suggest that the decrease in Chl biosynthesis and reduced Chl degradation led to a decrease in Chl content in roots under salinity stress. In the carotenoid biosynthetic pathway, all of the identified carotenoid biosynthetic genes, except for *CRTISO2*, showed lowest expression levels at 12 h, but some of these genes were up-regulated at 4 h in response to salinity stress (Figure 6B). For example, the β-carotene 3-hydroxylase gene *CRTZ* had a mean FPKM value of 11 at 0 h, 57 at 4 h and 7 at 12 h (Figure 6B). However, Chl *a*, Chl *b* and total carotenoid content were not different between the control and salinity stress at 24 h (Supplementary Figure 2), but pigment content, especially Chl *a*, decreased significantly after two weeks' exposure to salt stress (Figure 6C). These results indicate that photosynthetic pigments were reduced after salt stress, even a long time after exposure. We also isolated the DEGs that encode photosynthetic proteins, including those associated with the photosystem I complex, light harvesting complex I, photosystem II complex, light harvesting

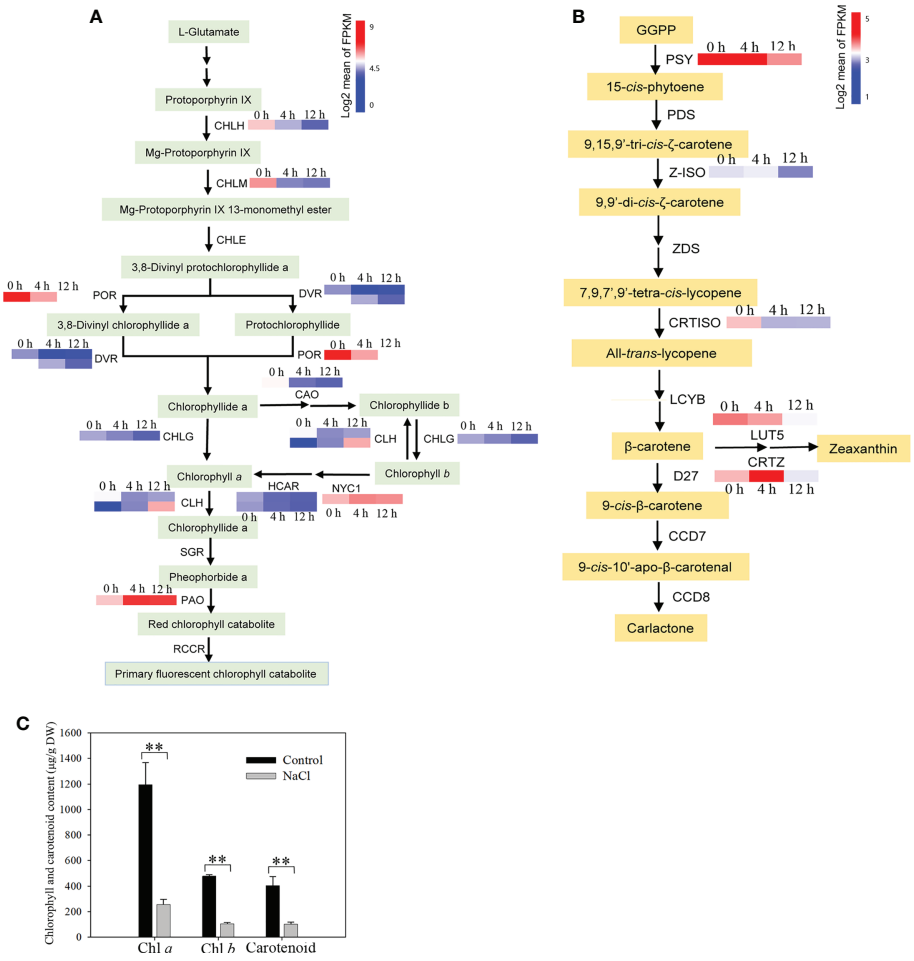


FIGURE 6
Analysis of the content of photosynthetic pigments and their metabolic pathway genes after salt treatment. **(A)** Analysis of expression of genes related to the Chl metabolic pathway. CHLH, magnesium chelatase subunit H; CHLM, magnesium-protoporphyrin *O*-methyltransferase; CHLE, magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase; POR, protochlorophyllide reductase; DVR, divinyl chlorophyllide *a* 8-vinyl-reductase; CHLG, chlorophyll/bacteriochlorophyll *a* synthase; CAO, chlorophyllide *a* oxygenase; CLH, chlorophyllase; NYC1, chlorophyll(ide) *b* reductase 1; HCAR, 7-hydroxymethyl chlorophyll *a* reductase; SGR, magnesium dechelatase; PAO, pheophorbide *a* oxygenase; RCCR, red chlorophyll catabolite reductase. **(B)** Changes in carotenoid biosynthetic pathway genes in response to salt stress. GGPP, geranylgeranyl diphosphate; PSY, 15-*cis*-phytoene synthase; PDS, 15-*cis*-phytoene desaturase; Z-ISO, ζ-carotene isomerase; ZDS, ζ-carotene desaturase; CRTISO, carotene *cis*-*trans* isomerase; LCYB, lycopene β-cyclase; LUT5, β-ring hydroxylase; CRTZ, β-carotene 3-hydroxylase; D27, β-carotene isomerase; CCD7, 9-*cis*-β-carotene 9′,10′-cleaving dioxygenase; CCD8, carlactone synthase/all-*trans*-10′-apo-β-carotenal 13,14-cleaving dioxygenase. The Chl metabolic pathway and carotenoid biosynthetic pathway were redrawn based on 'porphyrin and chlorophyll metabolism' and 'carotenoid biosynthesis' pathways in the KEGG database (https://www.kegg.jp/). Genes whose expression changed significantly in at least one time point after salt treatment are shown. **(C)** Content of chlorophyll (Chl) *a*, Chl *b* and carotenoids. Roots were harvested from plantlets after 250 mM NaCl treatment for two weeks. Bars indicate means ± standard deviation of three replicates. ** indicates significant differences at $P < 0.01$ according to the Dunnett test. DW, dry weight.

complex II, and cytochrome (Cyt) b_6f complex, as well as DEGs related to carbon fixation. A small number of photosynthetic genes were up-regulated, such as *PsaC*, *PsbA*, *PsbK* and *petD*, while all of the remaining genes were down-regulated in response to salt stress (Figure 7A). In addition, the expression of genes in the Calvin cycle decreased (Figure 7B). For example, three *RBCS* genes, which encode ribulose bisphosphate carboxylases, were strongly expressed before salinity treatment, but were down-regulated more than 2.5-fold after salt stress (Figure 7B). *D. officinale* is regarded as a facultative crassulacean acid metabolism (CAM) species (Zhang et al., 2014). Hence, the genes encoding key enzymes for CAM photosynthesis were identified and analyzed. Two *CA* genes that encode carbonic anhydrase were down-regulated (Figure 7C). Other genes involved in CAM photosynthesis, such as the malate dehydrogenase gene *MDH* and pyruvate phosphate dikinase

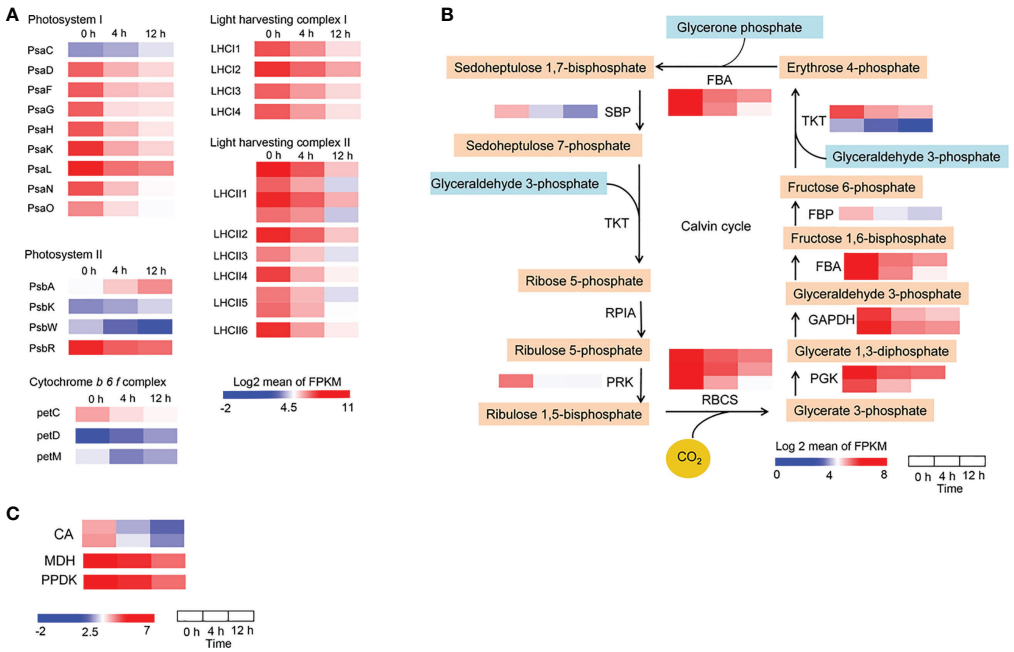


FIGURE 7
Changes in the expression of photosynthetic genes in response to salt treatment. **(A)** Expression pattern of genes encoding photosynthetic proteins, including the photosystem I complex, light harvesting complex I, photosystem II complex, light harvesting complex II, and cytochrome b_6f complex. **(B)** Heatmaps of genes involved in the Calvin cycle. SBP, sedoheptulose-1,7-bisphosphatase; TKT, transketolase; RPIA, ribose 5-phosphate isomerase A; PRK, phosphoribulokinase; RBCS, ribulose-bisphosphate carboxylase small chain; PGK, phosphoglycerate kinase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; FBA, fructose-bisphosphate aldolase; FBP, fructose-1,6-bisphosphatase I **(C)** Heatmaps of genes involved in the CAM pathway. CA, carbonic anhydrase; MDH, malate dehydrogenase; PPDK, pyruvate, phosphate dikinase. Genes that changed significantly at least one time point after salt treatment are shown.

gene *PPDK*, were highly expressed in the control, but were suppressed rapidly after salinity treatment (Figure 7C). These results indicate that photosynthesis was considerably repressed by salinity stress and that the CAM pathway was not active at an early stage of salt stress in *D. officinale* roots.

Global analysis of the amino acid metabolic pathway genes and composition of free amino acids

In plants, amino acids serve as precursors for the synthesis of a wide range of biologically important compounds, but they also play a role in stress response. A total of 218 DEGs related to the amino acid metabolic pathway were found. They were clustered into two main groups: down-regulated DEGs after salt stress, and up-regulated DEGs at least one time point after treatment (Supplementary Figure 3). In the 'Arginine biosynthesis' pathway, the genes involved in glutamate and ornithine biosynthesis were up-regulated after salinity stress (Figure 8A). For example, *ALT*, *glnA* and *gdnA*, which are related to glutamate biosynthesis, increased at least 2-fold at 4 h and 3-fold at 12 h after salinity stress (Figure 8A). The *arg* gene, which encodes an arginase that converts arginine to urea and ornithine, was up-regulated about 3-fold after salinity stress (Figure 8A). The Asparagine synthetase gene *ASNS* and the L-aspartate oxidase gene *nadB* were up-regulated in response to salt stress (Figure 8B). In addition, two *lysC* genes that encode aspartate kinases catalyzing the synthesis of asparagine from aspartate were up-regulated more than 3-fold at 12 h after salinity stress (Figure 8C). In lysine metabolism, the gene *AASS*, which encodes α-aminoadipic semialdehyde synthase involved in lysine catabolism, was up-regulated (Figure 8C). In proline biosynthesis, one *PRODH* gene was up-regulated while another was down-regulated (Figure 8D). Moreover, two *P4HA* genes involved in proline catabolism were down-regulated (Figure 8D). The dry weight content of free amino acids in roots decreased about 2-fold between control and salinity stress, even two weeks after salinity treatment (Table 1). The content of all of the main free amino acids (asparaginate, *O*-phosphoethanolamine, arginine, lysine, and aspartate), except for *O*-phosphoethanolamine, declined after salinity stress (Table 1). The contents of proline and ornithine, which are involved in osmotic stress responses such as salt and drought stresses in plants (Hussein et al., 2019), were significantly up-regulated after salinity

10.3389/fpls.2022.1028245

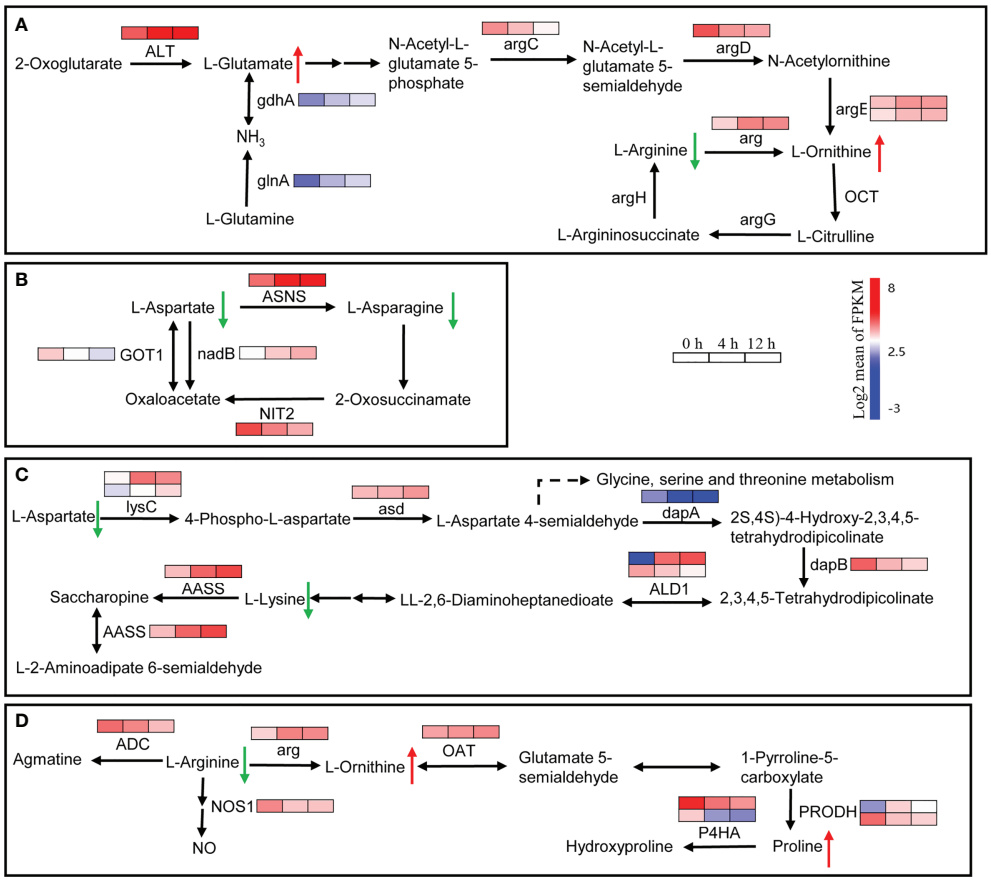


FIGURE 8
Expression heatmaps of genes involved in the selective amino acid metabolism pathway. **(A)** Changes in the expression of genes involved in glutamine, ornithine and arginine metabolism. ALT, alanine transaminase; argC, *N*-acetyl-γ-glutamyl-phosphate reductase; argD, acetylornithine aminotransferase; argE, acetylornithine deacetylase; argG, argininosuccinate synthase; arg, arginase; OCT, ornithine carbamoyltransferase; argH, argininosuccinate lyase; glnA, glutamine synthetase; gdhA, glutamate dehydrogenase. **(B)** Changes in the expression of genes involved in aspartate and asparagine metabolism. ASNS, asparagine synthase (glutamine-hydrolyzing); nadB, L-aspartate oxidase; GOT1, aspartate aminotransferase, cytoplasmic; NIT2, ω-amidase. **(C)** Changes in the expression of genes involved in aspartate and lysine metabolism. lysC, aspartate kinase; asd, aspartate-semialdehyde dehydrogenase; dapA, 4-hydroxy-tetrahydrodipicolinate synthase; dapB, 4-hydroxy-tetrahydrodipicolinate reductase; ALD1, LL-diaminopimelate aminotransferase; AASS, α-aminoadipic semialdehyde synthase. **(D)** Changes in the expression of genes involved in arginine, ornithine and proline metabolism. ADC, arginine decarboxylase; arg, arginase; OAT, ornithine aminotransferase; PRODH, proline dehydrogenase; P4HA, prolyl 4-hydroxylase; NOS1, nitric oxide synthase. Genes that changed significantly in at least one time point after salt treatment are shown.

stress (Table 1). These results suggest that the biosynthesis of main free amino acids such as aspartate and arginine is blocked, while the biosynthesis of stress-related amino acids like proline and ornithine is promoted in response to salinity stress.

Total flavonoids decreased in roots under salinity stress

In the flavonoid metabolic pathway, the expression level of two genes coding for flavonoid 3′-hydroxylase (*CYP75A* and *CYP75B1*) increased after salinity stress, suggesting that the conversion of dihydrokaempferol to dihydroquercetin or dihydromyricetin was enhanced (Figure 9A). Flavonoid biosynthetic genes like *HTC* and *CCOAMT* were also affected after salinity stress and their expression was up-regulated (Figure 9A). However, expression of the genes coding for chalcone isomerase (*CHI*) and chalcone synthase (*CHS*), which are key enzyme genes in the synthetic pathway of flavonoids, declined after salinity stress (Figure 9A). Total flavonoid content decreased in roots after salinity stress (Figure 9B). These results suggest that the biosynthesis of flavonoids is suppressed by salinity stress in *D. officinale* roots.

TABLE 1 Content of free amino acids (mg/100 mg DW) in *Dendrobium officinale* roots in control and salinity stress treatment.

Free amino acid	Control Mean ± SE	Salinity stress Mean ± SE	Regulation
O-Phospho-L-serine	0.0202 ± 0.0014	0.0207 ± 0.0018	
Taurine	0.0022 ± 0.0009	0.0110 ± 0.0010**	Up
O-Phosphoethanolamine	0.2150 ± 0.0176	0.2323 ± 0.0468	
L-Aspartic acid	0.0600 ± 0.0006	0.0453 ± 0.0030*	Down
L-Threonine	0.0216 ± 0.0004	0.0187 ± 0.0017	
L-Serine	0.0406 ± 0.0002	0.0740 ± 0.0045**	Up
L-Asparagine	3.8720 ± 0.0968	1.5855 ± 0.0665**	Down
L-Glutamic acid	0.0467 ± 0.0009	0.0610 ± 0.0050*	Up
D,L-α-Aminoadipic acid	0.0012 ± 0.0003	0.0012 ± 0.0002	
Glycine	0.0148 ± 0.0002	0.0127 ± 0.0013	
L-Alanine	0.0348 ± 0.0010	0.0430 ± 0.0040*	Up
L-Citrulline	0.0050 ± 0.0004	0.0103 ± 0.0060	
L-α-Amino-*n*-butanoic acid	0.0017 ± 0.0003	0.0020 ± 0.0006	
L-Cystine	0.0276 ± 0.0029	0.0615 ± 0.0029**	Up
L-Methionine	0.0027 ± 0.0006	0.0227 ± 0.0017**	Up
L-Isoleucine	0.0135 ± 0.0006	0.0148 ± 0.0015	
L-Leucine	0.0226 ± 0.0055	0.0200 ± 0.0029	
L-Tyrosine	0.0208 ± 0.0019	0.0252 ± 0.0037	
L-Phenylalanine	0.0175 ± 0.0009	0.0130 ± 0.0023	
β-Alanine	0.0010 ± 0.0000	0.0014 ± 0.0002	
D,L-β-Aminoisobutyric acid	0.0056 ± 0.0034	0.0008 ± 0.0002	
γ-Amino-*n*-butyric acid	0.0238 ± 0.0060	0.0468 ± 0.0093	
L-Histidine	0.0350 ± 0.0011	0.0323 ± 0.0024	
3-Methyl-L-histidine	0.0004 ± 0.0002	0.0005 ± 0.0003	
1-Methyl-L-histidine	ND	0.0004 ± 0.00024*	Up
L-Carnosine	0.0034 ± 0.0002	0.0017 ± 0.0005*	Down
L-Tryptophan	0.0022 ± 0.0007	0.0012 ± 0.0002	
Ornithine	0.0040 ± 0.0003	0.0062 ± 0.0005*	Up
L-Lysine	0.0805 ± 0.0016	0.0262 ± 0.0033**	Down
L-Arginine	0.9513 ± 0.0527	0.3297 ± 0.0182**	Down
L-Proline	0.0225 ± 0.0003	0.0240 ± 0.0003*	Up
Total	5.5923 ± 0.1356	2.7270 ± 0.2180**	Down

Mean values with an asterisk* in the same row are significantly different: * $P < 0.05$, ** $P < 0.01$, according to the Dunnet test. ND, not detected.

Global analysis of transcription factors and characterization of a *WRKY* gene in improvement of seed germination under salinity stress

TFs are widely involved in the regulation of metabolism and physiological processes, including stress response, in plants. A total of 226 and 243 TFs were up-regulated while 174 and 210 TFs were down-regulated at 4 and 12 h, respectively after salinity stress (Figure 10A). Among the down- or up-regulated genes at both time points, the largest number of down-regulated TFs was from the MYB family (22 genes) and the largest number of up-regulated TFs was from the APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) family (30 genes) (Supplementary Figure 4). Of note, all of the identified heat stress transcription factors were up-regulated at 4 h and 12 h. Only two WRKY genes were down-regulated while 17 WRKY genes were up-regulated at the detected time points after salt stress (Supplementary Figure 4).

To investigate the role of TFs in response to salinity stress, *DoWRKY69* with significantly up-regulated expression in WRKY family was selected and analyzed. *DoWRKY69* belongs to Group IIb (Supplementary Figure 5) and was up-regulated by about 4-fold at 4 h and by about 10-fold 12 h after salinity stress in roots (Supplementary Figure 6). The coding sequence (CDS) of *DoWRKY69* without a stop codon was cloned into pCAMBIA1302 at the *Nco*I site and is driven by the 35S promoter (Figure 10B). Ten independent transgenic lines for

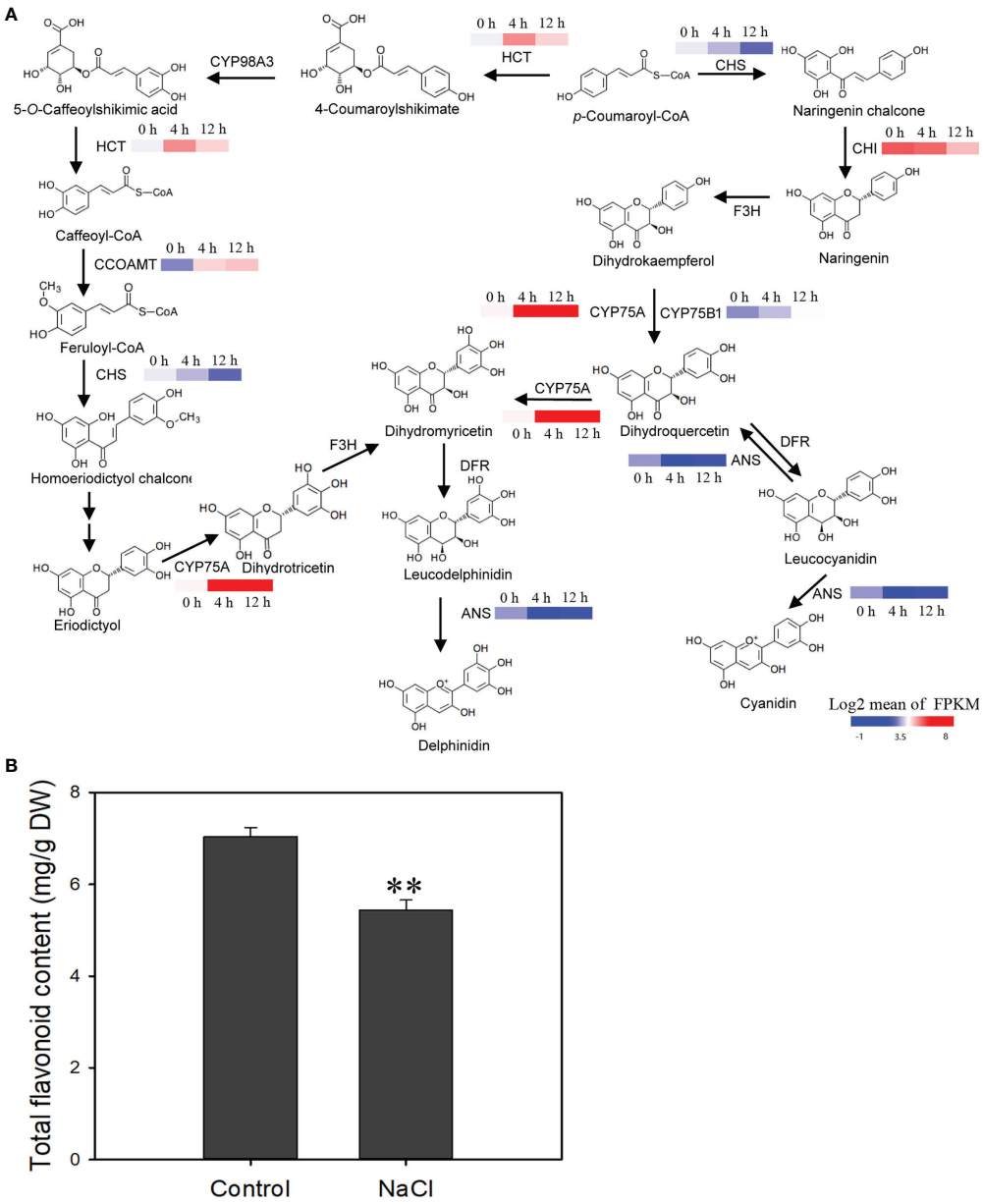


FIGURE 9
Changes in the expression of flavonoid biosynthetic genes and reduction of total flavonoids in response to salt treatment. **(A)** NaCl-stress induced changes in the expression of flavonoid biosynthetic genes. CHS, chalcone synthase; HCT, shikimate *O*-hydroxycinnamoyltransferase; CYP98A, 5-O-(4-coumaroyl)-D-quinate 3′-monooxygenase; CCOAMT, caffeoyl-CoA *O*-methyltransferase; CHI, chalcone isomerase; F3H, naringenin 3-dioxygenase; CYP75A, flavonoid 3′,5′-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase. **(B)** Estimation of total flavonoid content after salt stress. The roots of plantlets were collected after 250 mM NaCl treatment for two weeks and used to measure total flavonoid content. Bars indicate means ± standard deviation of three replicates. ** indicates significant differences at $P < 0.01$ according to the Dunnett test. DW, dry weight. Genes that changed significantly in at least one time point after salt treatment are shown.

DoWRKY69 were generated. Lines 7, 9 and 10 were used to analyze expression level and seed germination in response to three NaCl concentrations (150, 200, and 250 mM). The *DoWRKY69* gene was detected in all three transgenic lines but was not detected in WT plants (Figures 10C, D). In WT and transgenic lines, seeds germinated well and their germination percentage was nearly 100% in control medium 24 h after incubation (Figure 10E). However, the difference between WT

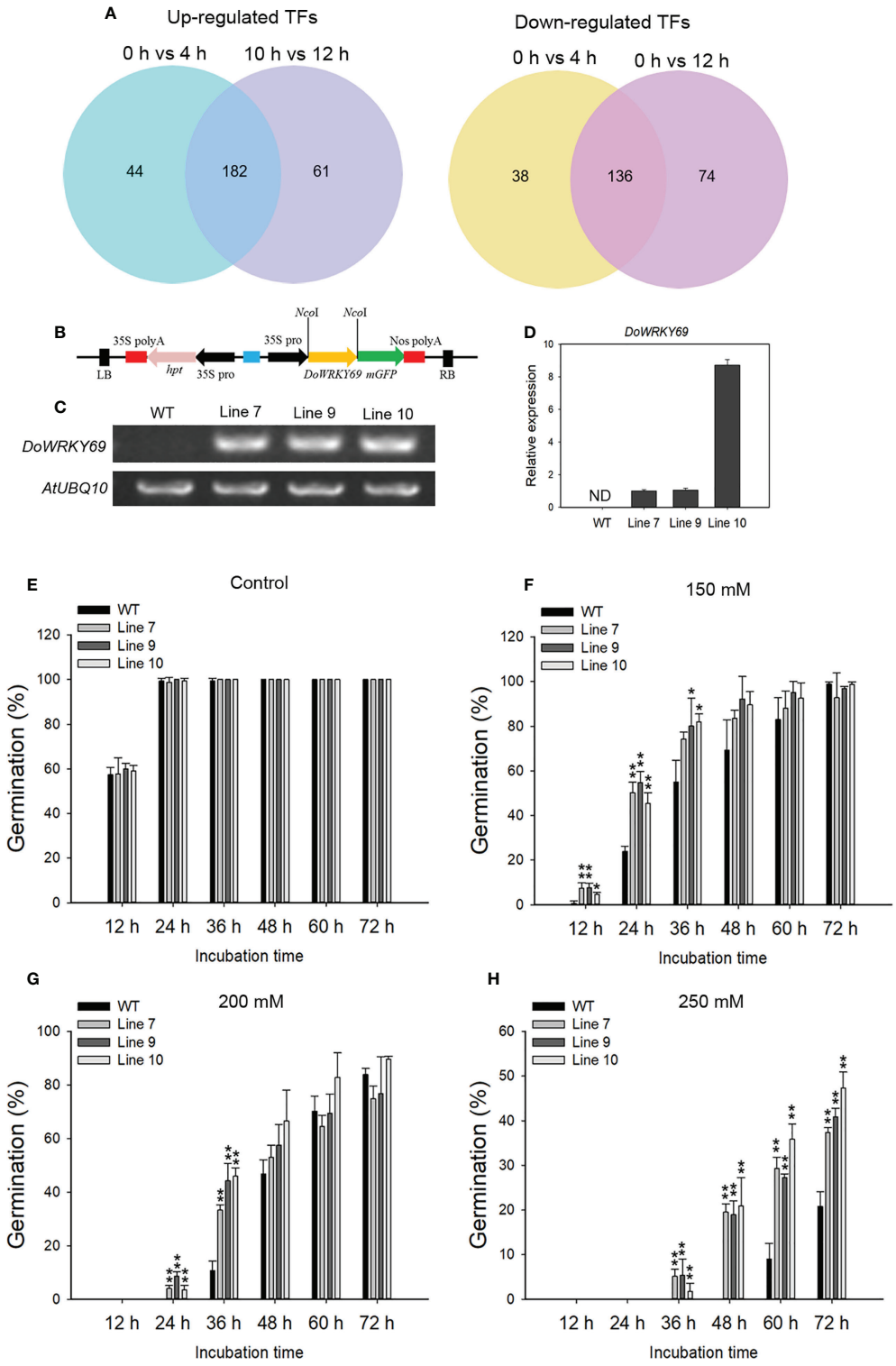


FIGURE 10
Identification of NaCl-induced TFs and characterization of one WRKY gene in the salt stress response. **(A)** Venn diagram of up-regulated TFs (left) and down-regulated TFs (right) from 0 h *vs* 4 h and 0 h *vs* 12 h comparisons. **(B)** The overexpression vector containing the *DoWRKY69* gene is shown. Analysis of the expression of the *DoWRKY69* gene in wild-type (WT) and *35S::DoWRKY69* lines by using RT-PCR **(C)** and qRT-PCR **(D)**. **(E)** Germination of WT and *35S::DoWRKY69* lines in the control (without NaCl), 150 mM NaCl **(F)**, 200 mM NaCl **(G)** and 250 mM NaCl **(H)**. Bars indicate means ± standard deviation of thee replicates. *, ** indicate significant differences at $P < 0.05$ and $P < 0.01$, respectively, according to the Dunnett test. About 60 seeds of each genotype were used.

and transgenic lines was obvious when the growth medium was supplemented with salt (Figures 10F-H). The germination rate of all transgenic lines was significantly higher than WT seeds within 24 h in response to 150 mM NaCl (Figure 10F). No WT seeds germinated within 24 h, while more than 3% of transgenic seeds of each transgenic line germinated with 24 h after exposure to 200 mM NaCl (Figure 10G). In addition, the germination of transgenic lines was significantly higher than WT seeds at 36 h after exposure to 200 mM NaCl (Figure 10H). Although no seeds germinated within 48 h in response to 250 mM NaCl, the germination of transgenic lines was significantly higher than WT seeds within 72 h after exposure to 250 mM NaCl (Figure 10H). These results suggest that *DoWRKY69* plays a positive role in salt stress tolerance in *A. thaliana*.

Discussion

Salinity is one of the most severe environmental factors limiting the productivity of agricultural crops. It is necessary to explore salinity tolerance mechanisms to help improve salt tolerance of crops *via* genetic engineering. In this study, we investigated transcriptomic reprogramming in *D. officinale* plantlets at an early phase of salinity stress, and characterized the ability of an up-regulated gene *DoWRKY69* to improve *A. thaliana* seed germination under salt stress. The roots of *D. officinale* plantlets are photosynthetic. The biosynthesis and signal transduction of hormones, such as IAA and cytokinin, which are essential for plant growth and development, were repressed, while stress-responsive hormones such as ABA, JA and SOS pathway genes were up-regulated at an early phase in response to salt stress. In order to survive and adapt to salinity stress, physiological and metabolic adjustments were made through extensive transcriptomic reprogramming. Our results suggest that physiological and metabolic processes and molecular functions are reprogrammed in the photosynthetic roots of this orchid when plants are exposed to salinity stress.

Plant hormones are crucial signaling molecules and their signaling depends on their spatio-temporal distribution (Waadt, 2020). Phytohormones coordinate all aspects of plant growth and development, as well as stress responses (Shan et al., 2012). Plant hormones interact with five key plant neurotransmitters, including serotonin, melatonin, dopamine, acetylcholine and γ-aminobutyric acid, and participate in many physiological processes such as photosynthesis, oxidative stress and osmotic regulation (Raza et al., 2022b). Two groups of phytohormones, auxins and cytokinins, have been clearly demonstrated as the main regulators of plant development (Benjamins and Scheres, 2008; Werner and Schmülling, 2009; Schaller et al., 2015) and stress responses (Hare et al., 1997; Blakeslee et al., 2019). For example, the *YUC* gene, which encodes indole-3-pyruvate monooxygenase, plays an important role in auxin (IAA) biosynthesis (Zhao, 2012). Overexpression of the *YUC* gene exhibited a drought-resistant phenotype in *A. thaliana* (Lee et al., 2012), and conferred water-deficit tolerance in potato (Kim et al., 2013b). However, improvement of drought tolerance by overexpression of the *YUC6* gene was not due to an increase in auxin synthesis, but rather due to *YUC6*-processed thiol-reductase, which inhibited the generation of ROS (Cha et al., 2015). Only one *YUC* gene was identified in this study and its expression was down-regulated after salt stress, suggesting that the synthesis of IAA might decrease in *D. officinale* roots in response to salt stress (Figure 3A). Moreover, the IAA biosynthesis pathway gene *IAA1* and auxin signal-mediated genes (*AUX1* and *ARF*) were down-regulated after salt stress, suggesting a reduction in auxin signals in response to salinity stress. Cytokinin biosynthesis pathway genes (three *IPT* genes and *CYP735A*) and the expression of cytokinin signal transduction pathway genes (*HK*, *AHP* and *A-ARR*, as well as one *B-ARR* gene) declined in *D. officinale* roots after salt stress, indicating the inhibition of cytokinin signaling in roots after salt stress. Biosynthetic genes such as *CYP735A2* from *A. thaliana* and *IPT* (a key enzyme for cytokinin biosynthesis) from tomato were predominantly expressed in roots, suggesting that roots are a major site for cytokinin synthesis (Takei et al., 2004; Ghanem et al., 2011). The down-regulated expression of *IPT* genes and *CYP735A* in *D. officinale* roots indicates a decrease in cytokinin. Some studies have shown that cytokinin signaling plays a negative role in stress responses. For example, two histidine kinase genes (*AHK2* and *AHK3*) negatively controlled osmotic stress responses in *A. thaliana*, and their single or double mutants displayed strong tolerance to drought and salt stress (Tran et al., 2007). These results indicate that cytokinin was a negative regulator of stress and that repression of the cytokinin signal in *D. officinale* roots might help plants to cope with salt stress. Most of the genes related to the biosynthesis of ABA, ethylene and JA and their signal transduction pathways were up-regulated after salt stress in *D. officinale* roots (Figures 3, 4). ABA, ethylene and JA signal transduction pathways after exposure to salt stress were described in greater detail in a fairly recent review (Kaleem et al., 2018). This indicates the conserved role of ABA, ethylene and JA in the response of plants to salt stress.

Photosynthesis is an essential process, converting solar energy into chemical energy. Salt stress plays a negative role in the photosynthesis of leaves (Sudhir and Murthy, 2004; Silva et al., 2011). Genes that encode the components of both the light and dark reactions of photosynthesis in the leaves of salt-treated soybean seedlings were slightly repressed or maintained at the early phase within 4 h, but were inhibited at a later phase (after 24 h) (Liu et al., 2019). Chloroplasts are important organelles central to plant photosynthesis and are affected by salt-induced toxicity, although different plant species and development stages display different degrees of resistance to salt stress (Suo et al., 2017; Hameed et al., 2021). In this study, an anatomical analysis demonstrated that *D. officinale* plantlet roots contained

chloroplasts (Figure 1), suggesting that the roots were capable of photosynthesis. The genes involved in the biosynthesis of photosynthetic pigments, as well as the genes encoding photosynthetic components, were strongly repressed in roots of salt-treated *D. officinale* plantlets within 12 h. These results indicate that photosynthesis is inhibited in both roots and leaves in response to salt stress, while the influence of time in both organs is different. Photosynthetic pigments were not discernibly altered under short-term salt stress (within 24 h) in the roots of *D. officinale* plantlets, although the expression of biosynthetic genes dropped drastically within 12 h (Figure 3). It is possible that the photosynthetic pigments were not damaged by salt stress in the first 12 h of exposure.

A decrease in photosynthetic rates under stress (dehydration, salt, extended darkness) may lead to an insufficient supply of carbohydrates. Free amino acids, which are tightly linked to carbohydrate metabolism, can be used as alternative substrates for mitochondrial respiration or act as precursors for the biosynthesis of secondary metabolites or immune signaling metabolites (Hildebrandt et al., 2015; Chen et al., 2018; Hildebrandt, 2018). In this study, the biosynthesis of arginine was repressed and its breakdown was induced, leading to a decrease of arginine in *D. officinale* roots after exposure to salt stress, suggesting the arginine might be used as a precursor for the biosynthesis of other compounds. For example, polyamines, which are synthesized from arginine, are responsive to plant stresses (Alcázar et al., 2006; Groppa and Benavides, 2008). Auxin can be synthesized from tryptophan (Wang et al., 2015a) and ethylene is synthesized from methionine (Sauter et al., 2013). There were no differences in tryptophan content while the expression of auxin biosynthetic genes was repressed by salt stress (Table 1 and Figure 8A). The increased expression of methionine and the deregulation of ethylene biosynthetic genes after salt stress (Table 1 and Figure 8D) suggest that ethylene synthesis is related to the methionine cycle. In addition, the higher level of stress-related proline and organic acids has been observed in salt cress (*Eutrema salsugineum*) exposed to extreme salt stress (Li et al., 2022). In the roots of *D. officinale*, proline biosynthetic genes were induced by salt and their expression was correlated with the increase in proline content.

Flavonoids, a diverse group of bioactive polyphenolic compounds, are catalyzed by a series of enzymes. CHS is first rate-limiting enzyme in flavonoid biosynthesis and catalyzes the production of naringenin chalcone using *p*-coumaroyl-CoA as the starting substrate (Martens et al., 2010; Dao et al., 2011). Under salinity stress, *CHS* and *CHI* were down-regulated, and this might have led to a decrease of total flavonoid production in *D. officinale* roots. Flavonoids are effective antioxidants that are involved in the response mechanisms of plants under adverse environments, including biotic and abiotic stresses (Agati et al., 2012; Petrussa et al., 2013). Studies have shown that salinity stress can enhance total flavonoid concentration in plant leaves. For example, flavonoid biosynthesis pathway genes such as the phenylalanine ammonia lyase gene *PAL*, the chalcone synthase gene *CHS* and the flavonol synthase gene *FLS*, as well as total flavonoids, were up-regulated in the leaves of *Solanum nigrum* under salt stress (Ben Abdallah et al., 2016). In our previous study, flavonoid biosynthesis genes (such as *CHS*, *CHI* and *F3H*) were up-regulated and total flavonoid content was enhanced in *D. officinale* leaves under salt stress (Zhang et al., 2021). Flavonoids act as antioxidants by scavenging ROS in stressed plant leaves (Agati et al., 2012). However, the content of flavonoids decreased in *D. officinale* roots under salt stress, suggesting that the path or mechanism by which ROS is scavenged in the roots of *D. officinale* is different from that in leaves.

Transcriptional activation of functional genes involved in stress responses by TFs leads to the adjustment of specific metabolism of metabolites, which is one strategy to cope with stress in plants. Different studies have identified many TFs involved in regulating the response to salt stress (Kaleem et al., 2018). Most up-regulated TFs belong to AP2/ERF, MYB, NAC and WRKY families (Supplementary Figure 3). Accumulating evidence has shown that WRKY genes play positive or negative roles in the regulation of the salt stress response. Maize (*Zea mays*) *WRKY17* and *WRKY114* negatively regulate salt stress tolerance (Cai et al., 2017; Bo et al., 2020). *GhWRKY17* from cotton (*Gossypium hirsutum*) and *WRKY46* from *A. thaliana* contribute to salt stress tolerance (Yan et al., 2014; Ding et al., 2015). A WRKY gene *DoWRKY69* was up-regulated after salt stress treatment. Over-expression of *DoWRKY69* demonstrated its role in improving *A. thaliana* seed germination under salt stress. These results indicate that stress-related TFs can be induced by salt stress treatment and might be responsible for the stress response.

Conclusion

During the salt stress response mechanism of *D. officinale in vitro* plantlets, the expression of genes involved in photosynthesis and flavonoid biosynthesis, as well as in the biosynthesis of auxin and cytokinin, declined, whereas the expression of genes coding for stress-related plant hormones (ABA, ethylene and JA), the signal transduction pathway, the SOS pathway, and the biosynthesis of amino acids related to osmotic adjustment, were activated (Figure 11). This indicates that *D. officinale* adapted to salt stress by reducing growth, the accumulation of compatible solutes, and increasing the exclusion of excess Na^+ in roots (Figure 11). Our findings illustrate a response mechanism of the roots of a facultative CAM plant to salinity stress. Our study also provides a large number of

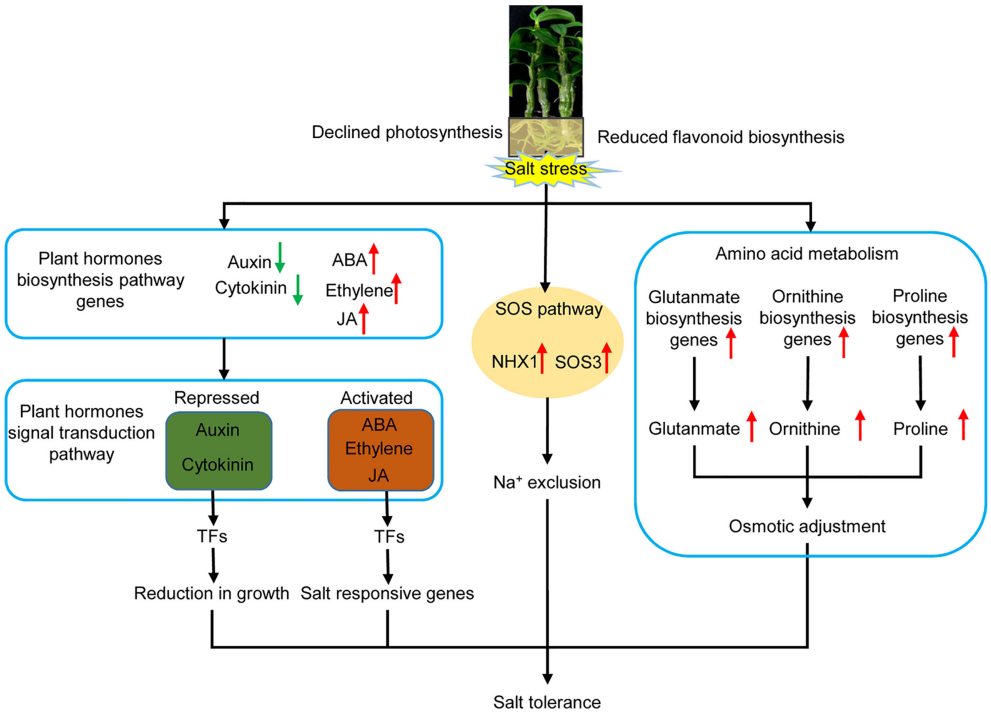


FIGURE 11
Schematic diagram of the response mechanism of *D. officinale* roots under salinity stress. Upward-facing red arrows represent activated/up-regulated targets. Downward-facing green arrows represent repressed/down-regulated genes.

candidate stress-responsive genes that could be used to develop and strengthen economically important crops using biotechnological approaches, allowing them to cope with salt stress.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: PRJNA715099

Author contributions

CH and MZ supervised the project. MZ and CH conceived the research and designed the experiments. RD performed histological analysis. MZ, CH, NL, XL, YY, JT and JD collectively interpreted the results and wrote all drafts of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (32071819 and 31800204), the Youth Science and Technology Talent Growth Project of Education Department of Guizhou Province of China (No. [2022]099), the Science and Technology Department Foundation of Guizhou Province of China (No. [2020]1Y120), and the Project of High-Level Talents Introduction in Qiannan Normal University for Nationalities (2021qnsyrc06).

Acknowledgments

We thank Dr. Minglei Zhao (South China Agricultural University) for helpful suggestions and Jian Liu (South China Botanical Garden, Chinese Academy of Sciences) for analyzing free amino acid content.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1028245/full#supplementary-material

References

Abdullakasim, S., Kongpaisan, P., Thongjang, P., and Saradhuldhat, P. (2018). Physiological responses of potted *Dendrobium* orchid to salinity stress. *Hortic. Environ. Biotechnol.* 59, 491–498. doi: 10.1007/s13580-018-0057-4

Abouelsaad, I., and Renault, S. (2018). Enhanced oxidative stress in the jasmonic acid-deficient tomato mutant *def-1* exposed to NaCl stress. *J. Plant Physiol.* 226, 136–144. doi: 10.1016/j.jplph.2018.04.009

Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as antioxidants in plants: Location and functional significance. *Plant Sci.* 196, 67–76. doi: 10.1016/j.plantsci.2012.07.014

Alcázar, R., Marco, F., Cuevas, J. C., Patron, M., Ferrando, A., Carrasco, P., et al. (2006). Involvement of polyamines in plant response to abiotic stress. *Biotechnol. Lett.* 28, 1867–1876. doi: 10.1007/s10529-006-9179-3

Apse, M. P., Aharon, G. S., Snedden, W. A., and Blumwald, E. (1999). Salt tolerance conferred by overexpression of a vacuolar Na^+/H^+ antiport in. *Arabidopsis. Sci.* 285, 1256–1258. doi: 10.1126/science.285.5431.1256

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Aziz, A., and Larher, F. (1995). Changes in polyamine titers associated with the proline response and osmotic adjustment of rape leaf discs submitted to osmotic stresses. *Plant Sci.* 112, 175–186. doi: 10.1016/0168-9452(95)04264-4

Ben Abdallah, S., Aung, B., Amyot, L., Lalin, I., Lachâal, M., Karray-Bouraoui, N., et al. (2016). Salt stress (NaCl) affects plant growth and branch pathways of carotenoid and flavonoid biosyntheses in *Solanum nigrum*. *Acta Physiol. Plant* 38, 72. doi: 10.1007/s11738-016-2096-8

Benjamins, R., and Scheres, B. (2008). Auxin: The looping star in plant development. *Annu. Rev. Plant Biol.* 59, 443–465. doi: 10.1146/annurev.arplant.58.032806.103805

Bhaskaran, S., and Savithramma, D. L. (2011). Co-Expression of *Pennisetum glaucum* vacuolar Na^+/H^+ antiporter and *Arabidopsis* H^+-pyrophosphatase enhances salt tolerance in transgenic tomato. *J. Exp. Bot.* 62, 5561–5570. doi: 10.1093/jxb/err237

Bielach, A., Hrtyan, M., and Tognetti, V. B. (2017). Plants under stress: Involvement of auxin and cytokinin. *Int. J. Mol. Sci.* 18, 1427. doi: 10.3390/ijms18071427

Blakeslee, J., Rossi, T., and Kriechbaumer, V. (2019). Auxin biosynthesis: Spatial regulation and adaptation to stress. *J. Exp. Bot.* 70, 5041–5049. doi: 10.1093/jxb/erz283

Bo, C., Chen, H., Luo, G., Li, W., Zhang, X., Ma, Q., et al. (2020). Maize *WRKY114* gene negatively regulates salt-stress tolerance in transgenic rice. *Plant Cell Rep.* 39, 135–148. doi: 10.1007/s00299-019-02481-3

Cai, R., Dai, W., Zhang, C., Wang, Y., Wu, M., Zhao, Y., et al. (2017). The maize WRKY transcription factor ZmWRKY17 negatively regulates salt stress tolerance in transgenic *Arabidopsis* plants. *Planta* 246, 1215–1231. doi: 10.1007/s00425-017-2766-9

Cha, J.-Y., Kim, W.-Y., Kang, S. B., Kim, J. I., Baek, D., Jung, I. J., et al. (2015). A novel thiol-reductase activity of *Arabidopsis* YUC6 confers drought tolerance independently of auxin biosynthesis. *Nat. Commun.* 6, 8041. doi: 10.1038/ncomms9041

Chen, Y. C., Holmes, E. C., Rajniak, J., Kim, J. G., Tang, S., Fischer, C. R., et al. (2018). *N*-hydroxy-pipecolic acid is a mobile metabolite that induces systemic disease resistance in Arabidopsis *Proc. Natl. Acad. Sci. U.S.A.* 115, E4920–E4929. doi: 10.1073/pnas.1805291115

Clough, S. J., and Bent, A. F. (1998). Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Dao, T. T. H., Linthorst, H. J. M., and Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochem. Rev.* 10, 397–412. doi: 10.1007/s11101-011-9211-7

Deinlein, U., Stephan, A. B., Horie, T., Luo, W., Xu, G., and Schroeder, J. I. (2014). Plant salt-tolerance mechanisms. *Trends Plant Sci.* 19, 371–379. doi: 10.1016/j.tplants.2014.02.001

Ding, Z. J., Yan, J. Y., Li, C. X., Li, G. X., Wu, Y. R., and Zheng, S. J. (2015). Transcription factor WRKY46 modulates the development of arabidopsis lateral roots in osmotic/salt stress conditions *via* regulation of ABA signaling and auxin homeostasis. *Plant J.* 84, 56–69. doi: 10.1111/tpj.12958

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al. (2003). *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33, 751–763. doi: 10.1046/j.1365-313x.2003.01661.x

Ernst, J., and Bar-Joseph, Z. (2006). STEM: A tool for the analysis of short time series gene expression data. *BMC Bioinf.* 7, 191. doi: 10.1186/1471-2105-7-191

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2013). Pfam: The protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2021). COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* 49, D274–D281. doi: 10.1093/nar/gkaa1018

Ghanem, M. E., Albacete, A., Smigocki, A. C., Frébort, I., Pospíšilová, H., Martínez-Andújar, C., et al. (2011). Root-synthesized cytokinins improve shoot growth and fruit yield in salinized tomato (*Solanum lycopersicum* L.) plants. *J. Exp. Bot.* 62, 125–140. doi: 10.1093/jxb/erq266

Givnish, T. J., Spalink, D., Ames, M., Lyon, S. P., Hunter, S. J., Zuluaga, A., et al. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. Biol. Sci.* 282, 20151553. doi: 10.1098/rspb.2015.1553

Groppa, M. D., and Benavides, M. P. (2008). Polyamines and abiotic stress: Recent advances. *Amino Acids* 34, 35–45. doi: 10.1007/s00726-007-0501-8

Hameed, A., Ahmed, M. Z., Hussain, T., Aziz, I., Ahmad, N., Gul, B., et al. (2021). Effects of salinity stress on chloroplast structure and function. *Cells* 10, 2023. doi: 10.3390/cells10082023

Hare, P. D., Cress, W. A., and van Staden, J. (1997). The involvement of cytokinins in plant responses to environmental stress. *Plant Growth Regul.* 23, 79–103. doi: 10.1023/a:1005954525087

He, K., Li, C., Zhang, Z., Zhan, L., Cong, C., Zhang, D., et al. (2022). Genome-wide investigation of the *ZF-HD* gene family in two varieties of alfalfa (*Medicago sativa* L.) and its expression pattern under alkaline stress. *BMC Genomics* 23, 150. doi: 10.1186/s12864-022-08309-x

Hildebrandt, T. M. (2018). Synthesis versus degradation: Directions of amino acid metabolism during arabidopsis abiotic stress response. *Plant Mol. Biol.* 98, 121–135. doi: 10.1007/s11103-018-0767-0

Hildebrandt, T. M., Nesi, A. N., Araújo, W. L., and Braun, H. P. (2015). Amino acid catabolism in plants. *Mol. Plant* 8, 1563–1579. doi: 10.1016/j.molp.2015.09.005

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hussein, H. A. A., Mekki, B. B., El-Sadek, M. E. A., and Lateef, E. E. E. (2019). Effect of L-ornithine application on improving drought tolerance in sugar beet plants. *Heliyon* 5, e02631. doi: 10.1016/j.heliyon.2019.e02631

Ji, H., Pardo, J. M., Batelli, G., Van Oosten, M. J., Bressan, R. A., and Li, X. (2013). The salt overly sensitive (SOS) pathway: Established and emerging roles. *Mol. Plant* 6, 275–286. doi: 10.1093/mp/sst017

Kaleem, F., Shabir, G., Aslam, K., Rasul, S., Manzoor, H., Shah, S. M., et al. (2018). An overview of the genetics of plant response to salt stress: Present status and the way forward. *Appl. Biochem. Biotechnol.* 186, 306–334. doi: 10.1007/s12010-018-2738-y

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi: 10.1093/nar/gkh063

Kazan, K. (2015). Diverse roles of jasmonates and ethylene in abiotic stress tolerance. *Trends Plant Sci.* 20, 219–229. doi: 10.1016/j.tplants.2015.02.001

Kieber, J. J., and Schaller, G. E. (2018). Cytokinin signaling in plant development. *Development* 145, dev149344. doi: 10.1242/dev.149344

Kim, J. I., Baek, D., Park, H. C., Chun, H. J., Oh, D.-H., Lee, M. K., et al. (2013b). Overexpression of *Arabidopsis YUCCA6* in potato results in high-auxin developmental phenotypes and enhanced resistance to water deficit. *Mol. Plant* 6, 337–349. doi: 10.1093/mp/sss100

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013a). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36. doi: 10.1186/gb-2013-14-4-r36

Knight, H., Trewavas, A. J., and Knight, M. R. (1997). Calcium signalling in *Arabidopsis thaliana* responding to drought and salinity. *Plant J.* 12, 1067–1078. doi: 10.1046/j.1365-313x.1997.12051067.x

Krasensky, J., and Jonak, C. (2012). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *J. Exp. Bot.* 63, 1593–1608. doi: 10.1093/jxb/err460

Kwon, O. K., Mekapogu, M., and Kim, K. S. (2019) Effect of salinity stress on photosynthesis and related physiological responses in carnation (*Dianthus caryophyllus*). *Hortic. Environ. Biotechnol.* 60, 831–839. doi: 10.1007/s13580-019-00189-7

Lee, M., Jung, J. H., Han, D. Y., Seo, P. J., Park, W. J., and Park, C. M. (2012). Activation of a flavin monooxygenase gene *YUCCA7* enhances drought resistance in. *Arabidopsis. Planta.* 235, 923–938. doi: 10.1007/s00425-011-1552-3

Li, C., Duan, C., Zhang, H., Zhao, Y., Meng, Z., Zhao, Y., et al. (2022). Adaptative mechanisms of halophytic *Eutrema salsugineum* encountering saline environment. *Front. Plant Sci.* 13, 909527. doi: 10.3389/fpls.2022.909527

Liu, H., Wang, L., Jing, X., Chen, Y., and Hu, F. (2021). Functional analysis of *CgWRKY57* from *Cymbidium goeringii* in ABA response. *PeerJ* 9, e10982. doi: 10.7717/peerj.10982

Liu, A., Xiao, Z., Li, M. W., Wong, F. L., Lam, H. M., and Yung, W. S. (2019). Transcriptomic reprogramming in soybean seedlings under salt stress. *Plant Cell Environ.* 42, 98–114. doi: 10.1111/pce.13186

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

Mansour, M. M. F., and Hassan, F. A. S. (2022). How salt stress-responsive proteins regulate plant adaptation to saline conditions. *Plant Mol. Biol.* 108, 175–224. doi: 10.1007/s11103-021-01232-x

Martens, S., Preuß, A., and Matern, U. (2010). Multifunctional flavonoid dioxygenases: Flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana* L. *Phytochemistry* 71, 1040–1049. doi: 10.1016/j.phytochem.2010.04.016

Moharramnejad, S., Sofalian, O., Valizadeh, M., Asgari, A., and Shiri, M. (2015). Proline, glycine betaine, total phenolics and pigment contents in response to osmotic stress in maize seedlings. *J. Biosci. Biotechnol.* 4, 313–319.

Munemasa, S., Hauser, F., Park, J., Waadt, R., Brandt, B., and Schroeder, J. I. (2015). Mechanisms of abscisic acid-mediated control of stomatal aperture. *Curr. Opin. Plant Biol.* 28, 154–162. doi: 10.1016/j.pbi.2015.10.010

Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Plant Physiol.* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S., et al. (2013). Plant flavonoids – biosynthesis, transport and involvement in stress responses. *Int. J. Mol. Sci.* 14, 14950–14973. doi: 10.3390/ijms140714950

Raza, A., Salehi, H., Rahman, M. A., Zahid, Z., Haghjou, M. M., Najafi-Kakavand, S., et al. (2022b). Plant hormones and neurotransmitter interactions mediate antioxidant defenses under induced oxidative stress in plants. *Front. Plant Sci.* 13, 961872. doi: 10.3389/fpls.2022.961872

Raza, A., Tabassum, J., Fakhar, A. Z., Sharif, R., Chen, H., Zhang, C., et al. (2022a). Smart reprograming of plants against salinity stress using modern biotechnological tools. *Crit. Rev. Biotechnol.* 15, 1–28. doi: 10.1080/07388551.2022.2093695

Ren, C., Wang, J., Xian, B., Tang, X., Liu, X., Hu, X., et al. (2020). Transcriptome analysis of flavonoid biosynthesis in safflower flowers grown under different light intensities. *PeerJ* 8, e8671. doi: 10.7717/peerj.8671

Sauter, M., Moffatt, B., Saechao, M. C., Hell, R., and Wirtz, M. (2013). Methionine salvage and *S*-adenosylmethionine: Essential links between sulfur, ethylene and polyamine biosynthesis. *Biochem. J.* 451, 145–154. doi: 10.1042/BJ20121744

Schaller, G. E., Bishopp, A., and Kieber, J. J. (2015). The yin-yang of hormones: Cytokinin and auxin interactions in plant development. *Plant Cell* 27, 44–63. doi: 10.1105/tpc.114.133595

Shalmani, A., Muhammad, I., Sharif, R., Zhao, C., Ullah, U., Zhang, D., et al. (2019). Zinc finger-homeodomain genes: Evolution, functional differentiation, and expression profiling under flowering-related treatments and abiotic stresses in plants. *Evol. Bioinform.* 15, 1–16. doi: 10.1177/1176934319867930

Shan, X., Yan, J., and Xie, D. (2012). Comparison of phytohormone signaling mechanisms. *Curr. Opin. Plant Biol.* 15, 84–91. doi: 10.1016/j.pbi.2011.09.006

Silva, E. N., Ribeiro, R. V., Ferreira-Silva, S. L., Viégas, R. A., and Silveira, J. A. G. (2011). Salt stress induced damages on the photosynthesis of physic nut young plants. *Sci. Agric.* 68, 62–68. doi: 10.1590/S0103-90162011000100010

Sudhir, P., and Murthy, S. D. S. (2004). Effects of salt stress on basic processes of photosynthesis. *Photosynthetica* 42, 481–486. doi: 10.1007/S11099-005-0001-6

Sumanta, N., Haque, C. I., Nishika, J., and Suprakash, R. (2014). Spectrophotometric analysis of chlorophylls and carotenoids from commonly grown fern species by using various extracting solvents. *Res. J. Chem. Sci.* 4, 63–69. doi: 10.1055/s-0033-1340072

Suo, J., Zhao, Q., David, L., Chen, S., and Dai, S. (2017). Salinity response in chloroplasts: Insights from gene characterization. *Int. J. Mol. Sci.* 18, 1011. doi: 10.3390/ijms18051011

Takei, K., Yamaya, T., and Sakakibara, H. (2004). *Arabidopsis CYP735A1* and *CYP735A2* encode cytokinin hydroxylases that catalyze the biosynthesis of *trans*-zeatin. *J. Biol. Chem.* 279, 41866–41872. doi: 10.1074/jbc.M406337200

Teixeira da Silva, J. A., and Ng, T. B. (2017). The medicinal and pharmaceutical importance of *Dendrobium* species. *Appl. Microbiol. Biotechnol.* 101, 2227–2239. doi: 10.1007/s00253-017-8169-9

The UniProt Consortium (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 45, D158–D169. doi: 10.1093/nar/gkw1099

Tran, L. S., Urao, T., Qin, F., Maruyama, K., Kakimoto, T., Shinozaki, K., et al. (2007). Functional analysis of AHK1/ATHK1 and cytokinin receptor histidine kinases in response to abscisic acid, drought, and salt stress in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20623–20628. doi: 10.1073/pnas.0706547105

Tuteja, N. (2007). Mechanisms of high salinity tolerance in plants. *Methods Enzymol.* 428, 419–438. doi: 10.1016/S0076-6879(07)28024-3

Tütüncü Konyar, S., Dane, F., and Tütüncü, S. (2013). Distribution of insoluble polysaccharides, neutral lipids, and proteins in the developing anthers of *Campsis radicans* (L.) Seem. (Bignoniaceae). *Plant Syst. Evol.* 299, 743–760. doi: 10.1007/s00606-013-0758-1

Verma, D., Jalmi, S. K., Bhagat, P. K., Verma, N., and Sinha, A. K. (2020). A bHLH transcription factor, MYC2, imparts salt intolerance by regulating proline biosynthesis in arabidopsis. *FEBS J.* 287, 2560–2576. doi: 10.1111/febs.15157

Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16, 86. doi: 10.1186/s12870-016-0771-y

Waadt, R. (2020). Phytohormone signaling mechanisms and genetic methods for their modulation and detection. *Curr. Opin. Plant Biol.* 57, 31–40. doi: 10.1016/j.pbi.2020.05.011

Wang, F., Chen, H.-W., Li, Q.-T., Wei, W., Li, W., Zhang, W.-K., et al. (2015b). GmWRKY27 interacts with GmMYB174 to reduce expression of *GmNAC29* for stress tolerance in soybean plants. *Plant J.* 83, 224–236. doi: 10.1111/tpj.12879

Wang, B., Chu, J., Yu, T., Xu, Q., Sun, X., Yuan, J., et al. (2015a). Tryptophan-independent auxin biosynthesis contributes to early embryogenesis in arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4821–4826. doi: 10.1073/pnas.1503998112

Wang, C., Wang, L., Lei, J., Chai, S., Jin, X., Zou, Y., et al. (2022). *IbMYB308*, a sweet potato R2R3-MYB gene, improves salt stress tolerance in transgenic tobacco. *Genes* 13, 1476. doi: 10.3390/genes13081476

Wasternack, C. (2015). How jasmonates earned their laurels: Past and present. *J. Plant Growth Regul.* 34, 761–794. doi: 10.1007/s0034-015-9526-5

Weigel, D., and Glazebrook, J. (2006). Transformation of *Agrobacterium* using the freeze-thaw method. *CSH Protoc.* 2006, pdb.prot4666. doi: 10.1101/pdb.prot4666

Werner, T., and Schmülling, T. (2009). Cytokinin action in plant development. *Curr. Opin. Plant Biol.* 12, 527–538. doi: 10.1016/j.pbi.2009.07.002

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483

Xu, X.-B., Pan, Y.-Y., Wang, C.-L., Ying, Q.-C., Song, H.-M., and Wang, H.-Z. (2014). Overexpression of *DnWRKY11* enhanced salt and drought stress tolerance of transgenic tobacco. *Biologia* 69, 994–1000. doi: 10.2478/s11756-014-0398-0

Yan, H., Jia, H., Chen, X., Hao, L., An, H., and Guo, X. (2014). The cotton WRKY transcription factor GhWRKY17 functions in drought and salt stress in transgenic *Nicotiana benthamiana* through ABA signaling and the modulation of reactive oxygen species production. *Plant Cell Physiol.* 55, 2060–2076. doi: 10.1093/pcp/pcu133

Zhang, Z., He, D., Niu, G., and Gao, R. (2014). Concomitant CAM and C3 photosynthetic pathways in *Dendrobium officinale* plants. *J. Amer. Soc Hortic. Sci.* 139, 290–298. doi: 10.21273/JASHS.139.3.290

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

Zhang, M., Yu, Z., Zeng, D., Si, C., Zhao, C., Wang, H., et al. (2021). Transcriptome and metabolome reveal salt-stress responses of leaf tissues from. *Dendrobium. officinale. Biomolecules* 11, 736. doi: 10.3390/biom11050736

Zhao, Y. (2012). Auxin biosynthesis: A simple two-step pathway converts tryptophan to indole-3-acetic acid in plants. *Mol. Plant* 5, 334–338. doi: 10.1093/mp/ssr104

Zhu, J.-K. (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247–273. doi: 10.1146/annurev.arplant.53.091401.143329

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 October 2022
DOI 10.3389/fpls.2022.954467

OPEN ACCESS

EDITED BY
Jen-Tsung Chen,
National University of Kaohsiung, Taiwan

REVIEWED BY
Sezai Ercisli,
Atatürk University, Turkey
Phanikanth Jogam,
Kakatiya University, India
Pandiyan Muthuramalingam,
Gyeongsang National University, South Korea

*CORRESPONDENCE
Sanjay Kumar
sanjaykumar@ihbt.res.in
Ravi Shankar
ravish9@gmail.com

†These authors have contributed equally to this work

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

RECEIVED 27 May 2022
ACCEPTED 01 August 2022
PUBLISHED 18 October 2022

CITATION
Bhattacharyya P, Sharma T, Yadav A, Lalthafamkimi L, Ritu, Swarnkar MK, Joshi R, Shankar R and Kumar S (2022) *De novo* transcriptome based insights into secondary metabolite biosynthesis in *Malaxis acuminata* (*Jeevak*)–A therapeutically important orchid. *Front. Plant Sci.* 13:954467. doi: 10.3389/fpls.2022.954467

De novo transcriptome based insights into secondary metabolite biosynthesis in *Malaxis acuminata* (*Jeevak*)–A therapeutically important orchid

Paromik Bhattacharyya[1†], Tanvi Sharma[1†], Abhinandan Yadav[2†], Lucy Lalthafamkimi[3,4], Ritu[2,4], Mohit Kumar Swarnkar[1], Robin Joshi[5], Ravi Shankar[2,4*] and Sanjay Kumar[1,4*]

[1]Biotechnology Division, Council of Scientific and Industrial Research-Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh, India, [2]Studio of Computational Biology & Bioinformatics, The Himalayan Centre for High-throughput Computational Biology (HiCHiCoB, A BIC supported by DBT, India), Council for Scientific and Industrial Research (CSIR)-Institute of Himalayan Bioresource Technology (CSIR-IHBT), Palampur, Himachal Pradesh, India, [3]Agrotechnology and Rural Development Division (ARDD), CSIR-North East Institute of Science and Technology, Jorhat, Assam, India, [4]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, Uttar Pradesh, India, [5]Chemical Technology Division, Council of Scientific and Industrial Research-Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh, India

Malaxis acuminata D. Don [=*Crepidium acuminatum* (D. Don) Szlach.] is an endangered medicinal orchid of the *Ashtvarga* group of plants in *Ayurveda* (Indian system of traditional medicine). Using a combination of aromatic cytokinin [*meta*-Topolin (*m*T)], plant biostimulant (chitosan), auxin [indole-3-butyric acid (IBA)], and a phenolic elicitor [phloroglucinol (PG)], plants of *M. acuminata* were regenerated *in vitro* for mass multiplication. The present research reveals the first-ever transcriptome of *M. acuminata*. A total of 43,111 transcripts encoding 23,951 unigenes were assembled *de novo* from a total of 815.02 million reads obtained from leaf and pseudobulb of *in vitro* raised *M. acuminata*. Expression analysis of genes associated with β-sitosterol and eugenol biosynthesis in leaf and pseudobulb provided vital clues for differential accumulation of metabolites in *M. acuminata*. Ultra-performance liquid chromatography (UPLC) confirmed higher amounts of β-sitosterol and eugenol content in the leaf as compared to the pseudobulb. Differential expression of transcripts related to starch and sucrose metabolism, plant hormone signal transduction, diterpenoid biosynthesis, phenylalanine metabolism, stilbenoid, diarylheptanoid, and gingerol biosynthesis suggested the operation of differential metabolic pathways in leaf and pseudobulb. The present research provides valuable information on the biosynthesis of secondary metabolites in *M. acuminata*, which could be used for advanced metabolite bioprospection using cell suspension culture and bioreactor-based approaches. Data also suggested

that leaf tissues rather than pseudobulb can be used as an alternate source of bioactive metabolites thereby shifting the need for harvesting the pseudobulb. This will further facilitate the conservation and sustainable utilization of this highly valued medicinal orchid.

KEYWORDS

***Ashtvarga*, *Ayurveda*, eugenol, *meta*-Topolin (*m*T), transcriptome, orchid, β-sitosterol**

Introduction

In the recent times, phytomedicines have attracted increased attention for their use in the treatment of chronic ailments (Golechha, 2020). Medicinal plants display distinct bioactivities which enable their use in the food, cosmetic, and pharmaceutical industries (Dogan et al., 2014; Mollova et al., 2020; Kibar, 2021). Amongst various medicinal plants used in herbal preparations, orchids deserve special mention (Hossain, 2011; Gantait et al., 2021). Interestingly, orchids occupy a prominent position in the traditional Indian system of medicine "*Ayurveda*". Categorized amongst the "*Ashtvarga*" group of eight medicinal herbs, *Malaxis acuminata* D. Don [=*Crepidium acuminatum* (D. Don) Szlach.] or "*Jeevak*"- a prized medicinal orchid figures out prominently (Kaur and Bhutani, 2010). The dried pseudobulb of *M. acuminata* contains rich reserves of various bioactive compounds which have antioxidant, cardioprotective, and anti-aging properties (Bose et al., 2017a; Giri et al., 2017; Balkrishna et al., 2018). Also, the pseudobulb of *M. acuminata* forms an important ingredient of "*Chyawanprash*"- an age-old health tonic marketed by various pharmaceutical brands across the globe (Bose et al., 2017a; Balkrishna et al., 2018). Recent reports on *M. acuminata* reveal an alarming scenario with high rates of anthropogenic intrusions and the destruction of natural habitats (Cheruvathur et al., 2010; Bose et al., 2017b). With a rapid change in the global climatic patterns in the past few decades, significant changes have been recorded in the habitat and distribution of plants occurring in various biodiversity hotspots globally amongst which the Himalayas deserve special mention. Based on the habitat and species distribution data, Conservation Assessment and Management Plan-World Wide Fund for Nature has categorized *M. acuminata* as "vulnerable" whereas the Convention on International Trade in Endangered Species of Wild Fauna and Flora has included *M. acuminata* within the Appendix II of endangered species (Lohani et al., 2013).

However, in spite of its high medicinal potential, there exists a persistent information gap with respect to the secondary metabolite profile of *M. acuminata* and the mechanisms regulating their biosynthesis. With habitat distribution in inaccessible regions of Himalayan, sub-Himalayan, and trans-Himalayan mountain ranges, collection of *M. acuminata* is often challenging and difficult (Bose et al., 2017a; Giri et al., 2017). In recent times, comprehensive advancements have been made in the research domain of plant tissue culture (PTC) with special reference to medicinal plants and orchids (Gantait et al., 2021). PTC is being recognized as an essential biotechnological tool not only for the supply of quality planting material (QPM) necessary for therapeutic bioprospection of metabolites but also for the sustainable replenishment of endangered germplasm of plants with high medicinal importance (Nalawade et al., 2003; Espinosa-Leal et al., 2018; Niazian, 2019). However, in order to modulate and selectively enhance metabolite production within *in vitro* systems, prior knowledge about the biosynthetic pathways, metabolic fluxes and regulatory mechanism involved in the metabolite biosynthesis is a principle prerequisite that can be obtained by the next generation sequencing (NGS) based transcriptomic approach (Han et al., 2016; Yamazaki et al., 2018; Shih and Morgan, 2020).

The emergence of NGS in the recent past has offered a feasible alternative to researchers for analyzing the genomes and transcriptome data of non-model plant species including orchids (Dhiman et al., 2019; Sharma et al., 2021). Transcriptomes of medicinal orchids, such as *Fontainea picrosperma* (Shan et al., 2021), *Dendrobium officinale* (He et al., 2020; Zhang et al., 2021), *Dactylorhiza hatagirea* (Dhiman et al., 2019), and *Gastrodia elata* (Shan et al., 2021) have been sequenced and assembled. Transcriptome profiling has provided vital molecular insights into the genes involved in metabolite biosynthesis in prized non-model medicinal plants (Yamazaki et al., 2018; Shih and Morgan, 2020). Furthermore, NGS approaches have facilitated the generation of large simple sequence repeat-expressed sequence tag (SSR-EST) datasets, especially of the non-model medicinal plants with negligible or none genomic data available (Morozova et al., 2009; Simon et al., 2009; Gai et al., 2012).

In the past few decades, a few coordinated research endeavors have contributed critical insights into the pharmacological and phytochemical constitution of *M. acuminata* (Suyal et al., 2020). However, to date, no information exists to elucidate the metabolic pathways involved in the biosynthesis of the key secondary metabolites of

medicinal importance. Before the advent of NGS platforms, the identification and generation of SSR and ESTs were widely used for the understanding of biosynthetic pathways in non-model plants including orchids (Li et al., 2010; Sun et al., 2010). However, to the best of our knowledge, no information exist on the SSR and EST data at the National Centre for Biotechnology Information (NCBI) Gene Bank database.

With this background, the present study was carried out to analyze the comparative transcriptome profiles of the leaf and pseudobulb tissues of *M. acuminata* and their correlation with target metabolites, β-sitosterol, and eugenol, content. Data thus generated can further be utilized for metabolite up scaling by application of relevant biotechnological tools.

Materials and methods

Plant tissue culture of *M. acuminata*

Plants of *M. acuminata* were sampled from the natural habitat of upper Shillong (Meghalaya), India, and were maintained in the greenhouse of Council of Scientific and Industrial Research (CSIR)- Institute of Himalayan Bio-resource Technology (IHBT), Palampur, Himachal Pradesh. Transverse thin cell layer (*t*-TCL) explants were used for micropropagation of *M. acuminata* in Murashige and Skoog medium (Murashige and Skoog, 1962) supplemented with an optimized plant growth trigger comprising of 1.5 mg/L *meta*-Topolin (*m*T), 5.0 mg/L of chitosan for multiple shoot induction along with 1.5 mg/L indole-3-butyric acid (IBA), and 5.0 mg/L phloroglucinol (PG) for rooting (Bhattacharyya et al., 2022). The cultures were maintained under optimum growth conditions at 25 ± 2°C temperature, 80% relative humidity, and 12 h alternative photoperiod of light with 40 μmol m^{-2} s^{-1} irradiance provided by cool white fluorescent light (Phillips, India). For transcriptome sequencing, leaf and pseudobulb tissues were excised from the fully developed *in vitro* grown plants.

RNA extraction, library preparation, and sequencing

Total RNA was isolated following the protocol described by Li and Trick (2005). The quality and quantity of RNA were checked using a Qubit 2.0 Fluorometer (Invitrogen, United States) and NanoDrop 1000 (NanoDrop Technologies, United States). Libraries (paired-end) of leaf and pseudobulb tissues were prepared using TruSeq stranded mRNA sample prep kit v2 (Illumina Incorporation, United States). The libraries were quantified using Qubit 4.0 Fluorometer (Life Technologies, United States) and DNA 1000 chip on Bioanalyzer (Agilent 2100 technologies, United States). ExAMP chemistry was used to generate the clusters and 400 pM of each prepared library was loaded in an S2 flow cell lane of the 2 × 100 bp paired-end format of NovaSeq 6000 (Illumina Inc., United States). The raw sequenced data was de-multiplexed with a BCL2FASTQ converter which resulted in FASTQ data format of final sequence reads. The quality was checked using the Fast QC tool for final sequence reads and then subjected to filtering with the in-house developed tool filteR (Gahlan et al., 2012). Horizontal filtering was done to check the quality of each read, whereas vertical filtering was done to check the consistency of the quality score position in the read files. *De novo* assembly was performed for the reads which had a Phred score of ≥QV30 for more than 70% bases. The reads that did not fall under the "cutoff range" were removed.

De novo sequence assembly

The *de novo* assembly of transcripts was done using the SOAP*de novo*-Trans with 127 k-mer value and default parameters; the optimal k-mer was found after running the assembly for various k-mer with an insert length of 32–260 bp. Based on parameters like average length of contigs, coverage, N50, and percentage length greater than 1,000 bp, the best assembly result was selected. TGICL-CAP3 clustering with 90% identity cut-off for each terminal joining of assembled transcripts and terminal length criteria of 40 bp for patching up of scaffolds was used. The resultant final contigs and singletons which did not fit TGICL-CAP3 criteria were then combined to serve as inputs for CD-HIT with a cut-off of 95% similarity. Based on clustering using CD-HIT-EST to eliminate redundancy and by using a set of in-house developed python scripts, the contigs that had no sequence similarity were subjected to DS clustering (Gahlan et al., 2012), and finally, the contigs that did not show any remarkable hits were further transformed into six open reading frames (ORFs). Reverse PSI-BLAST (RPS Blast) tool was used to search conserved domain database (CDD) for remaining contigs.

Functional annotation of the assembled transcriptome

The annotation of finally assembled transcripts was done by similarity search against NCBI non-redundant (NR) database using the BLASTX algorithm. Databases like NCBI-nr, Uniprot, Gene Ontology (GO), Enzyme Classification (EC), and Kyoto Encyclopedia of Genes and Genomes (KEGG) were used for homology search. Multiple hits were obtained for maximum unigenes using the Annot8r tool. The top hits were chosen for each unigene based on the *E*-value cut-off of 10^{-05} and the best bit score obtained.

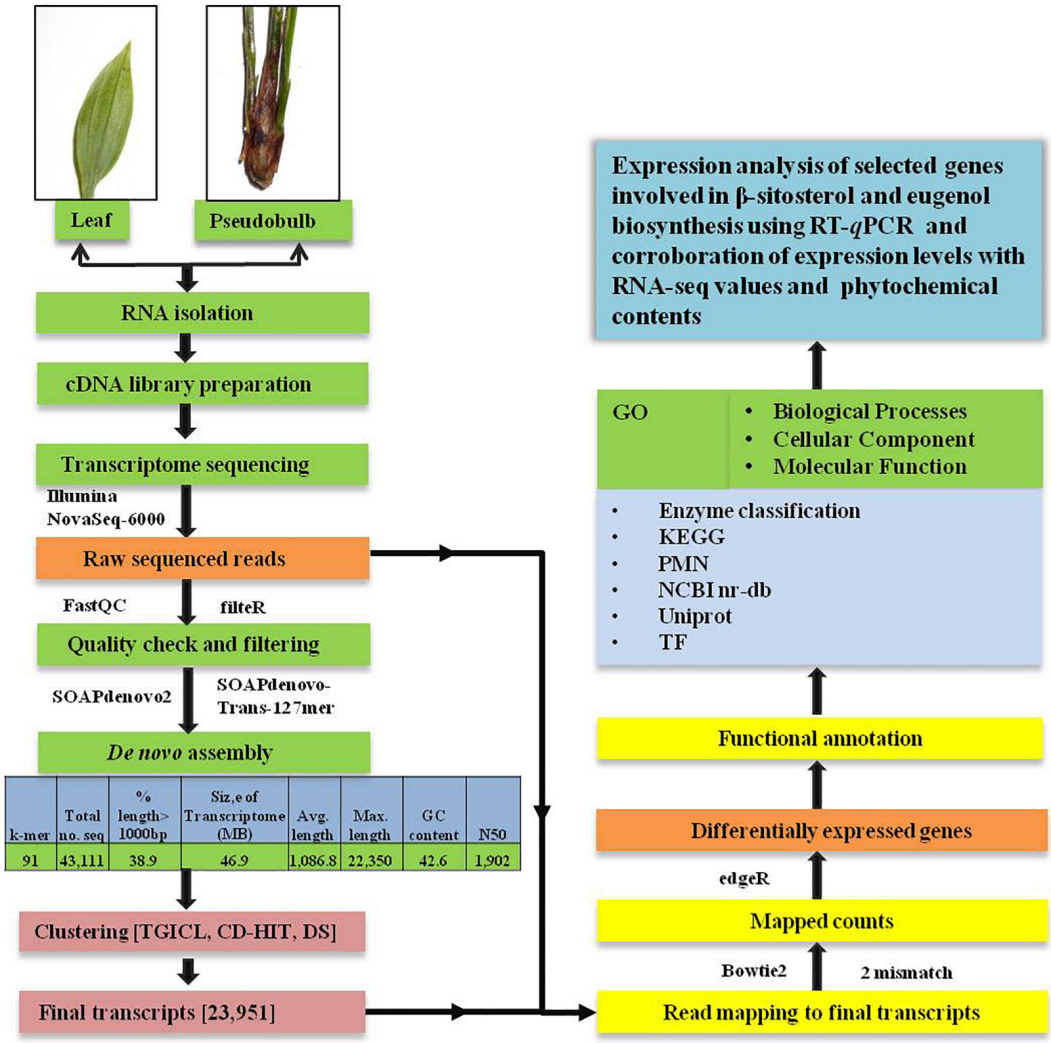


FIGURE 1
Schematic representation of workflow used for *de novo* transcriptome analysis and annotation of high throughput data of *M. acuminata*.

Differential expression analysis

The reads obtained from each of the samples were mapped using Bowtie to count the number of mapped reads having default scope for 2 mismatches. The counts of distinctly mapped reads were obtained using an in-house developed PERL script. For differential expression analysis between two individual plant tissues, i.e., leaf and pseudobulb, EdgeR was employed. The input files for each sample consisted of a matrix of mapped reads which was followed by a one-to-one analysis using a *p*-value at $\leq$0.05 cut-off for the reads from each sample. Finally contigs with log FC $\geq \pm 1$ and *p*-value at <0.05 were considered as differentially expressed unigenes. Expression values in terms of fragments per kilobase per million (FPKM), GO, EC, and KEGG annotations were determined by transforming the count matrix files into FPKM by using in-house developed scripts. Heatmaps and figures were prepared using ggplots, cluster package, and Biobase in R for all the differentially expressed genes (DEGs). Using AgriGOv2.0[1] (Tian et al., 2017), GO enrichment analyses were performed wherein GO enriched categories were identified using the Bonferroni multi-test adjustment method and the hygrogeometric statistical test at the significance level of 0.05. Transcripts encoding for transcription factors (TFs) in leaf and pseudobulb tissues were identified using the plant TFDBv5.0 database[2] (Tian et al., 2020) and the genes involved in different primary and secondary metabolic processes were also identified using Plant Metabolic Network (PMN) database.[3]

1 https://systemsbiology.cau.edu.cn/agriGOv2/c_SEA.php
2 http://planttfdb.cbi.pku.edu.cn/
3 https://plantcyc.org/

Estimation of β-sitosterol and eugenol content in leaf and pseudobulb

β-sitosterol and eugenol contents were estimated using hydrophilic interaction chromatography using an ACQUITY ultra-performance liquid chromatography (UPLC) BEH Shield RP18 (2.1 mm × 100 mm, 1.7 μm) column installed on UPLC system (Waters, India). Two-solvent systems (A; 0.05% formic acid in water and B; 0.1% formic acid in methanol) were used for the separation of samples with gradient programs; 10–20% B at 0–2 min, 20–40% at 2–4 min, 40–20% at 4–7 min, and 20–10% at 7–10 min. The column was equilibrated with the same mobile phase for 1 min at a flow rate of 0.25 μl min^{-1}. The injection volume was 2 μl. Elution was monitored at 254 nm using a photo diode array (PDA) detector with column temperature maintained at 30°C. β-sitosterol and eugenol were identified by comparing their respective retention times and quantified using calibration curves prepared with standard compounds.

Validation of RNA-seq data through reverse transcription quantitative polymerase chain reaction (RT-*q*PCR)

The robustness of transcriptome data was validated by using RT-*q*PCR. The extracted total RNA was reverse transcribed using a cDNA synthesis kit (Thermofisher Scientific, United States), following manufacturer's instructions. Primers for the validation of the selected genes were synthesized using Primer Express 3.0.1 primer design tool (Invitrogen, United States) (**Supplementary Table 1**). RT-*q*PCR was performed with the following conditions: 10 min at 94°C (initial denaturation), 40 cycles of 94°C for 15 s (denaturation), 58–60°C (primer annealing) for 30 s followed by melt curve analysis at 60°C for 30 s (Quantstudio, Applied Biosystems, United States). *Actin* was used as an internal control gene. The expression of the genes was estimated using the $2^{-\Delta\Delta\ CT}$ method (Livak and Schmittgen, 2001), and the values were transformed (log_2) to generate expression profiles.

A schematic diagram representing the workflow for *de novo* analysis of the transcriptome data obtained from leaf and pseudobulb tissues of *M. acuminata* has been provided in **Figure 1**.

Results and discussions

Plant tissue culture of *M. acuminata*

Orchids are primarily valued for their attractive flowers (Bhattacharyya et al., 2019) and also for their use in various herbal systems of medicines (Bhattacharyya et al., 2019; Gantait et al., 2021). Commonly known as "*Jeevak*", the pseudobulb of *M. acuminata* is used in the Indian traditional system of medicine. However, unchecked harvesting of plants from the wild imposes the risk of destruction of patched and scattered natural populations of *M. acuminata*. Keeping this into consideration, a sustainable PTC based protocol was developed to reduce pressure on the fragmented natural plant population. Combined usage of *m*T and chitosan at concentrations of 1.5 and 5.0 mg/L, respectively, induced healthy shoots of *M. acuminata*, which were further transferred to an optimized rooting medium supplemented with IBA and PG at 1.5 and 5.0 mg/L concentrations, respectively (Bhattacharyya et al., 2022; **Figures 2A–E**). The modulatory effect of the plant growth regulators especially *m*T and thidiazuron (TDZ) has been well documented for medicinal orchids *D. nobile*, *D. aphyllum*, *Habenaria edgeworthii*, and *Ansellia africana* (Giri et al., 2012; Bhattacharyya et al., 2016, 2018, 2020). In addition, synergistic application of plant biostimulants and plant growth regulators (PGRs) has been reported to effectively magnify the production of metabolites by stimulating the biosynthetic pathways (Conrath et al., 1989; Baskaran et al., 2016). In the present report, the combined impact of *m*T and chitosan has been studied for the first time. The study also elucidates how various tissues of an orchid can be effectively utilized under controlled *in vitro* conditions for the synthesis of desired biomolecules. Since no reports exist on the elucidation of biosynthetic pathways involved in secondary metabolite biosynthesis in *M. acuminata* and their subsequent corroboration with metabolite profiles, the present work endeavors to understand the involvement of the vital genes in metabolite biosynthesis through tissue-specific transcriptome analysis. The approach also opens an avenue for genetic manipulation *via* metabolic engineering.

Transcriptome sequencing and *de novo* assembly

Tissue-specific transcriptome analysis is an established approach to gain valuable insights into the molecular mechanisms of secondary metabolite biosynthesis (Gantait et al., 2021). However, limited transcriptome data is available for orchids except for dendrobes (Dhiman et al., 2019). In *D. huoshanense*, the genes associated with the secondary metabolite biosynthesis in the leaf, stem and roots were identified by transcriptome analysis (Zhou et al., 2020). Transcriptome profiling revealed genes involved in isoflavonoid biosynthesis in different tissues of *Pueraria lobata*, an orchid species of medicinal importance (Wang et al., 2021). Being a medicinal orchid of biopharmaceutical importance, tissue-specific transcriptome profiling of *M. acuminata* was crucial to identify key genes and regulatory mechanisms governing biosynthesis

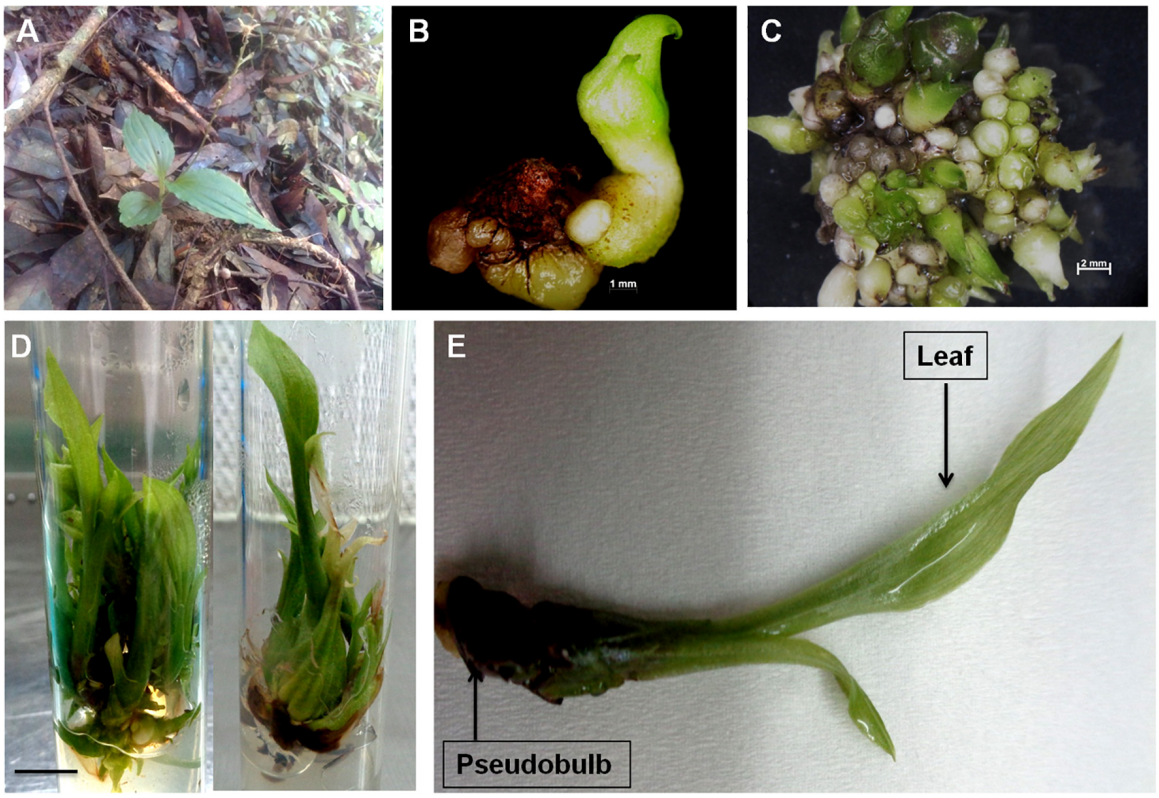


FIGURE 2
(A) Wild plants of *M. acuminata* growing in natural habitat; **(B)** initiation of aseptic cultures using transverse thin cell layer (*t*-TCL) explants (bar = 1 mm); **(C)** proliferation of protocorm like bodies (PLBs) from *t*-TCL explants and initiation of shoot buds (bar = 2 mm); **(D)** multiplication and growth of plants after 8 weeks of culture (bar = 1 cm); and **(E)** full grown plant of *M. acuminata* with differentiated organs.

and differential partitioning of secondary metabolites in *M. acuminata*. Since pseudobulb is widely used in herbal preparations and a previous report on metabolite profiling had revealed promising content of metabolites in leaf tissue of *M. acuminata* (Bose et al., 2017a), cDNA libraries prepared from leaf and pseudobulb tissues of *M. acuminata* plants were sequenced by using NovaSeq 6000 sequencing platform.

For knowing the number of reads, the read types (paired-end), GC content for individual libraries, possible contamination, and other issues, fast QC was used. The final *de novo* assembly of transcripts was obtained using the SOAP *de novo-Trans*-127 mer with default parameters. A total of 43,111 sequences covering 46.9 MB, an average length of 1,086.8 bp, a maximum length of 22,350 bp, GC content of 42.6% and an *N*50 value of 1.9 kb were obtained. These transcripts were clustered into contigs and scaffolds using TGICL and CDHIT followed by DS clustering for the contigs that had no sequence similarity but which belonged to a large gene's part, resulting in 23,951 unigenes (**Supplementary Datasheet 1**). Mapping of the unigenes onto the KEGG pathways facilitated the identification of pathway-specific unigenes involved in the β-sitosterol and eugenol biosynthesis.

Functional annotation of unigenes

Amongst the generated transcriptome dataset, annotation of the assembled sequence reads derived from leaf and pseudobulb tissues of *M. acuminata* revealed that the mapped unigenes were categorized into GO categories. Out of 23,951 transcripts, 69.31% were annotated to cellular component categories, 68.46% to biological process, and 66.63% to molecular function categories. Enzyme classifications were found for 40.20% of transcripts, out of which 38.75% were assigned to biosynthetic pathways. The unannotated transcripts obtained through transcriptome data suggested a lack of information aligning with *M. acuminata*. Similarly, GO analysis underlined the functional heterogeneity amongst the documented transcriptome dataset where the maximum abundance of functional gene transcripts was grouped under cellular component and biological processes category followed by molecular function. The top 20 classes from these three categories are represented in **Figure 3**. In the molecular function category, genes for protein binding, RNA binding and sequence-specific DNA binding transcription factor activity were highly represented. In the biological process category, response to

abscisic acid stimulus, defense response to the bacterium, and regulation of transcription were highly represented. Amongst cellular components, nucleus, plasma membrane, and cytosol were among the predominant categories.

UPLC based estimation of β-sitosterol and eugenol content in leaf and pseudobulb

The medicinal properties of plants are attributed to the occurrence of various types of bioactive compounds of which steroids occupy an important position (Paul et al., 2017; Hu et al., 2022). The steroids are sub-classified into various groups of which phytosterols have attracted the attention of researchers due to their broad-spectrum medicinal properties (Hu et al., 2022). Structurally analogous to cholesterol, phytosterols are of universal occurrence within various members of the plant kingdom including orchids (Singh et al., 2014; Bin Sayeed et al., 2016; Aboobucker and Suza, 2019). Ethnomedicinally, *M. acuminata* is reported to possess bioactive molecules like β-sitosterol and eugenol (Bose et al., 2017a; Suyal et al., 2020). Amongst the two mentioned metabolites, β-sitosterol is a phytosterol that is well known for its therapeutic attributes including anti-cancerous, estrogenic, anti-diabetic, and anti-aging activities whereas eugenol is a phenylpropanoid, which is known to possess anti-inflammatory activity. Realizing the importance of these plant-derived metabolites in the pharmaceutical, food, and cosmetic industries, these can be produced *in vitro* systems through PTC based approach.

In the present study, β-sitosterol and eugenol contents were determined in leaf and pseudobulb tissues of micropropagated *M. acuminata* by using UPLC (**Supplementary Figure 1**). Leaf tissue had promising content of both β-sitosterol (39.69 μg/100 mg) and eugenol (165.94 μg/100 mg) as compared to pseudobulb (β-sitosterol-36.25 μg/100 mg; eugenol-110.92 μg/100 mg), which is traditionally being utilized in herbal preparations (**Figure 4**). The present report is strongly corroborated with the findings of Bose et al. (2017a) in *M. acuminata*. Higher levels of β-sitosterol and eugenol in leaves suggested that leaves rather than pseudobulb may be used for medicinal purpose whereas, pseudobulb can be used for regeneration, thereby helping *in vitro* multiplication and conservation of *M. acuminata*.

Identification and expression analysis of genes involved in β-sitosterol and eugenol biosynthesis

β-sitosterol is a bioactive phytosterol derived from isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) formed *via* plastidic methylerythritol (MEP) and cytosolic mevalonate (MVA) pathways respectively (De-Eknamkul and Potduang, 2003; Suo et al., 2019). The expression pattern of genes involved in the MEP and MVA pathway was analyzed in both leaf and pseudobulb tissues (**Supplementary Datasheet 2**). Several enzymes of the MEP pathway, *1-deoxy-D-xylulose-5-phosphate synthase* (*dxs*), *dimethylallyltranstransferase* (*dmatt*), *4-[cytidine5(prime)-diphospho]-2-C-methyl-D-erythritol kinase* (*ispE*), *2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase* (*ispF*) were downregulated in pseudobulb as compared to that in leaf. *Diphosphomevalonate decarboxylase* (*mvdd*), *3-hydroxy-3-methylglutaryl coenzyme A synthase* (*hmgs*), and *hydroxymethylglutaryl-coenzyme A reductase* (*hmgr*) of the MVA pathway were also downregulated in pseudobulb as compared to that in leaf tissue. *Farnesyl diphosphate synthase* (*fpps*) and *cycloartenol synthase* (*cas*) were downregulated in pseudobulb as compared to that in leaf. Overall downregulation of enzymes involved in β-sitosterol biosynthesis corroborated with lower β-sitosterol content in pseudobulb as compared to that in leaf tissue.

Like phytosterols, eugenol and isoeugenol are bioactive phenylpropanoids synthesized by plants as attractants of pollinators or as defense compounds (Koeduka et al., 2006). Eugenol biosynthesis starts with amino acid phenylalanine which undergoes sequential nine enzyme-catalyzed steps to form eugenol (Singh et al., 2020). The expression of transcripts involved in eugenol biosynthesis was analyzed in leaf and pseudobulb. Amongst all the genes involved in eugenol biosynthesis, the expression of *phenylalanine ammonia-lyase* (*pal*), *cinnamyl alcohol dehydrogenase* (*cad*), and *cinnamyl alcohol reductase* (*ccr*) was similar in both the tissues. Similar expression of these genes was anticipated in both the tissues, as the enzymes coded by these genes play an important role in diverting metabolites toward the biosynthesis of phenylpropanoids and related compounds. However, transcripts encoding for *coumarate CoA ligase* (*4-cl*), *diacylglycerol O-acyltransferase* (*dgat*), and *alcohol acetyltransferase/2-acylglycerolO-acyltransferase* (*acetyl-tag*) were downregulated in pseudobulb as compared to that in leaf. *4-cl* and *dgat* are involved in the specific partitioning of metabolites towards eugenol biosynthesis. Downregulation of these enzymes in pseudobulb as compared to that in leaf tissue corroborated with higher eugenol content in leaf and further suggested that leaf rather than pseudobulb may be used for medicinal preparations.

Analysis of differentially expressed genes in leaf and pseudobulb

The differentially expressed genes in leaf and pseudobulb were analyzed. In the biological process category, transcripts

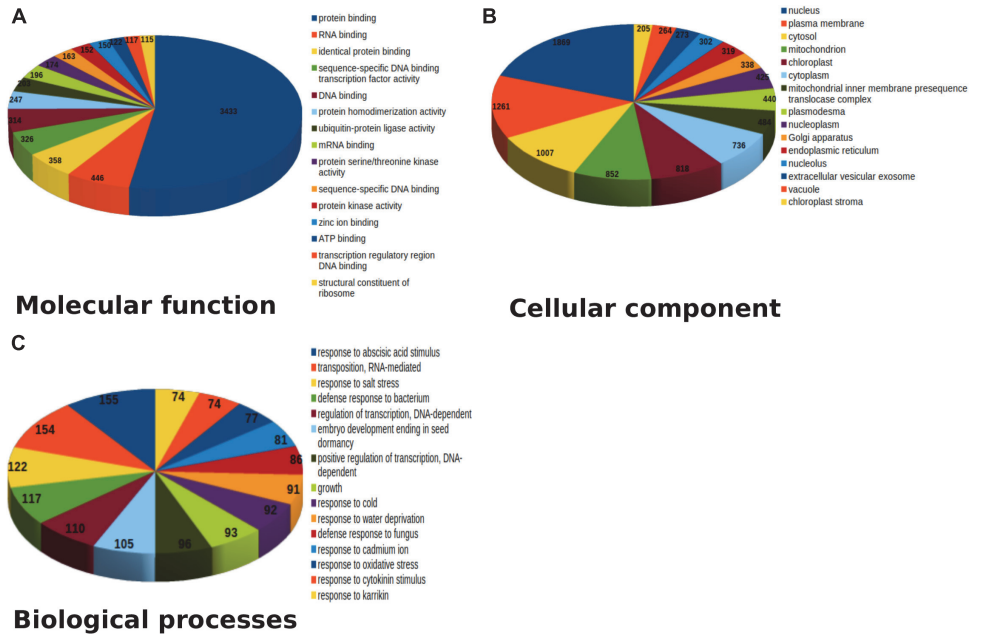


FIGURE 3
Top 15 pathway annotation of transcript in gene ontology (GO) database in *M. acuminata* **(A)** Molecular function; **(B)** Cellular component; **(C)** Biological process.

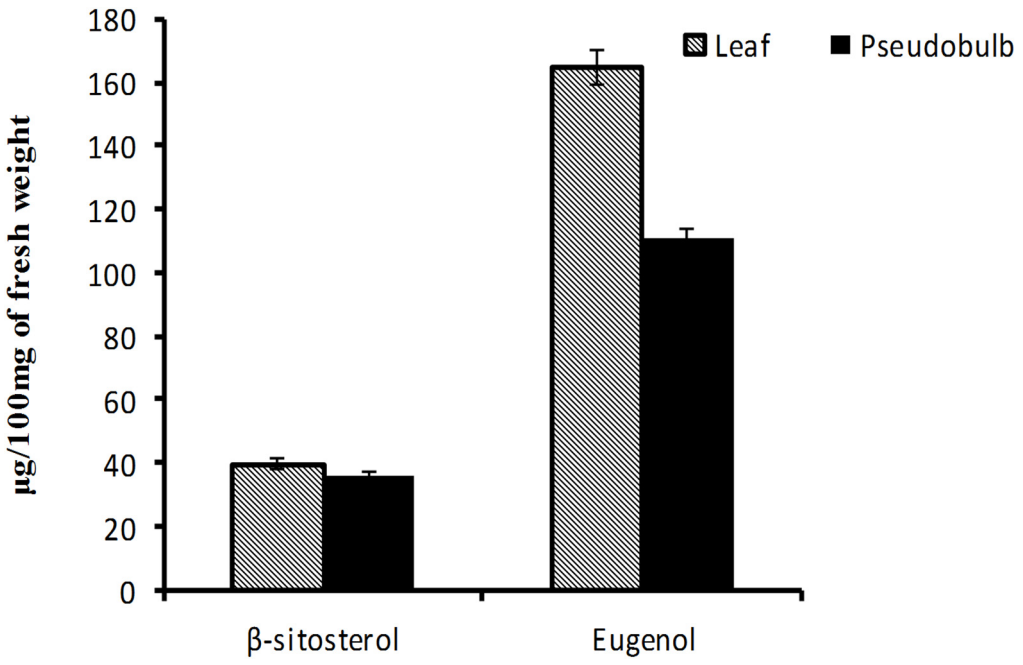


FIGURE 4
Eugenol and β-sitosterol content in leaf and pseudobulb of *in vitro* raised *M. acuminata*. Value in the bar represents mean of three independent biological replicates with standard error of mean.

associated with response to wounding, defense response to fungus, cell growth, and response to abscisic and jasmonic acid stimulus were upregulated in pseudobulb as compared to that in leaf (Figure 5). Pseudobulb is responsible for the regeneration of *in vitro* and naturally occurring populations of *M. acuminata*. An upregulation of defense and plant

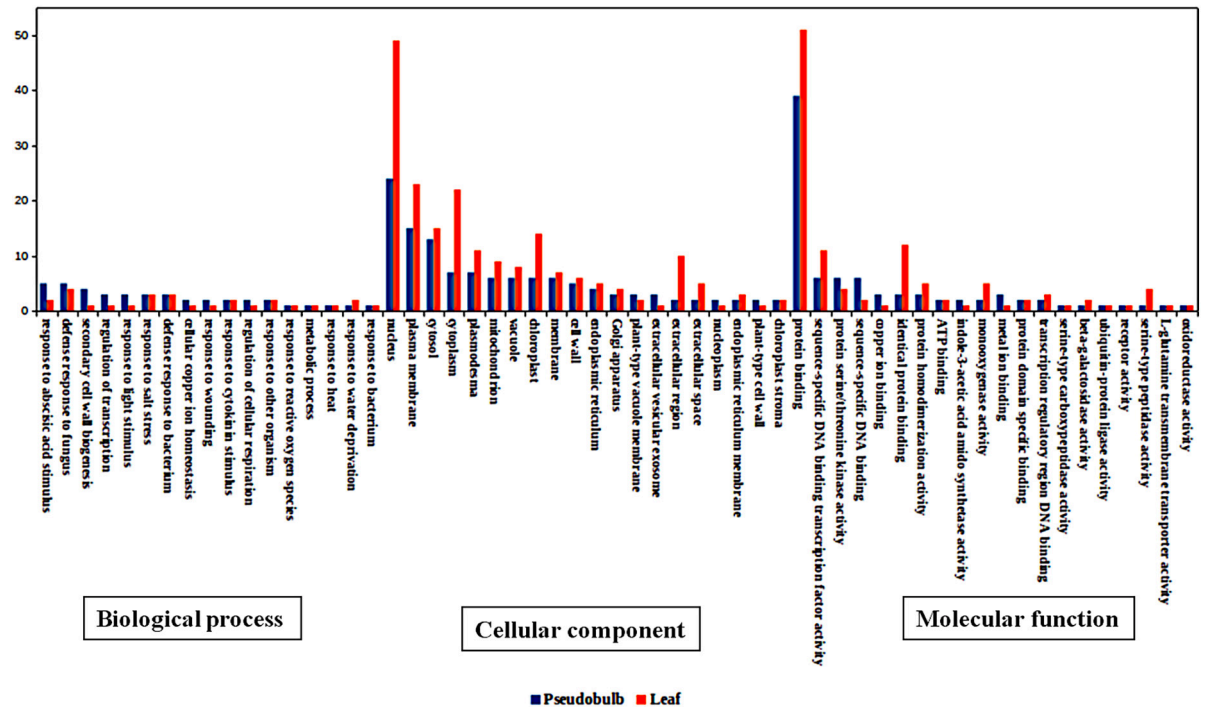


FIGURE 5
Top 20 gene ontology (GO) categories of differentially expressed genes in *M. acuminata* for biological process, cellular component, and molecular function.

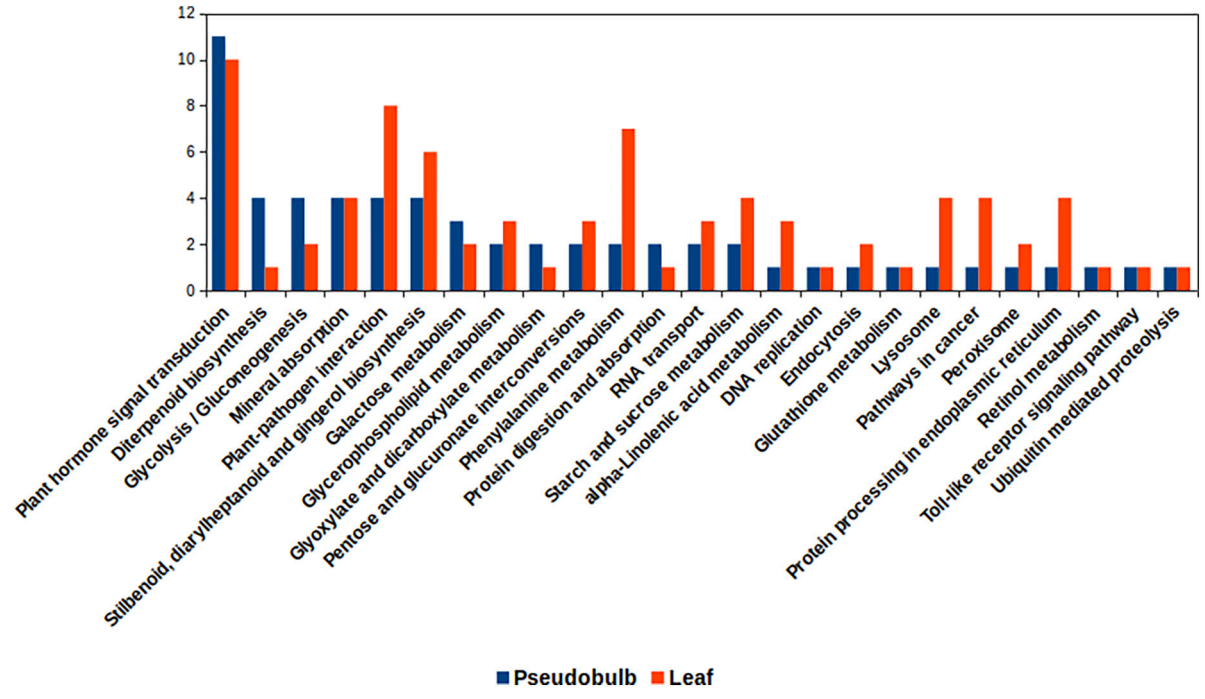


FIGURE 6
Kyoto encyclopedia of genes and genomes (KEGG) pathways associated with differentially expressed genes for leaf and pseudobulb in *M. acuminata*.

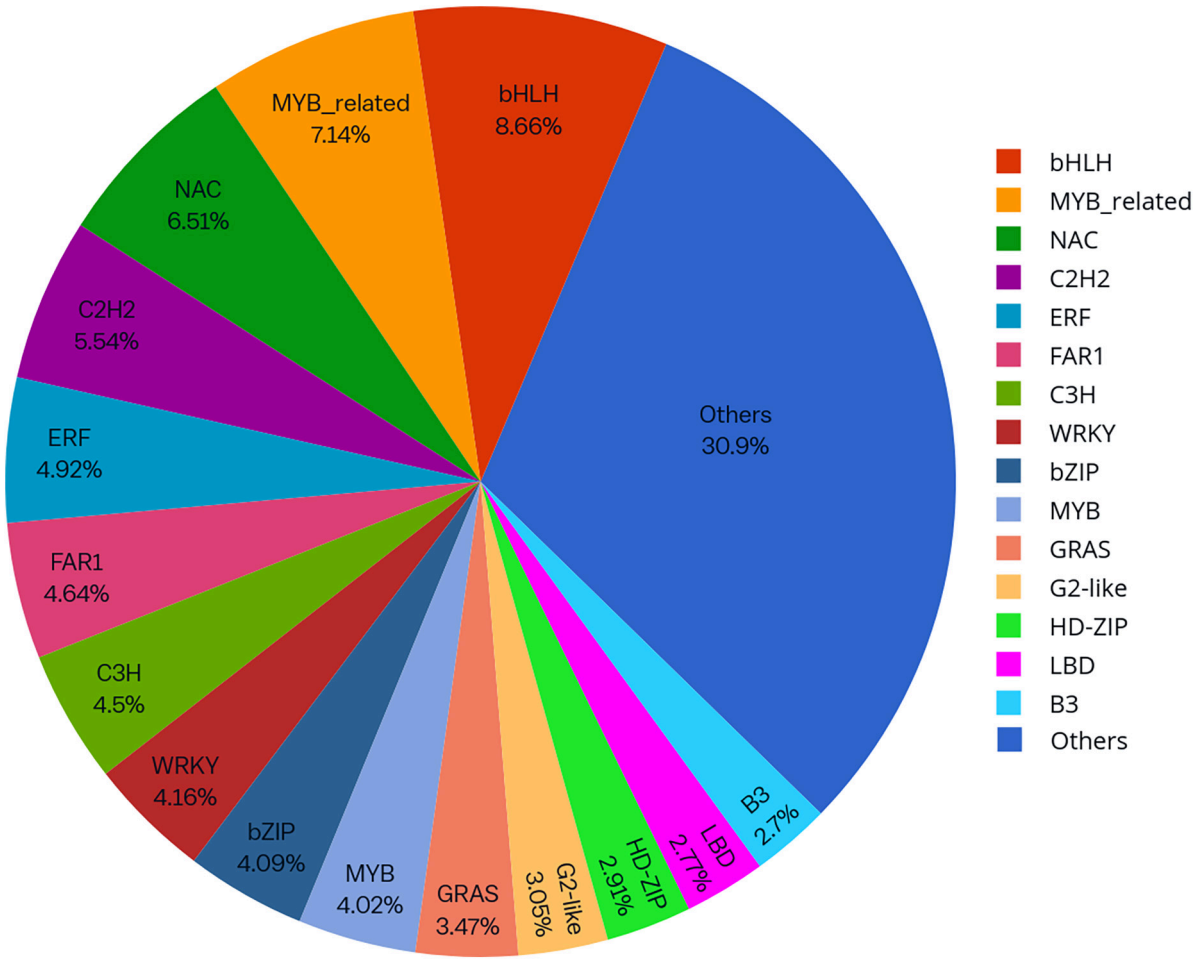


FIGURE 7
Distribution of top transcription factors (TFs) families identified in *M. acuminata*.

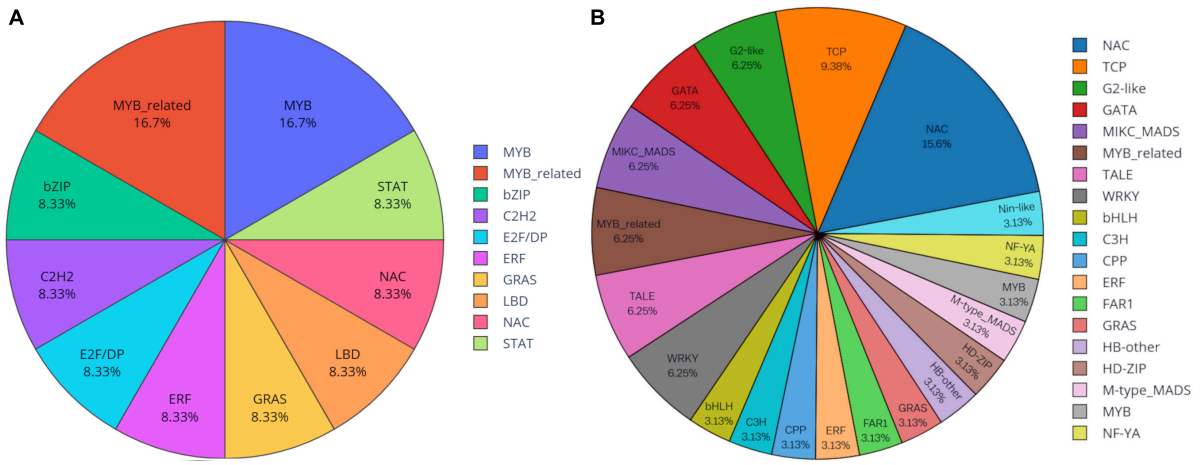


FIGURE 8
Differentially expressed transcription factors (TFs) in **(A)** pseudobulb and **(B)** leaf tissues of *M. acuminata*.

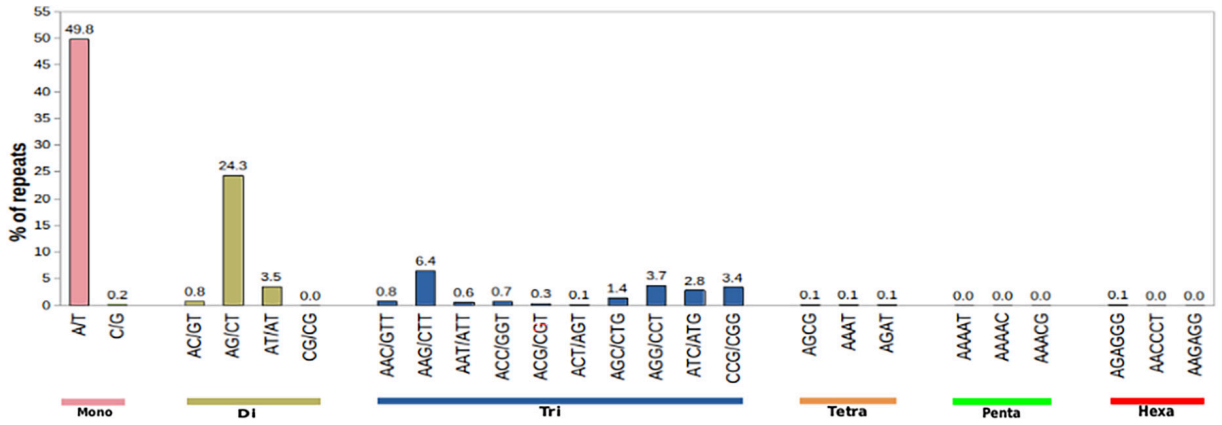


FIGURE 9
Distribution of simple sequence repeats (SSRs) in *M. acuminata* transcriptome.

hormone-related pathways was envisaged as these might play an important role in coping with various environmental stresses. Amongst cellular components, transcripts associated with the nucleus, cell membrane, and chloroplast were downregulated in pseudobulb as compared to that in leaf. In the molecular function category, protein binding and sequence-specific transcription factor binding activity were downregulated in pseudobulb as compared to that in leaf suggesting a high degree of protein interaction and transcription factor-mediated gene regulation in leaf as compared to that in pseudobulb (**Figure 5**).

In the KEGG pathway category, transcripts related to starch and sucrose metabolism, phenylalanine metabolism, stilbenoid, diarylheptanoid, and gingerol biosynthesis were upregulated in leaf as compared to that in pseudobulb (**Figure 6**). Heatmaps depicted expression pattern of unigenes associated with differentially expressed KEGG pathways (**Supplementary Figure 2** and **Supplementary Datasheet 3**). Phenylalanine being a substrate for biosynthesis of phenylpropanoids and related compounds, an upregulation of transcripts associated with phenylalanine metabolism in leaf corroborated with higher eugenol content in leaves. However, transcripts related to plant hormone signal transduction and diterpenoid biosynthesis were upregulated in pseudobulb as compared to that in the leaf. Upregulation of hormone signal transduction pathways in pseudobulb over leaf suggested an important role of phytohormones in mediating *in vitro* regeneration of *M. acuminata* plantlets from pseudobulb explants. The role of plant in hormones *in vitro* regeneration of plants is well documented. Also, chemicals like chitosan and *m*T have been reported to modulate phytohormone levels in plants (Bhattacharyya et al., 2021). Overall differential expression across the various KEGG pathways in leaf and pseudobulb suggested tissue-specific modulation of metabolic processes in *M. acuminata.*

Transcription factors

Knowledge of transcriptional regulation of secondary metabollite biosynthesis, along with an explanation of the biosynthetic pathways is much anticipated (Chezem and Clay, 2016; Singh et al., 2017). TFs regulate the expression of target genes by binding to specific *cis*-regulatory elements located in the promoter region of genes eventually regulating a vast network of biological and metabolic processes (Patra et al., 2013). In the present study, a total of 1,443 unigenes belonging to 56 TF families were identified as TFs (**Figure 7**). Analysis of differentially expressed gene datasets in leaf and pseudobulb of *M. acuminata* revealed that as compared to pseudobulb, higher numbers of TFs were expressed in leaf suggesting a high level of TF mediated gene regulation in the leaf. However, MYB, bHLH, GRAS, NAC, and ERF families were well represented in both leaf and pseudobulb tissue of *M. acuminata* (**Figure 8** and **Supplementary Datasheet 4**). Amongst the various classes of TFs, MYB is one of the largest transcriptional regulators playing a key role in the regulation of secondary metabolism in medicinal plant species (Cao et al., 2020). Apart from MYB, the role of GRAS, bHLH, ERF, and NAC TFs in the regulation of biosynthesis of secondary metabolites is also well documented (Rohrmann et al., 2011; Zhao et al., 2013; Chezem and Clay, 2016; Duraisamy et al., 2016; Fu et al., 2017). Enrichment of these transcripts in both the tissues suggests their involvement in the biosynthesis of valuable metabolites (**Figure 8**).

Further, profiling of target metabolites (β-sitosterol and eugenol) by UPLC suggested that apart from predominance in the pseudobulb, metabolites were also present and accumulated in leaf tissues which are supported by the findings of Bose et al. (2017a) in *M. acuminata*. The present finding suggested that from pseudobulb, metabolites are possibly transported to the other tissues such as the leaf. Furthermore, the metabolite production

can be further increased by the administration of a suitable elicitor in cell suspension cultures derived from leaf/pseudobulb protocorm like bodies (PLBs) or callus. This will pave the path for both conservations and commercial utilization. Similar findings were reported for *Achyranthes bidentata* and *T. govanianum,* also (Li et al., 2016; Singh et al., 2017).

Identification of simple sequence repeats

Apart from providing critical insights into the pathways of secondary metabolite biosynthesis in medicinal plants, transcriptome sequencing also facilitates the detection of 2–6 base pair tandem repeat nucleotide motifs defined as SSR or microsatellites. EST-SSRs play a vital role in genetic breeding by providing a wealth of molecular markers and can be used for the construction of genetic maps in plants (Varshney et al., 2005a,b; Ellis and Burke, 2007; Kalia et al., 2011). These have been used as a resource for candidate markers in population genetic studies. SSRs were mined using the microsatellite identification tool (MISA) program with the following parameters: di- ≥ 6 nt, tri- ≥ 5 nt, tetra- ≥ 5 nt, penta- ≥ 4 nt, and hexa- ≥ 4 nt. A total of 5,370 SSRs were identified in 23,951 transcripts. Amongst the identified SSRs, the most abundant were mononucleotide repeat (11.20%) followed by di (6.43%) and tri (4.52%) nucleotide repeat types (**Figure 9** and **Supplementary Table 2**). To date, no data exist on EST/SSRs and microsatellites of *M. acuminata* to the best of our knowledge. Being an endangered terrestrial orchid species of broad-spectrum therapeutic importance, knowledge of genetic diversity is pivotal in the formulation of effective conservation strategies for its sustainable utilization (Bhattacharyya and Van Staden, 2018). Therefore, the identification and development of highly polymorphic EST-SSRs can be used for distinguishing closely related populations of *M. acuminata* and could thereby aid in the identification of superior strains for improvement.

Validation of transcriptome data by RT-*q*PCR

Reverse transcription quantitative polymerase chain reaction was used to validate the expression of genes involved in β-sitosterol and eugenol biosynthesis as obtained from transcriptome data. The results revealed a significant correlation between the expression of genes obtained by RT-*q*PCR and RNA-seq data (**Figure 10**). Further, the expression patterns of the selected genes were correlated with the assayed metabolite fractions, i.e. β-sitosterol and eugenol. Higher expression of genes involved in β-sitosterol and eugenol biosynthesis and relatively higher content of target metabolites (β-sitosterol and eugenol) in leaf tissue as compared to pseudobulb was obtained. Data suggested that leaf tissues rather than pseudobulb can be used as an alternate source of bioactive metabolites in *M. acuminata*.

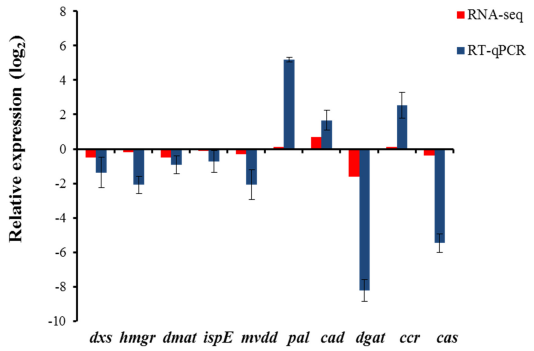


FIGURE 10
Expression of genes involved in biosynthesis of β-sitosterol and eugenol in *M. acuminata*. Relative expression ratios in pseudobulb have been calculated with reference to the leaf tissue. The relative expression values were $\log_2$ transformed. *dxs*, 1-deoxy-D-xylulose-5-phosphate synthase, *hmgr, hydroxy 3 methylglutaryl coenzyme A reductase*, *dmat, dimethylallyltranstransferase*, *ispE*, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase, *mvdd, diphosphomevalonate decarboxylase*, *pal*, *phenylalanine ammonia lyase*, *cad*, cinnamyl alcohol dehydrogenase, *dgat, diacylglycerol O-acyltransferase*, *ccr*, *cinnamoyl coA reductase*, *cas*, *cycloartenol synthase*. *Actin* was used as endogenous control to normalize the data. Each value in the bar diagram represents mean of three independent biological replicates and error bar represents standard error of mean.

Conclusion

Terrestrial orchids are largely unexplored medicinal plant species. The present report provides the first-ever insights into the tissue-specific transcriptome of *M. acuminata* (*Jeevak*), a key ingredient of *Ashtvarga* in *Ayurveda*. The *de novo* assembled transcriptome provides a comprehensive resource information for further research aimed at enhancing secondary metabolite biosynthesis in *M. acuminata*. The research findings also firmly establish the utility of *in vitro* propagules in providing an alternative source of metabolite prospection. The expression of genes involved in β-sitosterol and eugenol biosynthesis in leaf and pseudobulb corroborated with target metabolite contents and substantiates the utility of leaf tissues as a valuable alternative source for metabolites in *M. acuminata*. The present research approach can serve as a research model for the PTC-mediated conservation and RNA sequencing aided gene mining in non-model medicinal plants, especially orchids. Functional characterization of regulatory genes/enzymes/TFs involved in secondary metabolite biosynthetic pathways could facilitate a metabolic engineering program in *M. acuminata*.

Data availability statement

The original contributions presented in this study are publicly available. This data can be found here: NCBI, PRJNA842241.

Author contributions

PB contributed to PTC of *M. acuminata*, RNA extraction library preparation, and drafting of the manuscript. TS performed the RT-*q*PCR gene expression analysis, data analysis, and drafting of the manuscript. AY performed the bioinformatics analysis. LL assisted in RNA extraction and gene expression analysis. Ritu assisted in uploading the transcriptome data in SRA database also assisted in revising and rechecking of the bioinformatics data presented in the research manuscript. MS performed the transcriptome sequencing and assisted in library preparation. RJ performed UPLC profiling and helped in analyzing the phytochemical data. RS supervised bioinformatics analysis and finalized the manuscript. SK planning, conceptualization, overall supervision, and finalization of the manuscript.

Funding

PB acknowledges the Science and Education Research Board (SERB), GOI for providing financial support in the form of National Postdoctoral Fellowship (N-PDF) viding file number: PDF/2016/000340. RS was thankful to the Department of Biotechnology (DBT) for providing support to HiCHiCoB as DBT BIC.

Acknowledgments

The authors acknowledge Council for Scientific and Industrial Research (CSIR) for logistic and instrumentation support. The authors acknowledge Nitesh Kumar Sharma for sharing technical expertise in the significant improvisation of this manuscript. PB also acknowledges Dr. Prasenjit Paul, Lecturer, St Edmund's College, Shillong, Meghalaya, India for helping in the collection and identification of the plant germplasm from wild habitats. CSIR-IHBT publication number for this manuscript is 5118.

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.2022.954467/full#supplementary-material

References

Aboobucker, S. I., and Suza, W. P. (2019). Why do plants convert sitosterol to stigmasterol? *Front. Plant Sci.* 10:354. doi: 10.3389/fpls.2019.00354

Balkrishna, A., Mishra, R. K., Sharma, N., Sharma, V. K., and Misra, L. (2018). Phytochemical, botanical and biological paradigm of astavarga plants- The Ayurvedic rejuvenators. *J. Nat. Ayurvedic Med.* 18:6.

Baskaran, P., Kumari, A., Ncube, B., and Van Staden, J. (2016). Acetylcholinesterase-inhibition and antibacterial activity of *Mondia whitei* adventitious roots and *ex vitro*-grown somatic embryogenic-biomass. *Front. Pharmacol.* 7:335. doi: 10.3389/fphar.2016.00335

Bhattacharyya, P., Kumaria, S., and Tandon, P. (2016). High frequency regeneration protocol for *Dendrobium nobile*: A model tissue culture approach for propagation of medicinally important orchid species. *South Afr. J. Bot.* 104, 232–243. doi: 10.1016/j.sajb.2015.11.013

Bhattacharyya, P., Paul, P., Kumaria, S., and Tandon, P. (2018). Transverse thin cell layer (*t*-TCL)-mediated improvised micropropagation protocol for endangered medicinal orchid *Dendrobium aphyllum* Roxb: An integrated phytomolecular approach. *Acta Physiol. Plant.* 40:137. doi: 10.1007/s11738-018-2703-y

Bhattacharyya, P., Kumar, V., Grúz, J., Doležal, K., and Van Staden, J. (2019). Deciphering the phenolic acid reserves and antioxidant activity within the protocorm like bodies of *Ansellia africana*: A vulnerable medicinal orchid. *Ind. Crops Prod.* 135, 21–29. doi: 10.1016/j.indcrop.2019.03.024

Bhattacharyya, P., Gupta, S., and Van Staden, J. (2020). "Phytochemistry, pharmacology, and conservation of *Ansellia africana*: A vulnerable medicinal orchid of Africa," in *Orchids Phytochemistry, Biology and Horticulture. Reference Series in Phytochemistry*, eds J. M. Merillon and H. Kodja (Cham: Springer). doi: 10.1007/978-3-030-11257-8_17-1

Bhattacharyya, P., Lalthafamkimi, L., and Van Staden, J. (2021). "Insights into the biostimulatory effects of chitosan in propagation of orchid bioresources," in *Biostimulants for Crops from Seed Germination to Plant Development: A Practical Approach*, eds S. Gupta and J. Van Staden (London: Academic Press), 197–210. doi: 10.1016/B978-0-12-823048-0.00020-4

Bhattacharyya, P., Lalthafamkimi, L., Sharma, T., Kumar, D., Kumar, A., Kumar, D., et al. (2022). Metabolic and tissue-specific expression profiling in micropropagated plants of *Malaxis acuminata*: An endangered medicinal orchid. *Plant Cell Tissue Organ. Culture.* doi: 10.1007/s11240-022-02369-3

Bhattacharyya, P., and Van Staden, J. (2018). Molecular insights into genetic diversity and population dynamics of five medicinal *Eulophia* species: A threatened orchid taxa of Africa. *Physiol. Mol. Biol. Plants* 24, 631–641. doi: 10.1007/s12298-018-0523-6

Bin Sayeed, M., Karim, S., Sharmin, T., and Morshed, M. (2016). Critical analysis on characterization, systemic effect, and therapeutic potential of β-Sitosterol: A plant-derived orphan phytosterol. *Medicines* 3:29. doi: 10.3390/medicines3040029

Bose, B., Choudhury, H., Tandon, P., and Kumaria, S. (2017a). Studies on secondary metabolite profiling, anti-inflammatory potential, *in vitro* photoprotective and skin-aging related enzyme inhibitory activities of *Malaxis acuminata*, a threatened orchid of nutraceutical importance. *J. Photochem. Photobiol. B Biol.* 173, 686–695. doi: 10.1016/j.jphotobiol.2017.07.010

Bose, B., Kumaria, S., Choudhury, H., and Tandon, P. (2017b). Insights into nuclear DNA content, hydrogen peroxide and antioxidative enzyme activities during transverse thin cell layer organogenesis and *ex vitro* acclimatization of *Malaxis wallichii*, a threatened medicinal orchid. *Physiol. Mol. Biol. Plants* 23, 955–968. doi: 10.1007/s12298-017-0474-3

Cao, Y., Li, K., Li, Y., Zhao, X., and Wang, L. (2020). MYB transcription factors as regulators of secondary metabolism in plants. *Biology (Basel)* 9:61. doi: 10.3390/biology9030061

Cheruvathur, M. K., Abraham, J., Mani, B., and Thomas, T. D. (2010). Adventitious shoot induction from cultured internodal explants of *Malaxis acuminata* D. Don, a valuable terrestrial medicinal orchid. *Plant Cell Tissue Organ Cult.* 101, 163–170. doi: 10.1007/s11240-010-9673-0

Chezem, W. R., and Clay, N. K. (2016). Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. *Phytochemistry* 131, 26–43. doi: 10.1016/j.phytochem.2016.08.006

Conrath, U., Domard, A., and Kauss, H. (1989). Chitosan-elicited synthesis of callose and of coumarin derivatives in parsley cell suspension cultures. *Plant Cell Rep.* 8, 152–155. doi: 10.1007/BF00716829

De-Eknamkul, W., and Potduang, B. (2003). Biosynthesis of β-sitosterol and stigmasterol in *Croton sublyratus* proceeds via a mixed origin of isoprene units. *Phytochemistry* 62, 389–398. doi: 10.1016/S0031-9422(02)00555-1

Dhiman, N., Sharma, N. K., Thapa, P., Sharma, I., Kumar Swarnkar, M., Chawla, A., et al. (2019). *De novo* transcriptome provides insights into the growth behaviour and resveratrol and *trans-stilbenes* biosynthesis in *Dactylorhiza hatagirea*-an endangered alpine terrestrial orchid of western Himalaya. *Sci. Rep.* 9:13133. doi: 10.1038/s41598-019-49446-w

Dogan, H., Ercişli, S., Temim, E., Hadziabulic, A., Tosun, M., Yilmaz, S., et al. (2014). Diversity of chemical content and biological activity in flower buds of a wide number of wild grown caper (*Capparis ovata* Desf.) genotypes from Turkey. *Comptes Rendus L Acad. Bulg. Des Sci.* 67, 1593–1600.

Duraisamy, G. S., Mishra, A. K., Kocabek, T., and Matoušek, J. (2016). Identification and characterization of promoters and cis-regulatory elements of genes involved in secondary metabolites production in hop (*Humulus lupulus*. L). *Comput. Biol. Chem.* 64, 346–352. doi: 10.1016/j.compbiolchem.2016.07.010

Ellis, J. R., and Burke, J. M. (2007). EST-SSRs as a resource for population genetic analyses. *Heredity (Edinb.)* 99, 125–132. doi: 10.1038/sj.hdy.6801001

Espinosa-Leal, C. A., Puente-Garza, C. A., and García-Lara, S. (2018). *In vitro* plant tissue culture: Means for production of biological active compounds. *Planta* 248, 1–18. doi: 10.1007/s00425-018-2910-1

Fu, C., Han, Y., Kuang, J., Chen, J., and Lu, W. (2017). Papaya CpEIN3a and CpNAC2 co-operatively regulate carotenoid biosynthesis-related genes CpPDS2/4, CpLCY-e and CpCHY-b during fruit ripening. *Plant Cell Physiol.* 58, 2155–2165. doi: 10.1093/pcp/pcx149

Gahlan, P., Singh, H. R., Shankar, R., Sharma, N., Kumari, A., Chawla, V., et al. (2012). *De novo* sequencing and characterization of *Picrorhiza kurrooa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics* 13:126. doi: 10.1186/1471-2164-13-126

Gai, S., Zhang, Y., Mu, P., Liu, C., Liu, S., Dong, L., et al. (2012). Transcriptome analysis of tree peony during chilling requirement fulfillment: Assembling, annotation and markers discovering. *Gene* 497, 256–262. doi: 10.1016/j.gene.2011.12.013

Gantait, S., Das, A., Mitra, M., and Chen, J.-T. (2021). Secondary metabolites in orchids: Biosynthesis, medicinal uses, and biotechnology. *South Afr. J. Bot.* 139, 338–351. doi: 10.1016/j.sajb.2021.03.015

Giri, L., Belwal, T., Bahukhandi, A., Suyal, R., Bhatt, I. D., Rawal, R. S., et al. (2017). Oxidative DNA damage protective activity and antioxidant potential of Ashtvarga species growing in the Indian Himalayan Region. *Ind. Crops Prod.* 102, 173–179. doi: 10.1016/j.indcrop.2017.03.023

Giri, L., Dhyani, P., Rawat, S., Bhatt, I. D., Nandi, S. K., Rawal, R. S., et al. (2012). *In vitro* production of phenolic compounds and antioxidant activity in callus suspension cultures of *Habenaria edgeworthii*: A rare Himalayan medicinal orchid. *Ind. Crops Prod.* 39, 1–6. doi: 10.1016/j.indcrop.2012.01.024

Golechha, M. (2020). Time to realise the true potential of Ayurveda against COVID-19. *Brain Behav. Immun.* 87, 130–131. doi: 10.1016/j.bbi.2020.05.003

Han, R., Rai, A., Nakamura, M., Suzuki, H., Takahashi, H., Yamazaki, M., et al. (2016). *De novo* deep transcriptome analysis of medicinal plants for gene discovery in biosynthesis of plant natural products. *Methods Enzymol.* 576, 19–45. doi: 10.1016/bs.mie.2016.03.001

He, C., Liu, X., Teixeira da Silva, J. A., Liu, N., Zhang, M., and Duan, J. (2020). Transcriptome sequencing and metabolite profiling analyses provide comprehensive insight into molecular mechanisms of flower development in *Dendrobium officinale* (Orchidaceae). *Plant Mol. Biol.* 104, 529–548. doi: 10.1007/s11103-020-01058-z

Hossain, M. M. (2011). Therapeutic orchids: Traditional uses and recent advances—an overview. *Fitoterapia* 82, 102–140. doi: 10.1016/j.fitote.2010.09.007

Hu, Y., Suo, J., Jiang, G., Shen, J., Cheng, H., Lou, H., et al. (2022). The effect of ethylene on squalene and β-sitosterol biosynthesis and its key gene network analysis in *Torreya grandis* nuts during post-ripening process. *Food Chem.* 368:130819. doi: 10.1016/j.foodchem.2021.130819

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., and Dhawan, A. K. (2011). Microsatellite markers: An overview of the recent progress in plants. *Euphytica* 177, 309–334. doi: 10.1007/s10681-010-0286-9

Kaur, S., and Bhutani, K. K. (2010). Micropropagation of *Malaxis acuminata* D. Don: A rare orchid of high therapeutic value. *Open Access J. Med. Aromat. Plants* 1, 29–33.

Kibar, B. (2021). Influence of different drying methods and cold storage treatments on the postharvest quality and nutritional properties of *P. ostreatus* mushroom. *Turk. J. Agric. For.* 45, 565–579. doi: 10.3906/tar-2102-76

Koeduka, T., Fridman, E., Gang, D. R., Vassão, D. G., Jackson, B. L., Kish, C. M., et al. (2006). Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10128–10133. doi: 10.1073/pnas.0603732103

Li, J., Wang, C., Han, X., Qi, W., Chen, Y., Wang, T., et al. (2016). Transcriptome analysis to identify the putative biosynthesis and transport genes associated with the medicinal components of *Achyranthes bidentata* Bl. *Front. Plant Sci.* 7:1860. doi: 10.3389/fpls.2016.01860

Li, Y., Luo, H.-M., Sun, C., Song, J.-Y., Sun, Y.-Z., Wu, Q., et al. (2010). EST analysis reveals putative genes involved in glycyrrhizin biosynthesis. *BMC Genomics* 11:268. doi: 10.1186/1471-2164-11-268

Li, Z., and Trick, H. N. (2005). Rapid method for high-quality RNA isolation from seed endosperm containing high levels of starch. *Biotechniques* 38, 872–876. doi: 10.2144/05386BM05

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lohani, N., Tewari, L. M., Joshi, G. C., Kumar, R., Kishor, K., and Upreti, B. M. (2013). Population assessment and threat categorization of endangered medicinal orchid *Malaxis acuminata* D. Don. from north-west Himalaya. *Int. J. Conserv. Sci.* 4, 483–492.

Mollova, S., Fidan, H., Antonova, D., Bozhilov, D., Stanev, S., Kostova, I., et al. (2020). Chemical composition and antimicrobial and antioxidant activity of *Helichrysum italicum* (Roth) G. Don subspecies essential oils. *Turk. J. Agric. For.* 44, 371–378. doi: 10.3906/tar-1909-34

Morozova, O., Hirst, M., and Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annu. Rev. Genomics Hum. Genet.* 10, 135–151. doi: 10.1146/annurev-genom-082908-145957

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Nalawade, S. M., Sagare, A. P., Lee, C.-Y., Kao, C.-L., and Tsay, H.-S. (2003). Studies on tissue culture of Chinese medicinal plant resources in Taiwan and their sustainable utilization. *Bot. Bull. Acad. Sin* 44, 79–98.

Niazian, M. (2019). Application of genetics and biotechnology for improving medicinal plants. *Planta* 249, 953–973. doi: 10.1007/s00425-019-03099-1

Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S., and Yuan, L. (2013). Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta (BBA) Gene Regul. Mech.* 1829, 1236–1247. doi: 10.1016/j.bbagrm.2013.09.006

Paul, P., Joshi, M., Gurjar, D., Shailajan, S., and Kumaria, S. (2017). *In vitro* organogenesis and estimation of β-sitosterol in *Dendrobium fimbriatum* Hook.: An orchid of biopharmaceutical importance. *South African J. Bot.* 113, 248–252. doi: 10.1016/j.sajb.2017.08.019

Rohrmann, J., Tohge, T., Alba, R., Osorio, S., Caldana, C., McQuinn, R., et al. (2011). Combined transcription factor profiling, microarray analysis and metabolite profiling reveals the transcriptional control of metabolic shifts occurring during tomato fruit development. *Plant J.* 68, 999–1013. doi: 10.1111/j.1365-313X.2011.04750.x

Shan, T., Yin, M., Wu, J., Yu, H., Liu, M., Xu, R., et al. (2021). Comparative transcriptome analysis of tubers, stems, and flowers of *Gastrodia elata* Blume reveals potential genes involved in the biosynthesis of phenolics. *Fitoterapia* 153:104988. doi: 10.1016/j.fitote.2021.104988

Sharma, T., Sharma, N. K., Kumar, P., Panzade, G., Rana, T., Swarnkar, M. K., et al. (2021). The first draft genome of *Picrorhiza kurrooa*, an endangered medicinal herb from Himalayas. *Sci. Rep.* 11:14944. doi: 10.1038/s41598-021-93495-z

Shih, M.-L., and Morgan, J. A. (2020). Metabolic flux analysis of secondary metabolism in plants. *Metab. Eng. Commun.* 10:e00123. doi: 10.1016/j.mec.2020.e00123

Simon, S. A., Zhai, J., Nandety, R. S., McCormick, K. P., Zeng, J., Mejia, D., et al. (2009). Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.* 60, 305–333. doi: 10.1146/annurev.arplant.043008.092032

Singh, P., Kalunke, R. M., Shukla, A., Tzfadia, O., Thulasiram, H. V., and Giri, A. P. (2020). Biosynthesis and tissue-specific partitioning of camphor and eugenol in *Ocimum kilimandscharicum*. *Phytochemistry* 177:112451. doi: 10.1016/j.phytochem.2020.112451

Singh, P., Singh, G., Bhandawat, A., Singh, G., Parmar, R., Seth, R., et al. (2017). Spatial transcriptome analysis provides insights of key gene (s) involved in steroidal saponin biosynthesis in medicinally important herb *Trillium govanianum*. *Sci. Rep.* 7:45295. doi: 10.1038/srep45295

Singh, S., Pal, S., Shanker, K., Chanotiya, C. S., Gupta, M. M., Dwivedi, U. N., et al. (2014). Sterol partitioning by HMGR and DXR for routing intermediates toward withanolide biosynthesis. *Physiol. Plant.* 152, 617–633. doi: 10.1111/ppl.12213

Sun, C., Li, Y., Wu, Q., Luo, H., Sun, Y., Song, J., et al. (2010). *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics* 11:262. doi: 10.1186/1471-2164-11-262

Suo, J., Tong, K., Wu, J., Ding, M., Chen, W., Yang, Y., et al. (2019). Comparative transcriptome analysis reveals key genes in the regulation of squalene and β-sitosterol biosynthesis in *Torreya grandis*. *Ind. Crops Prod.* 131, 182–193. doi: 10.1016/j.indcrop.2019.01.035

Suyal, R., Rawat, S., Rawal, R. S., and Bhatt, I. D. (2020). "A review on phytochemistry, nutritional potential, pharmacology, and conservation of *Malaxis acuminata*: An orchid with rejuvenating and vitality strengthening properties," in *Orchids phytochemistry, biology and horticulture. Reference series in phytochemistry*, eds J. M. Mérillon and H. Kodja (Cham: Springer). doi: 10.1007/978-3-030-11257-8_15-1

Tian, F., Yang, D.-C., Meng, Y.-Q., Jin, J., and Gao, G. (2020). PlantRegMap: Charting functional regulatory maps in plants. *Nucleic Acids Res.* 48, D1104–D1113. doi: 10.1093/nar/gkz1020

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2. 0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Varshney, R. K., Graner, A., and Sorrells, M. E. (2005a). Genic microsatellite markers in plants: Features and applications. *Trends Biotechnol.* 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005

Varshney, R. K., Graner, A., and Sorrells, M. E. (2005b). Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10, 621–630. doi: 10.1016/j.tplants.2005.10.004

Wang, C., Xu, N., and Cui, S. (2021). Comparative transcriptome analysis of roots, stems, and leaves of *Pueraria lobata* (Willd.) Ohwi: Identification of genes involved in isoflavonoid biosynthesis. *PeerJ* 9:e10885. doi: 10.7717/peerj.10885

Yamazaki, M., Rai, A., Yoshimoto, N., and Saito, K. (2018). Perspective: Functional genomics towards new biotechnology in medicinal plants. *Plant Biotechnol. Rep.* 12, 69–75. doi: 10.1007/s11816-018-0476-9

Zhang, M., Yu, Z., Zeng, D., Si, C., Zhao, C., Wang, H., et al. (2021). Transcriptome and metabolome reveal salt-stress responses of leaf tissues from *Dendrobium officinale*. *Biomolecules* 11, 1–16. doi: 10.3390/biom11050736

Zhao, L., Gao, L., Wang, H., Chen, X., Wang, Y., Yang, H., et al. (2013). The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. *Funct. Integr. Genomics* 13, 75–98. doi: 10.1007/s10142-012-0301-4

Zhou, P., Pu, T., Gui, C., Zhang, X., and Gong, L. (2020). Transcriptome analysis reveals biosynthesis of important bioactive constituents and mechanism of stem formation of *Dendrobium huoshanense*. *Sci. Rep.* 10:2857. doi: 10.1038/s41598-020-59737-2

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fpls.2022.995734

OPEN ACCESS

EDITED BY
Katharina Nargar,
Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia

REVIEWED BY
Liangsheng Zhang,
Zhejiang University, China
Peng Zhao,
Northwest University, China

*CORRESPONDENCE
Zhong-Jian Liu
zjliu@fafu.edu.cn
Siren Lan
lkzx@fafu.edu.cn

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

RECEIVED 16 July 2022
ACCEPTED 07 November 2022
PUBLISHED 24 November 2022

CITATION
Wang Q-Q, Li Y-Y, Chen J, Zhu M-J, Liu X, Zhou Z, Zhang D, Liu Z-J and Lan S (2022) Genome-wide identification of YABBY genes in three *Cymbidium* species and expression patterns in *C. ensifolium* (Orchidaceae). *Front. Plant Sci.* 13:995734. doi: 10.3389/fpls.2022.995734

Genome-wide identification of YABBY genes in three *Cymbidium* species and expression patterns in *C. ensifolium* (Orchidaceae)

Qian-Qian Wang[1], Yuan-Yuan Li[1], Jiating Chen[1], Meng-Jia Zhu[2], Xuedie Liu[1], Zhuang Zhou[1,3], Diyang Zhang[1], Zhong-Jian Liu[1,2,3]* and Siren Lan[1,2]*

[1]Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at College of Landscape Architecture, Fujian Agriculture and Forestry University, Fuzhou, China, [2]College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, China, [3]Zhejiang Institute of Subtropical Crops, Zhejiang Academy of Agricultural Sciences, Wenzhou, China

Members of the YABBY gene family play significant roles in lamina development in cotyledons, floral organs, and other lateral organs. The Orchidaceae family is one of the largest angiosperm groups. Some YABBYs have been reported in Orchidaceae. However, the function of YABBY genes in *Cymbidium* is currently unknown. In this study, 24 YABBY genes were identified in *Cymbidium ensifolium*, *C. goeringii*, and *C. sinense*. We analyzed the conserved domains and motifs, the phylogenetic relationships, chromosome distribution, collinear correlation, and *cis*-elements of these three species. We also analyzed expression patterns of *C. ensifolium* and *C. goeringii*. Phylogenetic relationships analysis indicated that 24 YABBY genes were clustered in four groups, INO, CRC/DL, YAB2, and YAB3/FIL. For most YABBY genes, the zinc finger domain was located near the N-terminus and the helix-loop-helix domain (YABBY domain) near the C-terminus. Chromosomal location analysis results suggested that only *C. goeringii* YABBY has tandem repeat genes. Almost all the YABBY genes displayed corresponding one-to-one relationships in the syntenic relationships analysis. *Cis*-elements analysis indicated that most elements were clustered in light-responsive elements, followed by MeJA-responsive elements. Expression patterns showed that YAB2 genes have high expression in floral organs. RT-qPCR analysis showed high expression of *CeYAB3* in lip, petal, and in the gynostemium. *CeCRC* and *CeYAB2.2* were highly expressed in gynostemium. These findings provide valuable information of YABBY genes in *Cymbidium* species and the function in Orchidaceae.

KEYWORDS

YABBY genes, Orchidaceae, *Cymbidium*, expression pattern, genome-wide

Introduction

The seed plant-specific YABBY gene family, belonging to the zinc-finger superfamily, plays significant roles in lamina development in cotyledons, floral organs, and outer ovule integuments (Finet et al., 2016). YABBY genes encode transcription factors which contain two domains: a zinc finger domain located near the N-terminus and a helix-loop-helix domain (YABBY domain) located near the C-terminus (Bowman and Smyth, 1999). Six genes have been identified in *Arabidopsis thaliana*, and were clustered into five subfamilies: FIL/YAB3, CRC, INO, YAB2, and YAB5 (Siegfried et al., 1999). FIL, YAB2, YAB3, and YAB5 are expressed in leaf and floral organs and have been termed 'vegetative YABBYs'. CRC and INO are essential in developing carpels and ovules, respectively, and have been termed 'reproductive YABBYs' (Bowman and Smyth, 1999; Siegfried et al., 1999; Villanueva et al., 1999; Bartholmes et al., 2012; Soundararajan et al., 2019).

According to previous studies from expression characterization in *Arabidopsis* YABBY genes, FIL, YAB2 and YAB3 play essential roles in lateral organ development (Siegfried et al., 1999; Rudall and Bateman, 2002; Lora et al., 2011). CRC is restricted to carpels and nectaries in angiosperms (Siegfried et al., 1999). INO functions in the development of the outer integument of the ovule to the seed coat in *Arabidopsis*, and INO expresses in eudicots, eumagnoliids, and some basal angiosperms (Bowman, 2000; Yamada et al., 2003; McAbee et al., 2005; Lora et al., 2011; Yamada et al., 2011).

The genome-wide YABBY gene family has been identified in *Averrhoa carambola* (star fruit), *Cucumis sativus* (cucumber), *Lycopersicon esculentum* (tomato), *Oryza sativa* (rice), *Triticum aestivum* (wheat) and *Vitis vinifera* (grape) (Toriba et al., 2007; Han et al., 2015; Zhang et al., 2019; Hao et al., 2022; Li et al., 2022; Yin et al., 2022). In monocot plants, YABBY genes show functional divergence and are crucial for vegetative and reproductive development. For example, the YAB3 clade genes *ZYB9* and *ZYB14* play essential roles in flower development and regulate lateral outgrowth (Juarez et al., 2004). *OsDL*, a member of the CRC subfamily in *O. sativa*, is necessary for the development of the leaf midrib and the flower carpel specification (Nagasawa et al., 2003; Yamaguchi et al., 2004; Ohmori et al., 2008; Zhang et al., 2020). *OsYAB1*, belonging to the YAB2 clade, is mainly expressed in the primordia of the carpel and stamen (Jang et al., 2004). The *OsYAB3* gene may be necessary for the development of lateral organs and the growth and differentiation of leaf cells (Jang et al., 2004).

With an estimated > 28000 species, the Orchidaceae family is one of the largest angiosperm groups (Christenhusz and Byng, 2016). There are five subfamilies of Orchidaceae: Apostasioideae, Cypripedioideae, Vanilloideae, Orchidoideae, and Epidendroideae (Chase et al., 2003). The Orchidaceae show considerable diversity in epiphytic and terrestrial life forms and show unique flower morphologies and reproductive biology (Hsiao et al., 2011). Orchidaceae flowers show a variety of reliable floral morphological synapomorphies, such as a gynostemium (a fused structure of the pistils and stamens), a highly evolved petal termed labellum, and flowers with pollinia (Chase et al., 2003; Tsai et al., 2004). In the Orchidaceae family, genome-wide identification and expression patterns of YABBY genes were analyzed in *Apostasia shenzhenica* (Apostasioideae), *Dendrobium catenatum* (Epidendroideae), *Gastrodia elata* (Epidendroideae), and *Phalaenopsis equestris* (Epidendroideae) (Chen et al., 2020). However, studies of YABBY genes in the orchid tribe Cymbideae are still limited. *Cymbidium* is one of the most significant orchid genera for ornamental value because of its beautiful flowers (Ramya et al., 2019). Given the considerable role of YABBY genes in both vegetative and reproductive development, the identification of *Cymbidium ensifolium*, *C. goeringii*, and *C. sinense* will be employed, and the expression patterns of *C. ensifolium* will be analyzed in this study. This study provides new insights into the roles of YABBY genes and their contribution to the development of flower morphologies in *Cymbidium* subfamily of Orchidaceae.

Methods

Identification of YABBY genes from three *Cymbidium* species

The YABBY domain (PF04690) from PFAM was used as a query to search the protein database (El-Gebali et al., 2019). The genomes from *Cymbidium ensifolium*, *C. goeringii*, and *C. sinense* can be downloaded from their whole-genome sequencing data (Sun et al., 2021; Yang et al., 2021; Ai et al., 2021). HMM analysis (built in Tbtools) was used at an e value of 10^{-5} (Chen et al., 2018). BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was also used to search the protein database using *A. thaliana*'s YABBY sequences, which can be downloaded in the TAIR database (https://www.arabidopsis.org). Then, the CDD website (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) was used to confirm the retrieved putative sequences. The aliphatic index (AI), grand average of hydrophobicity (GRAVY), instability index (II), and isoelectric points (pI) of the YABBY proteins were predicted using the ExPASy website (https://www.expasy.org/; Artimo et al., 2012). AtSubP (http://bioinfo3.noble.org/AtSubP/) was used to predict the subcellular localization of YABBY genes (Kaundal et al., 2010). The secondary structure was predicted using the SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) program (Rozewicki et al., 2019).

Phylogenetic relationship analysis of YABBY genes

The TAIR database (https://www.arabidopsis.org/) was used to download the protein sequences of *Arabidopsis thaliana*. The sequences of *Oryza sativa*, *Phalaenopsis equestris*, *V. vinifera*, and *Zea mays* were downloaded from the NCBI website (https://www.ncbi.nlm.nih.gov/genbank/). The protein sequences of YABBY genes from *C. ensifolium*, *C. goeringii*, and *C. sinense* can be downloaded from their whole-genome sequencing data (Ai et al., 2021; Sun et al., 2021; Yang et al., 2021). Multiple alignments were carried out using the program MAFFT (Rozewicki et al., 2019). Maximum likelihood (ML) tree inference was carried out using RAxML (RAxML-HPC2 on XSEDE; Miller et al., 2011), and was under a GTRGAMMA substitution model with 1,000 bootstraps. The EVOLVIEW website (https://evolgenius.info/) was used for layouting the phylogenetic tree (He et al., 2016).

Motifs of YABBY proteins and sequence alignment in three *Cymbidium* species

Conserved domains of YABBY genes were analyzed using the CDD website (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi), and motifs were analyzed using the default parameters of the MEME website (http://meme-suite.org/) (Artimo et al., 2012). Fifteen motifs were identified in this study. To investigate the YABBY domains and C2C2 zinc-finger domain, the WEBLOGO tool (built in Tbtools) was employed. Multiple sequence alignments were carried out using MAFFT (Rozewicki et al., 2019).

Chromosome distribution and collinear correlation in three *Cymbidium* species

To analyze the chromosomal location of YABBY genes in three *Cymbidium* species, the Tbtools software was used to create gene distribution maps by uploading the YABBY sequence (Chen et al., 2018). To analyze syntenic relationships, one step MCScanx (built in Tbtools) was used to analyze YABBY genes of *C. ensifolium*, *C. goeringii*, and *C. sinense* (Chen et al., 2018).

Promoter element analysis of YABBY genes in *C. ensifolium*, *C. goeringii*, and *C. sinense*

The 2000 bp regions upstream of the YABBY genes in *C. ensifolium*, *C. goeringii*, and *C. sinense* were extracted by TBTOOLS (Chen et al., 2018). Then, the *cis*-acting elements were identified by the PlantCare website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; Zhang et al., 2018).

RNA extraction and RT–qPCR analysis

Flower organs (petal, lip, and gynostemium) and leaves of *C. ensifolium* were collected, frozen in liquid nitrogen, and stored at 80°C until use. Total RNA was extracted using the Biospin Plant Total RNA Extraction Kit (Bioer Technology, Hangzhou, China). TransScript® All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (TransGen Biotech, Beijing, China) was used to create first-strand DNA and remove genomic DNA. The reaction conditions were 30 s at 94 °C and 45 cycles of 5 s at 94°C and 30 s at 60°C. Primers for the RT–qPCR analysis were designed by Primer Premier 5 software. GAPDH (JL008987) was used for normalization. Three biological replicates were performed in this study, and the expression data were quantified *via* the 2-△△CT method (Livak and Schmittgen, 2001).

Results

YABBY gene identification and sequence analysis in three *Cymbidium* species

Seven YABBY genes were found in *C. ensifolium*, nine in *C. goeringii*, and eight in *C. sinense*. The deduced protein length of YABBY genes ranged from 63 to 243 amino acids. The theoretical isoelectric point (pI) ranged from 6.11 to 10.75, and instability index (II) ranged from 32.78 to 57.09. The deducted grand average of hydrophilic values (GRAVY) of YABBY genes ranged from -1.155 to -0.232, and we found all the YABBY proteins were hydrophilic. The molecular weight (Mw) ranged from 7744.05 to 27185.43, and the aliphatic index (AI) ranged from 52.92 to 83.98 (Table 1). Subcellular localization results showed that all the YABBY genes were located in the nucleus, indicating that the nucleus may be where the YABBY genes function (Supplementary Table S1; Kaundal et al., 2010). The results of secondary structure prediction revealed that the average of α-helices, extended strands, β-turns, and random coils comprised 27.61, 14.13, 5.65, and 52.6% of the structure, respectively (Supplementary Table S2; Geourjon and Deléage, 1995).

Phylogenetic relationship analysis of YABBY genes

To analyze the evolution patterns of YABBY genes in *Cymbidium* species, a phylogenetic tree was created by using

TABLE 1 A list of YABBY genes in three *Cymbidium* species.

Gene ID[1]	Name	AA[2](aa)	pI[3]	Mw[4](kDa)	AI[5]	II[6]	GRAVY[7]	Clade[8]	Localizetion[9]
JL015423	*CeCRC*	194	9.38	21643.69	59.38	45.46	-0.565	CRC	Nucleus
JL011339	*CeYAB2.1*	181	8.5	19855.55	83.98	43.57	-0.299	YAB2	Nucleus
JL000262	*CeYAB2.2*	181	7.74	19949.4	73.81	51.69	-0.401	YAB2	Nucleus
JL008521	*CeYAB3.1*	221	6.79	24350.89	79.5	57.09	-0.235	YAB2	Nucleus
JL005041	*CeYAB3.2*	221	7.7	24671.23	75.48	44.98	-0.329	YAB3	Nucleus
JL005324	*CeYAB2.3*	185	8.45	20710.15	66.43	47.37	-0.612	YAB2	Nucleus
JL012731	*CeINO*	157	9.32	17921.56	67.77	41.43	-0.543	INO	Nucleus
GL09549	*CgCRC.1*	188	9.11	21386.4	59.89	32.78	-0.625	CRC	Nucleus
GL08212	*CgCRC.2*	193	9.38	21643.69	59.38	45.46	-0.565	CRC	Nucleus
GL09374	*CgYAB3*	220	7.7	24698.26	75.48	44.98	-0.342	YAB3	Nucleus
GL12804	*CgYAB2.1*	242	9.73	27003.06	78.23	36.7	-0.408	YAB2	Nucleus
GL19435	*CgYAB2.2*	184	8.55	20658.15	64.32	46.16	-0.621	YAB2	Nucleus
GL30075	*CgYAB2.3*	78	9.8	8941.14	53.29	36.67	-0.995	YAB2	Nucleus
GL30077	*CgYAB2.4*	143	8.84	16732.82	52.92	47.9	-0.873	YAB2	Nucleus
GL30076	*CgYAB2.5*	63	10.75	7744.05	58.62	47.3	-1.155	YAB2	Nucleus
GL10103	*CgYAB2.6*	70	8.69	8099.94	64.79	33.43	-0.793	YAB2	Nucleus
Mol018025	*CsCRC.1*	243	9.1	27185.43	76.71	49.97	-0.27	CRC	Nucleus
Mol010228	*CsCRC.2*	194	9.38	21643.69	59.38	45.46	-0.565	CRC	Nucleus
Mol006632	*CsYAB2.1*	181	8.19	19887.51	81.82	41.79	-0.346	YAB2	Nucleus
Mol000581	*CsYAB2.2*	181	8.58	19968.49	73.81	52.75	-0.406	YAB2	Nucleus
Mol007225	*CsYAB3.1*	220	7.15	24195.67	79.86	54.95	-0.232	YAB2	Nucleus
Mol011195	*CsYAB2.3*	185	8.45	20710.15	66.43	47.37	-0.612	YAB2	Nucleus
Mol003404	*CsYAB3.2*	161	7.96	18198.82	73.42	43.89	-0.569	YAB3	Nucleus
Mol004846	*CsINO*	184	6.11	20622.28	69.46	41.53	-0.561	INO	Nucleus

[1]Gene ID is annotated in the genome; [2]AA, amino acid; [3]pI, theoretical isoelectric point; [4]Mw, molecular weight; [5]AI, aliphatic index; [6]II, instability index; [7]GRAVY, the grand average of hydrophobicity; [8]Clade is dependent on phylogenetic analysis, [9]Localization, predicted by AtSubP (Kaundal et al., 2010). Raw data are listed in Supplementary Tables S1 and S2.

the ML (maximum likelihood) method. Protein sequences from *C. ensifolium*, *C. goeringii*, *C. sinense*, *A. thaliana O. sativa*, *P. equestris*, *V. vinifera*, and *Z. mays* were used. The IDs of these species are listed in Supplementary Table S3. The results indicated that all *Cymbidium* species except *C. goeringii* have one member in the INO cluster. The number of YAB2 genes ranged from 3-6 (*C. ensifolium*: 3; *C. goeringii*: 6; *C. sinense*: 3). *C. goeringii* and *C. sinense* have two genes in the CRC subfamily, but *C. ensifolium* has only one. With the exception of *C. goeringii*, all *Cymbidium* species have two YAB3 genes (Figure 1).

Domains and motifs of YABBY genes

To analyze the conserved domains of YABBY genes, the sequence logo of YABBY domains and c2c2 zinc-finger domains from three *Cymbidium* species and *A. thaliana* was generated. The multiple sequence alignment was also generated. The results showed that *Cymbidum* species and *A. thaliana* have highly conserved c2c2 zinc-finger domains and YABBY domains. However, the YABBY domain is more conserved than the c2c2 domain in *Cymbidium* species (Figure 2). Additionally, the motifs, domains, and phylogenetic tree of three *Cymbidium* species were analyzed (Figure 2). Fifteen motifs were analyzed by MEME software (Supplementary Table S4; Bailey et al., 2009). The results indicated that all the *Cymbidium* species have YABBY domains, and most YABBY genes of *Cymbidium* have motif 2 and motif 4. The findings also revealed that the conserved motifs of YABBY genes in the same clusters are similar.

Chromosome distribution and collinear correlation analysis

To analyze the chromosome distribution of YABBY genes in three *Cymbidium* species, we create gene distribution maps. The results suggest that YABBY genes were distributed in seven chromosomes in *C. ensifolium*, *C. goeringii*, and *C. sinense* (Figure 3). In addition, YABBY genes were located in different chromosomes in *C. ensifolium* and *C. sinense*. Nevertheless, in *C. goeringii*, *CgYAB2.3*, *CgYAB2.4*, and *CgYAB2.5* were located on same chromosome (chr17). We also analyzed the syntenic relationships of YABBY genes in three *Cymbidium* species. There are seven, nine, and eight YABBY genes in *C. ensifolium*,

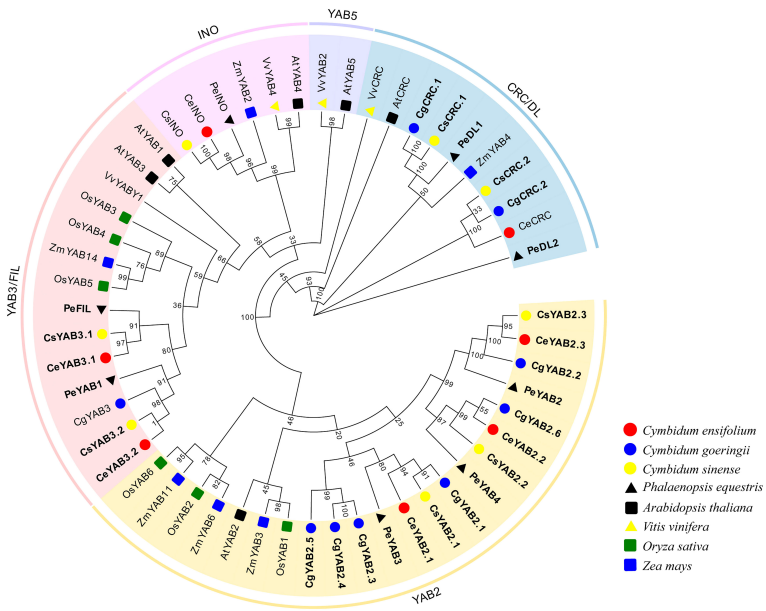


FIGURE 1
Phylogenetic tree of YABBY genes in eight plant species. The phylogenetic tree was created with the maximum-likelihood (ML) method using RAxML on the CIPRES Science Gateway web server (RAxML-HPC2 on XSEDE; Miller et al., 2011). Bootstrap values based on 1000 replicates are shown along the branches. Ce, *C. ensifolium*; Cg, *C. goeringii*; Cs, *C. sinensis*; At, *A. thaliana*; Os, *O. sativa*; Pe, *P. equestris*; Vv, *V. vinifera*; Zm, *Z. mays*; The duplicated genes are shown in bold.

C. goeringii, and *C. sinense* (Figure 4). The results indicated that almost all the YABBY genes displayed corresponding one-to-one relationships in these three *Cymbidium* species.

Cis-element analysis of *C. ensifolium*, *C. goeringii*, and *C. sinense*

To predict the regulatory function of YABBY genes, we retrieved a 2000-bp region upstream of 24 YABBY genes and analyze them in *C. ensifolium*, *C. goeringii*, and *C. sinense*. We identified 12 types of *cis*-elements: abscisic acid responsiveness element, anaerobic induction element, auxin responsiveness element, circadian control element, defense and stress responsiveness element, endosperm expression element, light responsive element, low-temperature responsiveness element, MeJA-responsiveness element, meristem expression element, salicylic acid responsiveness element, and zein metabolism regulation element. In total, we found 412 *cis*-elements in three *Cymbidium* species, and *C. sinense* has most of the cis-elements (192/412), followed by *C. goeringii* (120/412), and *C. ensifolium* (100/412). The results also indicated that most of the elements were clustered in light-responsive elements (199/412), followed by MeJA-responsive elements (64/412), anaerobic induction element (27/412), and abscisic acid responsiveness element (24/412). All YABBY genes have light-responsive elements, and *CsYAB3.1* contains the most (35/199). In addition, only *CeYAB2.1*, *CeYAB3.2*, and *CgYAB2.1* have circadian control elements (Figure S1).

Expression analysis of *C. ensifolium* and *C. goeringii*

To analyze the expression patterns of YABBY genes, we sampled vegetative and floral organs from *C. ensifolium* and *C. goeringii*. The results suggested that in *C. ensifolium*, *CeCRC* showed high expression in pseudobulbs and pedicel, *CeYAB2.1* and *CeYAB 2.2* showed high expression in leaf and gynostemium, and *CeYAB3.2* showed high expression in bud. *CeYAB2.1*, *CeYAB2.2*, *CeYAB3.1*, and *CeYAB3.2* had expression in both vegetative and floral organs (Figure 5). In *C. goeringii*, *CgCRC.1* showed high expression in gynostemium, and *CgCRC.2* showed high expression in pseudobulbs and gynostemium. *CgYAB3* showed high expression in pseudobulbs, leaves, and petals. *CgCRC.2*, *CgYAB2.1*, *CgYAB2.2*, *CgYAB2.6*, and *CgYAB3* had expression in both vegetative and floral organs.

Expression patterns in leaves and three floral organs in *C. ensifolium*

To analyze the expression patterns of YABBY genes, we collected three floral organs (petal, lip, and gynostemium) and

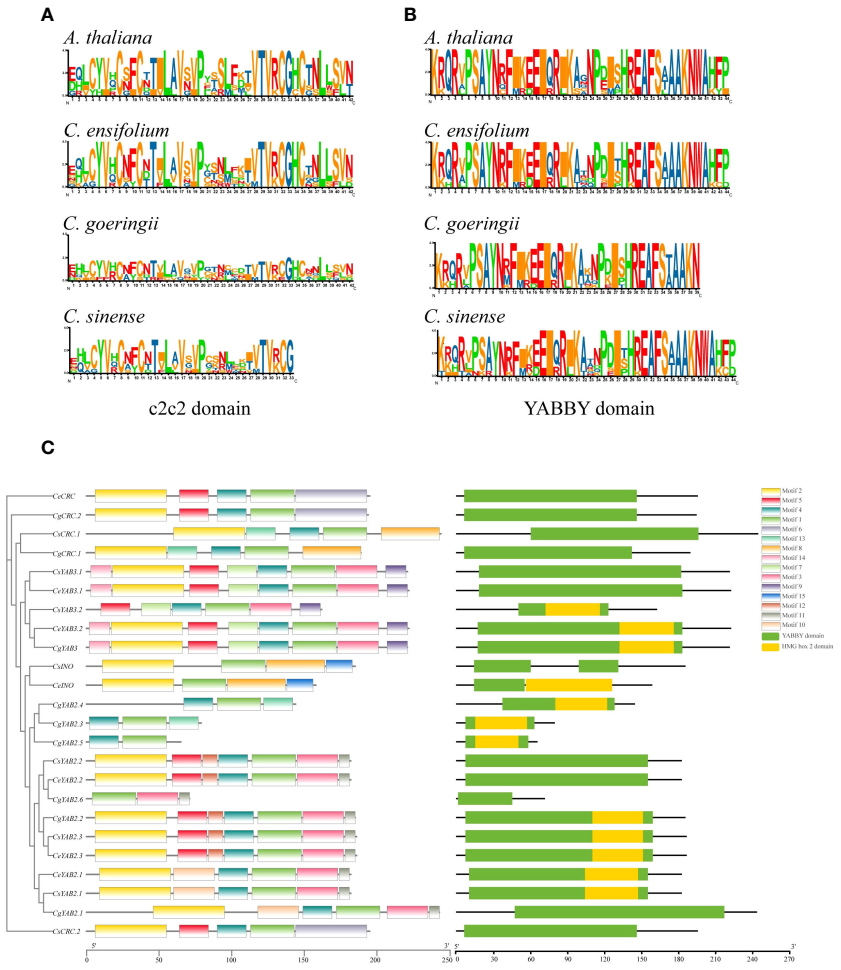


FIGURE 2
Conserved domains from three *Cymbidium* species and *A. thaliana*. **(A)** Sequence logo of the zinc-finger domain in the N-terminus. **(B)** Sequence logo of the YABBY domain in the C-terminus. **(C)** Motifs and conserved domains in the YABBY protein amino acid sequences in *Cymbidium* species.

leaves from *C. ensifolium*. Four YABBY genes, *CeCRC*, *CeINO*, *CeYAB2.2*, and *CeYAB3.1* were chosen for RT–qPCR analysis. The results showed that *CeYAB3* showed high expression in the lip, petal, and gynostemium. *CeCRC* and *CeYAB2.2* showed high expression in gynostemium. *CeCRC*, *CeYAB2.2*, and *CeYAB3.1* had higher expression levels in floral organs than in leaves. However, the expression levels in leaves were higher than those in floral organs from *CeINO* (Figure 6).

Discussion

YABBY genes, which include a zinc finger domain near the N-terminus and a helix-loop-helix domain (YABBY domain) near the C-terminus, play important roles in lamina development in cotyledons, floral organs, and outer ovule integuments (Finet et al., 2016). In monocots, eight genes have been identified in *O. sativa*; in core eudicots, six YABBY genes have been found in *A. thaliana* (Bowman and Smyth, 1999; Sawa et al., 1999; Villanueva et al., 1999). Orchidaceae, belonging to monocots, is one of the largest angiosperm families and show unique flower morphologies and reproductive biology (Hsiao et al., 2011; Christenhusz and Byng, 2016). Recent studies have indicated that six YABBY genes were identified in *A. shenzhenica*, eight in *D. catenatum*, five in *G. elata*, and eight in *P. equestris* (Chen et al., 2020). However, studies of YABBY genes in *Cymbidium* are still limited. In this study, YABBY genes were identified in three *Cymbidium* species and the number of YABBY genes ranged from 7-9 (*C. ensifolium*: 7; *C. goeringii*: 9; *C. sinense*: 8). These results indicated that the number of YABBY genes in *Cymbidum* orchids were comparable to those in monocot and dicot species. However, the absence of YABBY

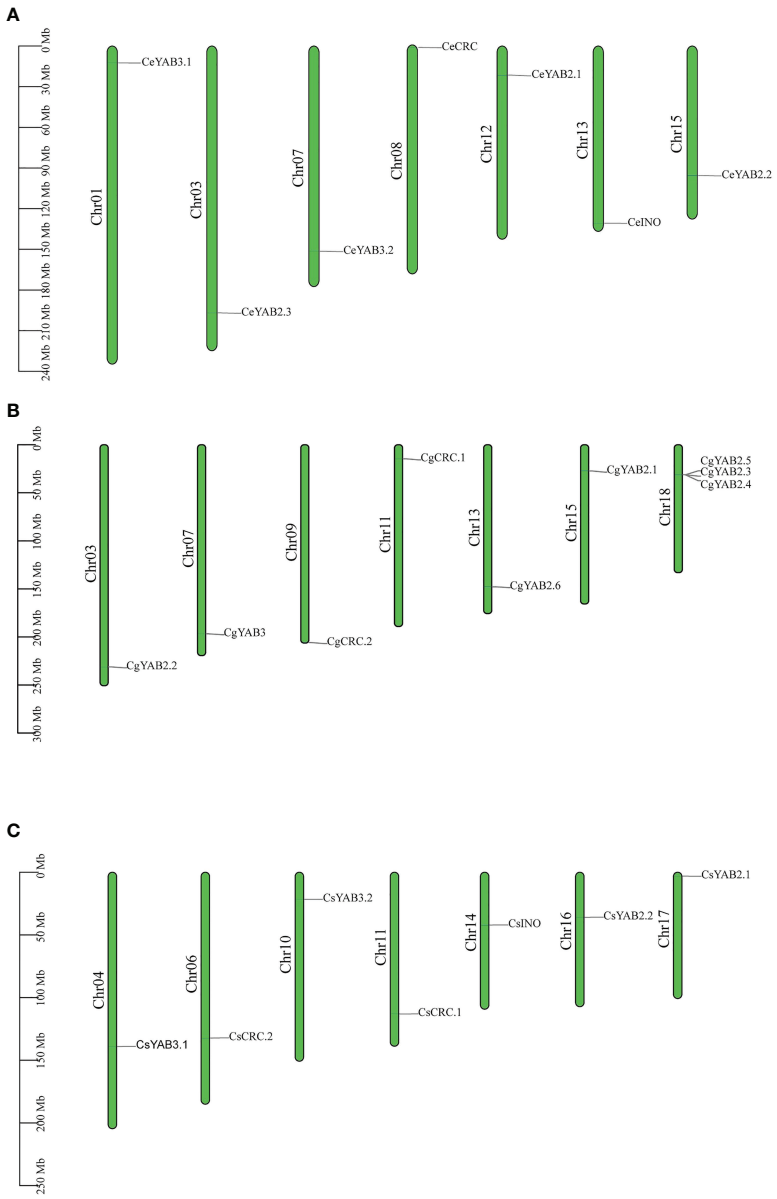


FIGURE 3
Chromosome distribution in three *Cymbidium* species. **(A)** Chromosome distribution in *C. ensifolium*. **(B)** Chromosome distribution in *C. goeringii*. **(C)** Chromosome distribution in *C. sinense*.

genes in YAB 5 subfamily in orchids and other monocots is an exception.

The phylogenetic analysis indicated that YABBY genes in *Cymbidium* species are clustered into four subfamilies: YAB2, CRC, YAB3, and INO. There were no YABBY genes that clustered in the YAB5 subfamily. The results were consistent with some monocot species, such as *A. shenzhenica*, *D. catenatum*, *G. elata*, pineapple, and rice (Toriba et al., 2007; Li et al., 2019; Chen et al., 2020). However, seven species of magnoliids and *A. thaliana* have YABBY genes clustered in the YAB 5 clade (Siegfried et al., 1999; Liu et al., 2021). Early in the evolution of angiosperms, the lineages of basal flowering plants diverged, and then the magnoliids, eudicots, and monocots underwent rapid diversification (Tang et al., 2014; Chen et al., 2019). Magnoliids have two cotyledons and pollen with a single pore, and they are not monocots or eudicots (Tang et al., 2014). Recent reports also studied the comparative development of the androecial form in the Zingiberales and

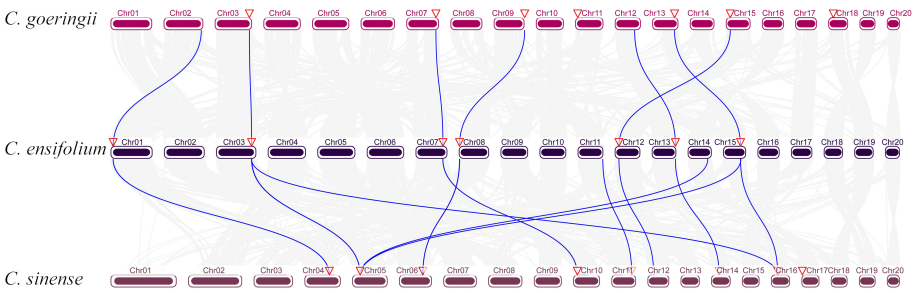


FIGURE 4
Collinear correlation analysis in three *Cymbidium* species.

found one YAB2 gene, which was less homologous to YAB5 (De Almeida et al., 2014). Based on this, they suggested that after the divergence of monocots and eudicots, duplication led to separate YAB2 and YAB5 gene lineages (De Almeida et al., 2014). The YAB5 clade was exclusively composed of basal angiosperms and eudicot in recent studies (Chen et al., 2017; Liu et al., 2021). These results suggested that YAB5 gene clade might have been lost in monocot plants.

INO are restricted to the development of the outer ovule integument (Villanueva et al., 1999). Interestingly, we found *C. ensifolium* and *C. sinense* only has one number in the INO clade. These results were consistent with *A. shenzhenica*, *A. thaliana*, *D. catenatum*, *G. elata*, *P. equestris*, and *V. vinifera*, and indicated INO clade genes might be conserved in angiosperm plants and play essential roles in the outer integument (Siegfried et al., 1999; Zhang et al., 2019; Chen et al., 2020).

YABBY genes include a zinc finger domain near the N-terminus and a helix-loop-helix domain (YABBY domain) near the C-terminus. The results showed that the YABBY domain is more conserved than the c2c2 domain in three *Cymbidium* species. Fifteen motifs were analyzed in three *Cymbidum* species, and most YABBY genes of *Cymbidium* have motif 4 and motif 2. These findings revealed that the gene structure of YABBY genes are conserved during evolution. In the evolution of gene families, two main methods are tandem duplication and fragment duplication (Cannon et al., 2004). Chromosomal location analysis results suggested that YABBY genes were located in different chromosomes in *C. ensifolium* and *C. sinense*. But in *C. goeringii*, *CgYAB2.3*, *CgYAB2.4*, and *CgYAB2.5* were located on same chromosome (chr17). The results indicated those genes might be tandem repeat genes. The syntenic relationships analysis indicated that almost every YABBY gene displayed

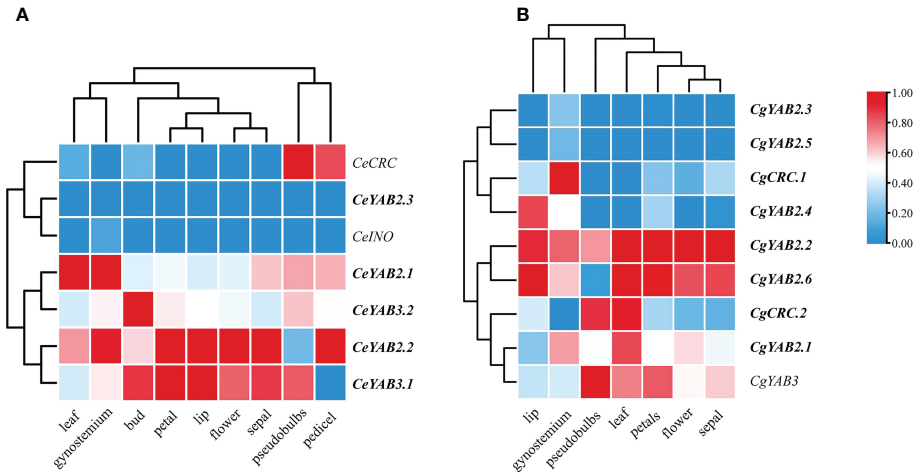


FIGURE 5
Expression patterns of YABBY genes in different organs from three *Cymbidium* species. **(A, B)** show the expression patterns of YABBY genes in different organs in *C. ensifolium* and *C. goeringii*. The heatmap was produced in Tbtools (Chen et al., 2018). The fragments per kilobase of transcript per million fragments (FPKM) values can be found in Supplementary Table S5. The duplicated genes are shown in bold.

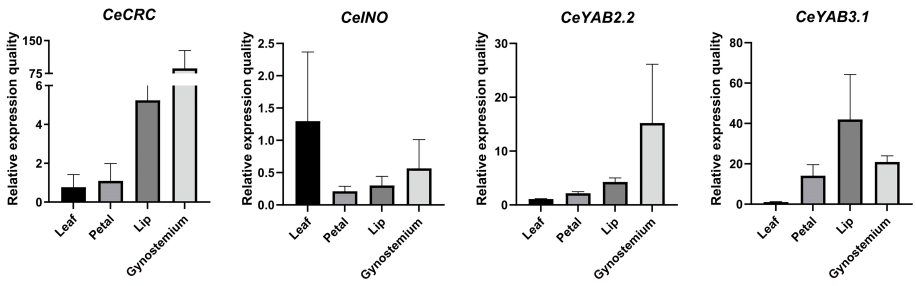


FIGURE 6
Relative expression patterns of YABBY genes in *C. ensifolium*. Raw data are listed in Supplementary Tables S6 and S7.

corresponding one-to-one relationships in these three *Cymbidium* species.

Cis-elements were found in promoter areas in YABBY genes. The results indicated that most of the elements were clustered in light-responsive elements (199/412), followed by MeJA-responsive elements (64/412), anaerobic induction elements (27/412), and abscisic acid responsiveness element (24/412). The MeJA (methyl jasmonate) is a phytohormone involved in defense signaling of plants (Howe, 2004). The results indicated YABBY genes might play essential roles in plant growth and stress.

The growth of lateral organs in *A. thaliana* is thought to be redundantly controlled by the genes YAB2 and FIL, which are expressed in the leaves, cotyledons, and floral organs (Siegfried et al., 1999; Rudall and Bateman, 2002). FIL gene orthologues have similarly acted in flower development in *Oryza* (Tanaka et al., 2017). Our study indicated that three *Cymbidium* species contained one or two FIL genes and had high expression in the floral organs of *C. ensifolium* and *C. goeringii*. The results suggested that FIL may play important roles in the development of floral organ in *Cymbidium* species. CRC showed high expression in pseudobulbs in *C. ensifolium* and *C. goeringii*, and CRC showed high expression in pedicels in *C. ensifolium*. CRC also showed high expression in gynostemium in *C. goeringii*. The results suggested that CRC in different *Cymbidum* had different expression patterns. INO expressed in the gynostemium and pedicel in *C. ensifolium*. It may play important roles in the development of gynostemium and pedicel. YAB2 genes (*CeYAB2.1*, *CeYAB2.2*, *CgYAB2.1*, *CgYAB2.2*, and *CgYAB2.6*) showed high expression in all organs in *Cymbidium* species, indicating that the YAB2 clade may have functions in both reproductive and vegetative organs.

The results indicated that YABBY genes in *Cymbidium* species showed higher expression in reproductive tissues than in vegetative tissues. The results were consistent with the expression patterns reported in *A. shenzhenica*, *D. catanum*, and *P. equestris* (Chen et al., 2020). RT–qPCR analysis showed that *CeCRC*, *CeYAB2.2*, and *CeYAB3.1* have higher expression levels in floral organs than in leaves. However, the expression levels in leaves were slightly higher than those in floral organs in *CeINO*. These findings indicated that YABBY genes play important roles in floral organ development in orchids. Orchids display unique flower morphologies, and their flowers possess several reliable floral morphological synapomorphies, including a gynostemium (a fused structure of the pistils and stamens) (Chase et al., 2003; Tsai et al., 2004). The results of this study indicated that *CeCRC* might play essential roles in floral organs, especially in gynostemium.

Data availability statement

The data presented in the study are deposited in the National Centre for Biotechnology Information (NCBI) and National Genomics Data Center (NGDC). The raw data can be found under the following accession numbers: SAMN20059972 (NCBI), PRJNA749652 (NCBI) and PRJCA005355 (NGDC).

Author contributions

SL, Z-JL, and DZ contributed to conceptualization and validation. Q-QW, Y-YL, and ZZ prepared the original draft. Q-QW, JC, and M-JZ analyzed data, Q-QW and XL make the images. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by Forestry Peak Discipline Construction Project of Fujian Agriculture and Forestry University (72202200205), The National Natural Science Foundation of China (no. 31870199), The Key Laboratory of

National Forestry and Grassland Administration for Orchid Conservation and Utilization Construction Funds (Grant 115/118990050, 115/KJG18016A), Natural Science Foundation of Zhejiang Province (Grant nos. LY20C160005, LY19C150003), Key Research and Development Program of Zhejiang Province (Grant no. 2021C02043), and Wenzhou Agricultural New Variety Breeding Cooperative Group Project (Grant no. 2019ZX004-3).

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.2022.995734/full#supplementary-material

References

Ai, Y., Li, Z., Sun, W. H., Chen, J., Zhang, D., Ma, L., et al. (2021). The cymbidium genome reveals the evolution of unique morphological traits. *Horticult. Res.* 8, 255. doi: 10.1038/s41438-021-00683-z

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 40, 597–603. doi: 10.1093/nar/gks400

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME suite: Tools for motif discovery and searching. *Nucleic Acids Res.* 37, 1–7. doi: 10.1093/nar/gkp335

Bartholmes, C., Hidalgo, O., and Gleissberg, S. (2012). Evolution of the YABBY gene family with emphasis on the basal eudicot eschscholzia californica (Papaveraceae). *Plant Biol.* 14, 11–23. doi: 10.1111/j.1438-8677.2011.00486.x

Bowman, J. L. (2000). The YABBY gene family and abaxial cell fate. *Curr. Opin. Plant Biol.* 3, 17–22. doi: 10.1016/S1369-5266(99)00035-7

Bowman, J. L., and Smyth, D. R. (1999). CRABS CLAW, a gene that regulates carpel and nectary development in arabidopsis, encodes a novel protein with zinc finger and helix-loop-helix domains. *Development* 126, 2387–2396. doi: 10.1242/dev.126.11.2387

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

Chase, M., Cameron, K., Barrett, R., and Freudenstein, J. V. (2003). DNA Data and orchidaceae systematics: A new phylogenetic classification. *Orchid Conserv.* 69 (89), 32. Available at: https://www.researchgate.net/publication/234814296.

Chen, C., Chen, H., He, Y., and Xia, R. (2018). TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. *bioRxiv*. 289660 (10.1101), 289660. doi: 10.1101/289660

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

Chen, Y. Y., Hsiao, Y. Y., Chang, S. B., Zhang, D., Lan, S. R., Liu, Z. J., et al. (2020). Genome-wide identification of yabby genes in orchidaceae and their expression patterns in phalaenopsis orchid. *Genes* 11, 1–17. doi: 10.3390/genes11090955

Chen, F., Liu, X., Yu, C., Chen, Y., Tang, H., and Zhang, L. (2017). Water lilies as emerging models for darwin's abominable mystery. *Horticult. Res.* 4, 17051. doi: 10.1038/hortres.2017.51

Christenhusz, M. J. M., and Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa* 261, 201–217. doi: 10.11646/phytotaxa.261.3.1

De Almeida, A. M. R., Yockteng, R., Schnable, J., Alvarez-Buylla, E. R., Freeling, M., and Specht, C. D. (2014). Co-Option of the polarity gene network shapes filament morphology in angiosperms. *Sci. Rep.* 4, 1–9. doi: 10.1038/srep06194

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432. doi: 10.1093/nar/gky995

Finet, C., Floyd, S. K., Conway, S. J., Zhong, B., Scutt, C. P., and Bowman, J. L. (2016). Evolution of the YABBY gene family in seed plants. *Evol. Dev.* 18, 116–126. doi: 10.1111/ede.12173

Geourjon, C., and Deléage, G. (1995). Sopma: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11, 681–684. doi: 10.1093/bioinformatics/11.6.681

Han, H. Q., Liu, Y., Jiang, M. M., Ge, H. Y., and Chen, H. Y. (2015). Identification and expression analysis of YABBY family genes associated with fruit shape in tomato (Solanum lycopersicum l.). *Genet. Mol. Res.* 14, 7079–7091. doi: 10.4238/2015.June.29.1

Hao, L., Zhang, J., Shi, S., Li, P., Li, D., Zhang, T., et al. (2022). Identification and expression profiles of the YABBY transcription factors in wheat. *PeerJ* 10, 1–15. doi: 10.7717/peerj.12855

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W. H., and Hu, S. (2016). Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* 44, W236–W241. doi: 10.1093/nar/gkw370

Howe, G. A. (2004). Jasmonates as signals in the wound response. *J. Plant Growth Regul.* 23 (3), 223–237. doi: 10.1007/s00344-004-0030-6

Hsiao, Y. Y., Pan, Z. J., Hsu, C. C., Yang, Y. P., Hsu, Y. C., Chuang, Y. C., et al. (2011). Research on orchid biology and biotechnology. *Plant Cell Physiol.* 52, 1467–1486. doi: 10.1093/pcp/pcr100

Jang, S., Hur, J., Kim, S. J., Han, M. J., Kim, S. R., and An, G. (2004). Ectopic expression of OsYAB1 causes extra stamens and carpels in rice. *Plant Mol. Biol.* 56, 133–143. doi: 10.1007/s11103-004-2648-y

Juarez, M. T., Twigg, R. W., and Timmermans, M. C. P. (2004). Specification of adaxial cell fate maize leaf development. *Development* 131, 4533–4544. doi: 10.1242/dev.01328

Kaundal, R., Saini, R., and Zhao, P. X. (2010). Combining machine learning and homology-based approaches to accurately predict subcellular localization in arabidopsis. *Plant Physiol.* 154, 36–54. doi: 10.1104/pp.110.156851

Li, C., Dong, N., Shen, L., Lu, M., Zhai, J., Zhao, Y., et al. (2022). Genome-wide identification and expression profile of YABBY genes in averrhoa carambola. *PeerJ* 9, 1–22. doi: 10.7717/peerj.12558

Li, Z., Li, G., Cai, M., Priyadarshani, S. V. G. N., Aslam, M., Zhou, Q., et al. (2019). Genome-wide analysis of the YABBY transcription factor family in pineapple and functional identification of AcYABBY4 involvement in salt stress. *Int. J. Mol. Sci.* 20, 1–17. doi: 10.3390/ijms20235863

Liu, X., Liao, X. Y., Zheng, Y., Zhu, M. J., Yu, X., Jiang, Y. T., et al. (2021). Genome-wide identification of the YABBY gene family in seven species of magnoliids and expression analysis in litsea. *Plants* 10, 1–18. doi: 10.3390/plants10010021

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lora, J., Hormaza, J. I., Herrero, M., and Gasser, C. S. (2011). Seedless fruits and the disruption of a conserved genetic pathway in angiosperm ovule development. *Proc. Natl. Acad. Sci. United States America* 108, 5461–5465. doi: 10.1073/pnas.1014514108

McAbee, J. M., Kuzoff, R. K., and Gasser, C. S. (2005). Mechanisms of derived unitegmy among impatiens species. *Plant Cell* 17, 1674–1684. doi: 10.1105/tpc.104.029207

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2011). The CIPRES science gateway: A community resource for phylogenetic analyses. Proceedings of the TeraGrid 2011 Conference: Extreme Digital Discovery, TG'11. doi: 10.1145/2016741.2016785

Nagasawa, N., Miyoshi, M., Sano, Y., Satoh, H., Hirano, H., Sakai, H., et al. (2003). SUPERWOMAN1 and DROOPING LEAF genes control floral organ identity in rice. *Development* 130, 705–718. doi: 10.1242/dev.00294

Ohmori, Y., Abiko, M., Horibata, A., and Hirano, H. Y. (2008). A transposon, ping, is integrated into intron 4 of the DROOPING LEAF gene of rice, weakly reducing its expression and causing a mild drooping leaf phenotype. *Plant Cell Physiol.* 49, 1176–1184. doi: 10.1093/pcp/pcn093

Ramya, M., Park, P. H., Chuang, Y. C., Kwon, O. K., An, H. R., Park, P. M., et al. (2019). RNA Sequencing analysis of cymbidium goeringii identifies floral scent biosynthesis related genes. *BMC Plant Biol.* 19, 1–14. doi: 10.1186/s12870-019-1940-6

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

Rudall, P., and Bateman, R. (2002). Roles of synorganisation, zygomorphy and heterotopy in floral evolution: The gynostemium and labellum of orchids and other lilioid monocots. *Biol. Rev.* 77 (3), 403–441. doi: 10.1017/S1464793102005936

Sawa, S., Ito, T., Shimura, Y., and Okada, K. (1999). FILAMENTOUS FLOWER controls the formation and development of arabidopsis inflorescences and floral meristems. *Plant Cell* 11, 69–86. doi: 10.1105/tpc.11.1.69

Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, D., Drews, G. N., and Bowman, J. L. (1999). Members of the YABBY gene family specify abaxial cell fate in arabidopsis. *Development* 126 (18), 4117–4128. doi: 10.1242/dev.126.18.4117

Soundararajan, P., Won, S. Y., Park, D. S., Lee, Y. H., and Sun Kim, J. (2019). Comparative analysis of the YABBY gene family of bienertia sinuspersici, a single-cell c4 plant. *Plants* 8. doi: 10.3390/plants8120536

Sun, Y., Chen, G., Huang, J., Liu, D., Xue, F., Chen, X., et al. (2021). The cymbidium goeringii genome provides insight into organ development and adaptive evolution in orchids. *Ornamental Plant Research* 1 (1), 1–13. doi: 10.48130/OPR-2021-0010

Tanaka, W., Toriba, T., and Hirano, H. Y. (2017). Three TOB1-related YABBY genes are required to maintain proper function of the spikelet and branch meristems in rice. *New Phytol.* 215, 825–839. doi: 10.1111/nph.14617

Tang, H., Lyons, E., and Schnable, J. C. (2014). "Earlay history of the angiosperms," in *Advances in botanical research*, (Academic Press) 69, 195–222. doi: 10.1016/B978-0-12-417163-3.00008-1

Toriba, T., Harada, K., Takamura, A., Nakamura, H., Ichikawa, H., Suzaki, T., et al. (2007). Molecular characterization the YABBY gene family in oryza sativa and expression analysis of OsYABBY1. *Mol. Genet. Genomics* 277, 457–468. doi: 10.1007/s00438-006-0202-0

Tsai, W. C., Kuoh, C. S., Chuang, M. H., Chen, W. H., and Chen, H. H. (2004). Four DEF-like MADS box genes displayed distinct floral morphogenetic roles in phalaenopsis orchid. *Plant Cell Physiol.* 45, 831–844. doi: 10.1093/pcp/pch095

Villanueva, J. M., Broadhvest, J., Hauser, B. A., Meister, R. J., Schneitz, K., and Gasser, C. S. (1999). INNER NO OUTER regulates abaxial-adaxial patterning in arabidopsis ovules. *Genes Dev.* 13, 3160–3169. doi: 10.1101/gad.13.23.3160

Yamada, T., Ito, M., and Kato, M. (2003). Expression pattern of INNER NO OUTER homologue in nymphaea (water lily family, nymphaeaceae). *Dev. Genes Evol.* 213, 510–513. doi: 10.1007/s00427-003-0350-8

Yamada, T., Yokota, S., Hirayama, Y., Imaichi, R., Kato, M., and Gasser, C. S. (2011). Ancestral expression patterns and evolutionary diversification of YABBY genes in angiosperms. *Plant J.* 67, 26–36. doi: 10.1111/j.1365-313X.2011.04570.x

Yamaguchi, T., Nagasawa, N., Kawasaki, S., Matsuoka, M., Nagato, Y., and Hirano, H. Y. (2004). The yabby gene drooping leaf regulates carpel specification and midrib development in oryza sativa. *Plant Cell* 16, 500–509. doi: 10.1105/tpc.018044

Yang, F., Gao, J., Wei, Y., Ren, R., Zhang, G., Lu, C., et al. (2021). The genome of cymbidium sinense revealed the evolution of orchid traits. *Plant Biotechnol. J.* 19 (12), 2501. doi: 10.1111/pbi.13676

Yin, S., Li, S., Gao, Y., Bartholomew, E. S., Wang, R., Yang, H., et al. (2022). Genome-wide identification of YABBY gene family in cucurbitaceae and expression analysis in cucumber (Cucumis sativus l.). *Genes* 13 (3), 467. doi: 10.3390/genes13030467

Zhang, T., Li, C., Li, D., Liu, Y., and Yang, X. (2020). Roles of YABBY transcription factors in the modulation of morphogenesis, development, and phytohormone and stress responses in plants. *J. Plant Res.* 133, 751–763. doi: 10.1007/s10265-020-01227-7

Zhang, J., Li, Y., Liu, B., Wang, L., Zhang, L., Hu, J., et al. (2018). Characterization of the populus rab family genes and the function of PtRabE1b in salt tolerance. *BMC Plant Biol.* 18, 1–15. doi: 10.1186/s12870-018-1342-1

Zhang, S., Wang, L., Sun, X., Li, Y., Yao, J., van Nocker, S., et al. (2019). Genome-wide analysis of the YABBY gene family in grapevine and functional characterization of VvYABBY4. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01207

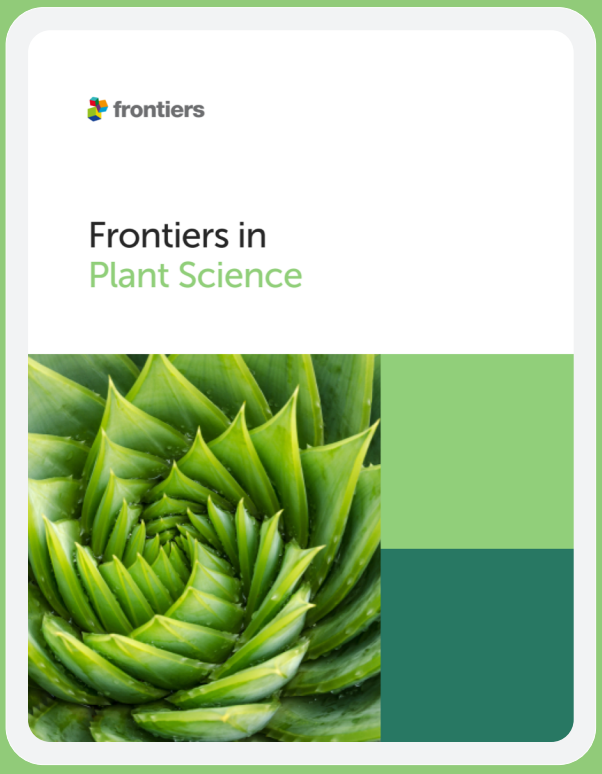
Frontiers in Plant Science

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