frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 March 2023
DOI 10.3389/fpls.2023.1123707

OPEN ACCESS

EDITED BY
Nikolai Borisjuk,
Huaiyin Normal University, China

REVIEWED BY
Mei-Yeh Jade Lu,
Academia Sinica, Taiwan
Tao Shi,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Yongpeng Ma
mayongpeng@mail.kib.ac.cn
Chengjun Zhang
zhangchengjun@mail.kib.ac.cn
Jihua Wang
wjh@yaas.org.cn

†These authors have contributed equally to this work

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

RECEIVED 14 December 2022
ACCEPTED 20 February 2023
PUBLISHED 21 March 2023

CITATION
Wu X, Zhang L, Wang X, Zhang R, Jin G, Hu Y, Yang H, Wu Z, Ma Y, Zhang C and Wang J (2023) Evolutionary history of two evergreen *Rhododendron* species as revealed by chromosome-level genome assembly.
Front. Plant Sci. 14:1123707.
doi: 10.3389/fpls.2023.1123707

Evolutionary history of two evergreen *Rhododendron* species as revealed by chromosome-level genome assembly

Xiaopei Wu[1,2†], Lu Zhang[3†], Xiuyun Wang[4], Rengang Zhang[5,6], Guihua Jin[1], Yanting Hu[1], Hong Yang[1,7], Zhenzhen Wu[1,2], Yongpeng Ma[5,6*], Chengjun Zhang[1,8*] and Jihua Wang[3*]

[1]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Flower Research Institute of Yunnan Academy of Agricultural Sciences, National Engineering Research Center for Ornamental Horticulture, Kunming, China, [4]Genomics and Genetic Engineering Laboratory of Ornamental Plants, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China, [5]Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [6]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming, China, [7]State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan, China, [8]Zhejiang Institute of Advanced Technology, Haiyan Engineering & Technology Center, Jiaxing, China

Background: The genus *Rhododendron* (Ericaceae), a species-rich and widely distributed genus of woody plants, is distinguished for the beautiful and diverse flowers. *Rhododendron delavayi* Franch. and *Rhododendron irroratum* Franch., are highly attractive species widely distributed in south-west China and abundant new varieties have been selected from their genetic resources.

Methods: We constructed chromosome-scale genome assemblies for *Rhododendron delavayi* and *Rhododendron irroratum*. Phylogenetic and whole-genome duplication analyses were performed to elucidate the evolutionary history of *Rhododendron*. Further, different types of gene duplications were identified and their contributions to gene family expansion were investigated. Finally, comprehensive characterization and evolutionary analysis of R2R3-MYB and NBS-encoding genes were conducted to explore their evolutionary patterns.

Results: The phylogenetic analysis classified *Rhododendron* species into two sister clades, 'rhododendrons' and 'azaleas'. Whole-genome duplication (WGD) analysis unveiled only one WGD event that occurred in *Rhododendron* after the ancestral γ triplication. Gene duplication and gene family expansion analyses suggested that the younger tandem and proximal duplications contributed greatly to the expansion of gene families involved in secondary metabolite biosynthesis and stress response. The candidate R2R3-MYB genes likely regulating anthocyanin biosynthesis and stress tolerance in *Rhododendron* will facilitate the breeding for ornamental use. NBS-encoding genes had undergone significant expansion and experienced species-specific gain and loss events in *Rhododendron* plants.

Conclusions: The reference genomes presented here will provide important genetic resources for molecular breeding and genetic improvement of plants in this economically important *Rhododendron* genus.

KEYWORDS

***Rhododendron delavayi*, *Rhododendron irroratum*, genome evolution, gene duplication, R2R3-MYB transcription factors, anthocyanin biosynthesis, NBS-encoding genes**

Introduction

Rhododendron L., the largest genus in Ericaceae family, contains more than 1000 species and is widely distributed in Asia, Europe, and North America, with the center of diversity in southern China (Fang and Min, 1995; Chamberlain et al., 1996). To date, approximately 600 species have been confirmed from China, including many new species recognized in recent years (Tian et al., 2019). Because of their distinctive and colorful flowers, species within this genus are widely cultivated as horticultural plants with high ornamental, aesthetic and economic value. *Rhododendron delavayi* Franch. and *Rhododendron irroratum* Franch., are both evergreen and alpine species of *Rhododendron* subgenus *Hymenanthes* (Chamberlain, 1982) and widely distributed in south-west China. The *R. delavayi* has large, scarlet flowers and *R. irroratum* is often characterized by yellow corollas with yellow-green to lavender-red spots (Figure 1), which make them highly attractive and abundant new varieties have been selected from the *R. delavayi* and *R. irroratum* genetic resources.

As important ornamental plants, mining key candidate genes conferring flower coloration and stress tolerance are important for future breeding to develop new horticultural cultivars, and to enhance stress tolerance of *Rhododendron* cultivars to unfavorable environmental factors. The R2R3-MYB proteins accounted for the largest proportion of MYB superfamily in plants and are critical in controlling the biosynthesis of flavonoids/anthocyanin, which is the

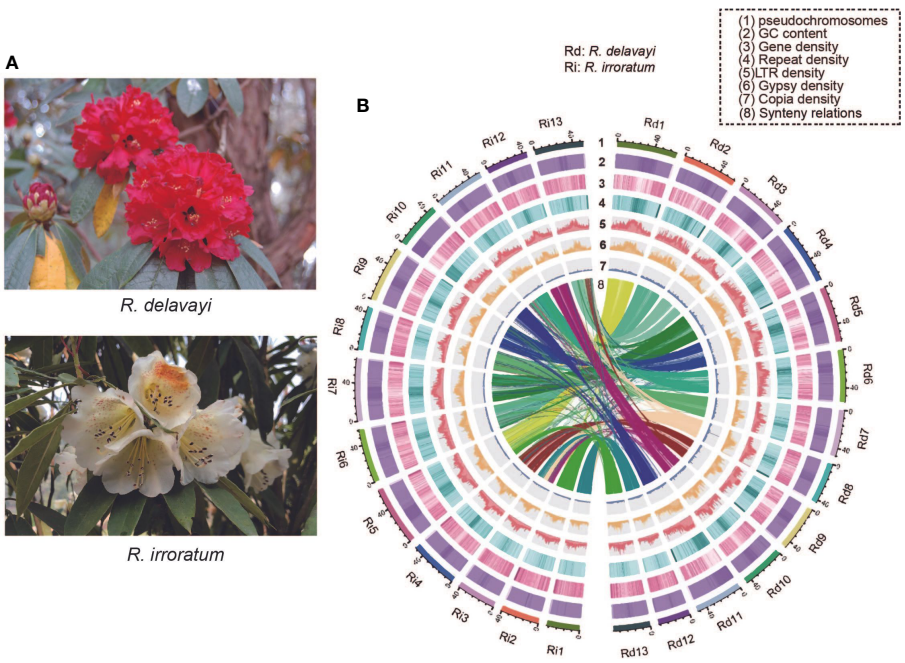

FIGURE 1
Flower morphology and genome features of *R. delavayi* and *R. irroratum*. **(A)** *R. delavayi* are characterized by bright red flowers and *R. irroratum* typically has pale yellow flowers. Photos: Hong Yang and Yan-Ting Hu, Baili Rhododendron Nature Reserve, Guizhou, China. **(B)** The landscape of the genome assemblies and annotations for *R. delavayi* and *R. irroratum*. The density was calculated in 100-kb sliding windows (Rd: *R. delavayi* and Ri: *R. irroratum*).

most important and widespread floral pigment (Tanaka et al., 2008; Wessinger and Rausher, 2012). Moreover, R2R3-MYB proteins are demonstrated to participant in responses to various stressful conditions, e. g., drought, cold and pathogen attack (Dubos et al., 2010). Nevertheless, the R2R3-MYB subfamily has not been well analyzed in *Rhododendron*. Thus, identifying and characterizing the *R2R3-MYB* genes in *Rhododendron* can provide reference for future functional verifications.

Through long periods of evolution, plants have evolved complicated defense mechanism to resist the invasion of various pathogens. The plant disease-resistance (R) genes play vital roles in recognition of virulence factors and inducing immune responses (Jones and Dangl, 2006; McDowell and Simon, 2006; Bent and Mackey, 2007). Nucleotide-binding site-leucine-rich repeat (*NBS-LRR*, NBS-encoding for short) genes make up the largest family (80%) of known plant R genes and are crucial in protecting plants against pathogenic organisms (Meyers et al., 1999; Dangl and Jones, 2001; Xiao et al., 2001; Meyers et al., 2003). The NBS-LRR proteins have been classified into three subclasses according to their phylogeny, that is, TIR-NBS-LRR(TNL), CC-NBS-LRR(CNL), and RPW8-NBS-LRR(RNL) (Shao et al., 2016; Shao et al., 2019). In recent years, *Rhododendron* flower rot, stem rot, root rot and leaf diseases are occurring with increasing frequently, causing serious losses of rhododendron plants and to local economies.

Up to date, several *Rhododendron* genomes have been released (Supplementary Table 1), including two draft rhododendron genomes (*R. delavayi* and *R. williamsianum*) obtained by Illumina short read sequencing (Zhang et al., 2017; Soza et al., 2019), two high-quality, chromosome-level azalea genomes (*R. simsii* and *R. ovatum*) generated by PacBio SMRT sequencing (Yang et al., 2020; Wang et al., 2021) and the recently sequenced high-quality genomes of *R. griersonianum* (Ma et al., 2021) and *R. henanense* (Zhou et al., 2022). These emerging genomic resources greatly promote the genomic research of *Rhododendron* and facilitate our understanding of the Ericaceae evolution. The available *R. delavayi* genome was generated using HiSeq Illumina sequencing, however the assembly was fragmented, which impedes the investigation of genome evolution and utilization of genetic resources. Therefore, the high-quality, chromosome-scale assembly for *R. delavayi* was constructed with Nanopore long read sequencing and Hi-C scaffolding in this study. Meanwhile, the *R. irroratum* was assembled to chromosome level by Combining PacBio SMRT (single-molecule real-time) and Hi-C sequencing. Phylogenetic analyses of *R. delavayi* and *R. irroratum* with other representative angiosperm species enabled us to further investigate the evolutionary relationships within the *Rhododendron* genus. Moreover, different patterns of gene duplications were searched and their contributions to gene family expansion were explored for the two sequenced genomes. Furthermore, the identification, comparative phylogenetic analysis and functional prediction of *R2R3-MYB* genes were performed in *Rhododendron* species. Finally, the systematic investigation of *Rhododendron* NBS-encoding genes were conducted to explore their chromosomal organization, evolutionary patterns and duplicated types of this gene family.

This study yielded insights into evolution history of the *Rhododendron* genus. It is hope that the comprehensive analysis of *R2R3-MYB* and NBS-encoding genes will provide basic and essential information for subsequent screening of the interesting flower coloration and disease resistance genes, which can facilitate genetic improvement of *Rhododendron* cultivars at molecular level.

Materials and methods

Plant materials and DNA sequencing

The plant materials of *R. delavayi* were collected from a 50-year-old tree growing in Jindian National Forest Park (Yunnan, China). This tree was transplanted from Cang Shan Mountain (Yunnan, China) in 1995. High-quality genomic DNA was isolated from fresh leaves with a QIAGEN® Genomic Kit and then purified using a QIAquick Gel Extraction Kit. The DNA library was constructed using the SQK_LSK109 Ligation Sequencing Kit and sequenced on a PromethION System (Oxford Nanopore Technologies, UK) using one Nanopore cell. After base-calling (Guppy v4.0.1) and data quality control, a total of ~55 Gb Nanopore long reads were obtained with an average length and N50 size of 17 Kb and 23 Kb, respectively (Supplementary Table 2). The Hi-C library was prepared with Dpn II restriction enzyme following the method of previous studies (van Berkum et al., 2010) and sequenced on an Illumina HiSeq X-Ten machine, finally generating ~68 Gb clean paired-end (PE) reads.

Fresh leaf tissues of *R. irroratum* were collected from a wild plant in Baili Rhododendron Nature Reserve (Guizhou, China). The materials were immediately frozen in liquid nitrogen and stored at −80°C for DNA extraction and library construction. The CTAB method was adopted to extract the genomic DNA. For short read sequencing, the short paired-end library (400 bp) was created and sequenced on the MGISEQ-2000 platform (PE100). Raw sequencing data was processed with fastp v0.20.1 (Chen et al., 2018) to obtain ~23 Gb clean data. For PacBio SMRT sequencing, the SMRTbell libraries with 20 Kb inserts were prepared and sequenced on the PacBio Sequel sequencer with 17 SMRT cells (Pacific Biosciences, CA, USA) using DNA Sequencing Reagent Kit 4.0 v2. After data filtering and preprocessing, ~71 Gb clean data were obtained, which had a mean subread length and N50 size of 7,681 bp and 12,150 bp, respectively (Supplementary Table 2). The method of Hi-C library preparation for *R. irroratum* was the same to *delavayi* using Dpn II restriction enzyme and ~74 Gb clean data were generated after sequencing on an Illumina Hiseq X-Ten platform.

Transcriptome sequencing and preprocessing

To support the gene annotation of *R. irroratum*, leaf and flower bud tissues were collected to prepare RNA samples and ~1 μg RNA

1 https://github.com/ben-lerch/IsoSeq-3.0

2 https://github.com/Nextomics/NextDenovo

were used to construct the sequencing libraries with NEBNext®Ultra™ RNA Library Prep Kit for Illumina®(NEB, USA) by following the manufacturer's instructions. The libraries were sequenced on the Illumina HiSeq 2000 platform (PE150) and ~42 Gb clean data were obtained (Supplementary Table 3). In addition, leaf and flower bud tissues from *R. irroratum* were mixed to perform PacBio Iso-Seq sequencing. The PacBio Iso-Seq3 pipeline[1] was applied to preprocess raw reads to generate high-quality, full-length and consistent isoform transcripts, finally producing 21,499 transcripts (Supplementary Table 3).

Estimation of genome size and heterozygosity

The genome size and heterozygosity for *R. delavayi* and *R. irroratum* were evaluated by *K*-mer analysis. In addition, flow cytometry measurement was also conducted to estimate the size of *R. irroratum* genome. Young leaves of *R. irroratum* were sampled at Kunming Botanical Garden (Yunnan, China) and nuclear DNA was isolated using a customed protocol. Then, nuclear DNA content was measured by flow cytometry using *Zea mays* 'B73' as reference standard. For *K*-mer analysis, Jellyfish v2.2.10 (Marcais and Kingsford, 2011) were applied to calculate *K*-mer counts and GenomeScope v1.0 (Vurture et al., 2017) estimated the genome size, repeat content, and heterozygosity.

Genome assembly and quality assessment

The Nanopore reads were applied to assemble contig sequences of *R. delavayi*. First, NextDenovo[2] was used to correct errors in the reads. Then, SMARTdenovo (Liu et al., 2020a), WTDBG[3], and NextDenovo were applied to assemble the corrected long reads, respectively. The three softwares yielded the assemblies of 765, 790 and 690 Mb, with contig N50 of 1,900, 500, and 11,000 Kb, respectively (Supplementary Table 4). Finally, the preassembly generated by NextDenovo was regarded as the optimal assembly for further analysis. Pilon (Walker et al., 2014) was applied to correct the contigs with pair-end short reads (~100 Gb) sequenced in the previous research (Zhang et al., 2017). After that, Juicer (Durand et al., 2016b) and 3D-DNA pipeline (Dudchenko et al., 2017) were used to anchor contig sequences into chromosome-scale scaffolds using Hi-C data. Boundaries of these scaffolds were manually adjusted in Juicebox (Durand et al., 2016a) and each scaffold was re-scaffolded in 3D-DNA, followed by manual adjustment within Juicebox to fix insert, bound, order, and mis-join errors. LR_Gapcloser (Xu et al., 2019) was used to close gaps twice by mapping the ONT long reads to Hi-C assembly. Subsequently, three rounds of genome polishing were conducted by NextPolish (Hu et al., 2020) with pair-end short reads. The heterozygous and redundant sequences in scattered contigs (contigs which were not anchored into chromosomes) were removed by Redundans (Pryszcz and Gabaldón, 2016). Finally, the *R. delavayi* genome sequences contained 13 pseudo-chromosomes, and 10 unplaced contigs. Three strategies were employed to assess the quality of final assembly. Firstly, the accuracy was evaluated by aligning Illumina and ONT reads to final genome using BWA-MEM (Li and Durbin, 2009) and minimap2 (Li, 2018), respectively. Next, BUSCO analysis (version 5.4.5, eudicots_odb10) was adopted to evaluate the completeness of the assembly. Finally, the continuity of assembly was evaluated by LAI index (Ou et al., 2018).

The *R. irroratum* genome sequence was assembled by integrating the PacBio, Hi-C and MGI-SEQ sequencing. To obtain an optimal primary assembly, SMARTdenovo, WTDBG, Canu (Koren et al., 2017) and FALCON/FALCON-Unzip (Chin et al., 2016) were initially used to assemble the PacBio subreads, producing four versions of primary assemblies (Supplementary Table 4). The assembly generated by FALCON and FALCON-Unzip was selected as the optimal assembly with the maximum N50 and fewest contigs. Subsequently, duplicated assembled haploid contigs were removed by Purge Haplotigs (Roach et al., 2018), reducing the assembled size from 853.63 to 700.47 Mb. The obtained curated contigs were scaffolded using the SSPACE-longread v1.1 (Boetzer and Pirovano, 2014). Then, ALLHiC pipeline (Zhang et al., 2019) anchored the scaffolds into super-scaffolds, which were subjected to inspection and manually adjusted in Juicebox v1.11.08 to generate chromosome-level scaffolds. Finally, the genome sequence was polished by Pilon with paired-end short reads. In result, the *R. irroratum* assembly was composed of 13 pseudo-chromosomes and 1440 unanchored contigs. Three methods used for quality control of the *R. delavayi* genome were also applied to evaluate the quality of *R. irroratum* assembly. PacBio subreads, pair-end short reads and full-length transcripts were mapped to final assembled sequences.

Genome annotation

Genome annotation of *R. delavayi* can be divided into two parts: annotation of repeat elements and annotation of gene structure and function. We employed a combined method of homology-based search and *ab initio* identification to annotate the repeat elements. For *ab initio* method, LTR_retriever (Ou and Jiang, 2018) and RepeatModeler v1.0.11 were used to find repeat elements, which were merged as the *ab initio* repeat library to identify repetitive sequences by RepeatMasker (Chen, 2004). The homology-based strategy was performed by searching a known repeat library (Repbase 20170127) with RepeatMasker. We combined repetitive sequences predicted by both methods and regarded them as the final repeat set. Then, the identified repeat elements were masked using RepeatMasker and gene models were predicted by MAKER2 pipeline (Holt and Yandell, 2011). The 83,515 transcripts from the mixed sample (containing five different tissues: flowers, flower buds, young leaves, mature leaves, and young stems) and 187,124 combined non-redundant protein sequences from *R. delavayi*, *R.*

3 https://github.com/ruanjue/wtdbg

4 http://transdecoder.github.io

simsii, *R. williamsianum*, and *Vaccinium corymbosum*, *Actinidia chinensis*, and *Arabidopsis thaliana* were considered as evidence of ESTs and homolog proteins, respectively.

The procedure of genome annotation for *R. irroratum* was similar to that of *R. delavayi*. Firstly, we identified the repeat elements by combining evidence from homology alignments and *ab initio* prediction. The homology-based approach was the same with *R. delavayi*. LTRharvest, LTR_FINDER (Xu and Wang, 2007) and RepeatModeler were applied to construct the *ab initio* repeat library. Then, the repeat database from the above two strategies was used to mask the genome using RepeatMasker. To perform gene structure predictions, the Illumina RNA-seq reads and high-quality full-length transcripts generated by Iso-Seq sequencing were aligned to the repeat-softmasked genome using STAR (Dobin et al., 2013) and GMAP (Wu and Watanabe, 2005), respectively. Proteins of *R. williamsianum*, *Solanum lycopersicum*, *Camellia sinensis*, *R. delavayi*, *A. chinensis* and *V. corymbosum* were selected for homologous protein alignment. The BRAKER2 pipeline v2.1.5 (Hoff et al., 2019), which combines GeneMark-ES/ET and AUGUSTUS, was applied to annotate gene models by integrating evidence from RNA-seq, Iso-Seq and homologous proteins. In addition, we also utilized the GeMoMa (Keilwagen et al., 2016; Keilwagen et al., 2018) to perform homology-based prediction by using genome sequences and gene models of *R. williamsianum*, *Solanum lycopersicum*, *Camellia sinensis*, *R. delavayi*, *A. chinensis* and *V. corymbosum*. Moreover, we employed RNA-seq and Iso-Seq data to conduct further RNA-seq-based prediction. Trinity v2.8.5 (Haas et al., 2013) was used to assemble the Illumina RNA-seq reads into transcripts. The obtained transcripts combined with full-length transcripts were aligned to final assembly by PASA (Campbell et al., 2006). Meanwhile, the candidate coding regions in the transcripts were identified by TransDecoder[4]. Finally, the consensus gene set was generated by combining the gene models predicted by different strategies with EVidenceModeler v1.1.1 (Haas et al., 2008)

Protein functional annotations were carried out by BLASTP (E-value: 1e−5) (Altschul et al., 1990) homology searches against the NCBI non-redundant protein database (2020-7-10) and Swiss-Prot database (2020-8-12). InterProScan package (version 5.29-68.0) was used to perform a comprehensive annotation such as Pfam, GO and KO. EggNOG-mapper v1.0.3 was used for functional annotation based on the eggNOG database. GO and KO information was retrieved from InterPro and eggNOG annotation.

Gene family, comparative genomics and evolutionary analysis

We selected 19 representative angiosperms (including *R. delavayi* and *R. irroratum*) (Supplementary Table 5) to identify orthogroups (orthologous gene families) by using the OrthoMCL pipeline[5] (parameter: min_length: 50, inflation: 1.5, https://github.com/apetkau/orthomcl-pipeline), which used the all-versus-all BLASTP (E-value: 1e-5) results of all proteins. The single-copy orthogroups (orthogroups that contained one only gene for each species) were selected for phylogenetic analysis with concatenated approach. First, the proteins of each single-copy orthogroup were aligned by MAFFT v3.8.1551 (Katoh and Standley, 2013) and the protein alignments were converted into corresponding coding sequence (CDS) alignments using PAL2NAL v14 (Suyama et al., 2006), followed by refinement with trimAI v1.480 (Capella-Gutierrez et al., 2009). Then, CDS alignments were concatenated to a super gene matrix. jModelTest v2.1.7 (Posada, 2008) was applied to select the optimal substitution model. Finally, a maximum likelihood (ML) tree was constructed using RAxML v8.2.12 (Stamatakis, 2014) with the concatenated CDS alignments based on GTR+I+GAMMA model (1000 bootstrap tests). Orthogroups expansion and contraction for each species were inferred by CAFE v4.2.1 (De Bie et al., 2006). Orthogroups were filtered out when one or more species contains more than 100 gene copies. Divergence times among the 19 species were estimated using MCMCtree implemented in PAML package (Yang, 1997) with 'JC69' model and parameters 'burn-in = 10,000,000, nsample = 200,000, and sample-frequency = 5,000'. Two fixed fossil ages: (1) Ericales (89.8 Mya) (Nixon and Crepet, 1993), (2) *Rhododendron* (56 Mya) (Collinson and Crane, 1978) and five soft-bound calibration points acquired from document (Eudicots: 125–161 Mya, Angiosperms: 199–167 Mya and Asterids–Rosid: 116–126 Mya) (Li et al., 2019), and TimeTree website (*A. comosus–O. sativa*: 102–120 Mya and *A. thaliana–V. vinifera*: 107–135 Mya) to were used to calibrate the age of the nodes in the tree.

Whole-genome duplication analysis

Three approaches were adopted to elucidate the WGDs in *Rhododendron*. First and foremost is the synonymous (*Ks*) substitution rates distribution. Collinear blocks (more than five collinear genes) within and between species were identified by MCScanX (Wang et al., 2012) based on the BLASTP (E-value: 1e-5) results of proteins. The protein-coding DNA alignments for collinear gene pairs were performed using ParaAT v2.0 (Zhang et al., 2012). *Ks* values of the collinear gene pairs were calculated by KaKs_Calculator v2.0 (Wang et al., 2010) with the YN model. Then, the microsynteny of grape and *R. delavayi* and of grape and kiwifruit were depicted by MCScan Python version[6]. Moreover, to explore whether kiwifruit and *Rhododendron* shared the same Ad-β WGD or experienced lineage-specific WGDs shortly after their divergence, we extracted collinear gene pairs falling in the *Ks* peak ± 1 s.d. as paralogous pairs that likely derived from Ad-β event in *R. delavayi* and kiwifruit, respectively, referring to the methods described by Teh (Teh et al., 2017). Next, gene family clustering was performed using proteins of kiwifruit, *R. delavayi* and *C. canephora* with the OrthoMCL pipeline and orthogroups that contained one such paralogous pair of *R. delavayi*, one such paralogous pair of kiwifruit, and one gene from *C. canephora* were extracted. In each of such orthogroups, each *R. delavayi* paralog was matched with its kiwifruit ortholog (among the two paralogs), by setting the *R. delavayi*–kiwifruit pair with the lowest synonymous

5 https://github.com/apetkau/orthomcl-pipeline

6 https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version)

distance as orthologs. RAxML constructed the phylogenetic relationships of the paralogs and orthologs in *R. delavayi* and kiwifruit, along the genes from *C. canephora*, which experienced no additional WGDs expect the ancient WGT-γ.

The time of β WGD event in *Rhododendron* was estimated according to the formula T = *Ks*/2r, where r is synonymous substitutions at each site per year, which was calculated according to the mean *Ks* peaks of BLASTP reciprocal best hit (RBH) pairwise sequences of kiwifruit vs. five *Rhododendron* species (*R. delavayi*, *R. williamsianum*, *R. irroratum*, *R. simsii* and *R. ovatum)* and kiwifruit - *Rhododendron* divergence time (median: 71 Mya) as an age constraint, following r=*Ks*/(2 × (divergence time)).

Analysis of gene duplication

To further understand the gene duplication and evolution in *R. delavayi* and *R. irroratum* genomes, we determined different classes of gene duplications with *DupGen_finder* pipeline (Qiao et al., 2019), which categorized the duplicated genes into 5 types: whole-genome duplicates, tandem duplicates (adjacent gene copies in the same chromosome), proximal duplicates (gene pairs located on nearby chromosomal region but separated by less than 10 genes), transposed duplicates (gene pairs originating from DNA-based transposition or retrotransposition) and dispersed duplicates (other types excluding the above mentioned). The *Ks* values for each type of duplicated pairs were computed by KaKsCalculator v2.0.

Identification and characterization of R2R3-MYB subfamily transcription factors

The R2R3-MYB subfamily members belonging to MYB superfamily were identified in four *Rhododendron* species (*R. delavayi*, *R. irroratum*, *R. ovatum*, and *R. simsii*) investigated in this study. *R. williamsianum* was excluded from the analysis due to the poor quality of its genome sequence. The Hidden Markov Model (HMM) profile of MYB DNA-binding domain (PF000249) was downloaded from the Pfam website[7] and used as a query to search the *Rhododendron* protein sequences using HMMER v3.1b2 (E-value: 1e-5). In addition, PlantRegMap[8] was also used to predict transcription factors with best hits in *A. thaliana*. The obtained MYB protein sequences were merged and redundant sequences were removed. Then, proteins that did not have the homologous SwissProt annotation (E-value: 1e-5) or included incorrect domains depending on TAPscan v.2 transcription factor database were filtered. The obtained protein sequences were examined by SMART[9], Pfam, and CD search[10] analyses to validate the existence of MYB domain. Finally, the candidate sequences that contained two DNA-binding repeats were categorized as the R2R3-MYB subfamily. The MEME online tool was used to analyze conserved motifs of R2R3-MYB proteins (Bailey et al., 2006) with the following parameter settings: maximum number of different motifs, 20; minimum motif width, 6; and maximum motif width, 50. The exon/intron organizations of *R2R3-MYB* genes were visualized by TBtools v1.098696 (Chen et al., 2020). Chromosomal distributions of *Rhododendron R2R3-MYB* genes were retrieved from genome annotation files.

To explore the subgroup classification and phylogenetic relationships of *R2R3-MYB* proteins in *Rhododendron*, 125 R2R3-MYB protein sequences from *Arabidopsis* were retrieved from the Arabidopsis Information Resource[11] (TAIR) and these proteins together with R2R3-MYB proteins from four *Rhododendron* species were aligned by MAFFT and a ML phylogenetic tree was constructed using IQ-TREE v1.6.12 (Nguyen et al., 2015) with 1000 bootstrap tests. The classification and categorization of *Arabidopsis R2R3-MYB* genes assisted the subdivision of *Rhododendron R2R3-MYB* genes. BLASTP (E-value: 1e-10) and MCScanX were applied to determine intraspecific synteny of the *Rhododendron R2R3-MYB* genes. TBtools plotted the 1:1 collinearity genes among four *Rhododendron* species.

The contributions of different gene duplications to *R2R3-MYB* subfamily gene expansion in *Rhododendron* species were further investigated by identifying the duplicated types of R2R3-MYB genes using *DupGen_finder*.

Identification and evolutionary analyses of NBS-encoding genes

The HMM searches were also conducted to identify candidate NBS-encoding genes in kiwifruit and four *Rhododendron* plants. The HMM profile of NB-ARC domain (PF00931) was downloaded from Pfam website and used to search the protein sequences of kiwifruit and four *Rhododendron* species with HMMER v 3.1b2. The high quality sequences (E-value: 1e-20) obtained were used for multiple alignment with ClustalW (Thompson et al., 1994) and species-specific NBS profiles were constructed using HMMER software with the "hmmbuild" module. Using the species-specific profiles, the candidate NBS-encoding genes were obtained for each species and then submitted to online Pfam database to further confirm the presence of NB-ARC domain. All candidate NBS-encoding genes were further inspected using Pfam, SMART, CD search to examine whether they had TIR, RPW8, or LRR domains. Moreover, COILS program (Lupas et al., 1991) identified the CC motifs using the threshold of 0.9. Chromosomal locations of the NBS-encoding genes were retrieved from genome annotation files and cluster assignment of NBS-encoding genes followed the criterion used for *Medicago truncatula* (Ameline-Torregrosa et al., 2008): if the interval of two adjacent genes on the same chromosome was less than 250 Kb, these two genes were considered to be in the same cluster. According to this standard, the NBS-encoding genes were assigned to singletons and clustered loci.

7 https://pfam.xfam.org/

8 http://plantregmap.gao-lab.org/

9 http://smart.embl-heidelberg.de/

10 https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

11 http://www.arabidopsis.org/

The amino acid sequences of NB-ARC domain were extracted from the NBS-encoding genes of four *Rhododendron* species and kiwifruit and subjected to multiple sequence alignment with ClustalW integrated in MEGA-X (Kumar et al., 2018). After removing too short and exceedingly divergent sequences from the alignment, the ML NBS-encoding gene phylogeny was constructed using IQ-TREE with bootstraps of 1000 replicates. To understand the evolutionary history of this gene family, two phylogenies (CNL and nCNL (RNL and TNL)) were reconstructed for four *Rhododendron* species and kiwifruit. Gene duplication and loss events in the CNL and nCNL phylogenetic trees were restored by reconciling these two gene trees with real species tree through Notung software v2.9 (Stolzer et al., 2012). On the reconciled trees, if the NBS-encoding genes constituted a monophyletic clade that originated from the common ancestor of four *Rhododendron* species, then this clade was defined as a *Rhododendron* NBS-encoding lineage. If a monophyletic NBS-encoding gene branch originated in the common ancestor of *Rhododendron* and kiwifruit, this branch was defined as an Ericales NBS-encoding gene lineage. The duplications and losses of NBS-encoding genes during different evolutionary periods were restored at each node of the phylogenetic trees.

The roles of different gene duplications in expanding the *Rhododendron* NBS-encoding genes were also investigated following the method described above.

Functional enrichment analysis

GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analyses were performed using R package clusterProfiler (Yu et al., 2012) by setting all annotated *R. delavayi* or *R. irroratum* genes as background gene set. We used the corrected *P*-value (q value) < 0.01 as threshold criteria to judge whether a specific GO term or KEGG pathway was significantly overrepresented.

Results

Genome assembly and annotation

The genome size of *R. delavayi* and *R. irroratum* were estimated to be ~647 Mb and ~597 Mb by *K*-mer analysis, respectively (Supplementary Figure S1). However, flow cytometry analysis estimated the haploid genome size of *R. irroratum* to be 730 Mb, which was higher than the *K*-mer method for yet unknown reasons (Supplementary Figure S1). Compared to *R. delavayi*, *R. irroratum* demonstrated a relatively high heterogeneity rate (Table 1; Supplementary Table 1). Here, ~ 55 Gb of Nanopore long reads and ~ 71 Gb of PacBio subreads were generated for *R. delavayi* and *R. irroratum*, respectively. Several assemblers were employed to conduct initial assembly and the optimal primary assemblies with a contig

TABLE 1 Genome survey, nuclear genome assembly and annotation features of *R. delavayi* and *R. irroratum*.

Feature	*R. delavayi*[17]	*R. delavayi* (this study)	*R. irroratum* (this study)
Genome survey			
Estimate genome size (Mb)	~697–717	647 (*K*-mer)	597 (*K*-mer)
Repeat content	–	65.6%	62.75%
Heterozygosity	0.9–1.1%	~1.1%	~1.55%
Genome assembly			
Assembly size (Mb)	695	662. 80	701.62
Contig No.	209, 969	216	1,549
Contig N50 (bp)	61, 801	11,000,000	695,292
Scaffold No.	193,091	23	1,413
Scaffold N50 (bp)	637,826	54,547,292	51,021,190
Chromosome-scale length (bp)	–	659,801,287 (99.56%)	652,297,987 (92.96%)
Complete BUSCOs	92.80%	97.4%	96.8%
LAI index	–	11.17	13.72
Genome annotation			
Repeat content	51.77%	51.00%	47.71%
No. of predicted protein coding genes	32,938	40,671	49,421
No. of functionally annotated (%)	85.91%	92.19%	88.92%
Complete BUSCOs	87.4%	90.9%	88.9%

Short line (-) indicated that data were not shown in the original article. Contig No. and Contig N50 were calculated from the optimal primary (contig-level) assembly. Scaffold No. and Scaffold N50 were calculated from the final chromosome-scale assembly. The assembly and annotation only considered the nuclear genome.

N50 of 11 Mb for *R. delavayi* and of 0.60 Mb for *R. irroratum*, were used for subsequent improvement (Supplementary Table 4). Through Hi-C scaffolding, 99.56% (659.80 Mb) of the sequences for *R. delavayi* and 92.96% (652.29 Mb) for *R. irroratum* were anchored, ordered and oriented into 13 pseudochromosomes (Table 1; Supplementary Figure S2). Three strategies were adopted to evaluate the quality of the assemblies. The accuracy of the assemblies were evidenced by the high genome coverage (> 98% for both *R. delavayi* and *R. irroratum*) and mapping rates (> 90% for both *R. delavay* and *R. irroratum*) (Supplementary Table 6). BUSCO analysis revealed that 2,264 (97.4%) and 2,251 (96.8%) of the 2,326 conserved single-copy orthologs of eudicots were completely recovered in the *R. delavayi* and *R. irroratum* genomes, respectively (Supplementary Table 7), indicating the completeness of the assemblies. Furthermore, high LTR Assembly Index (LAI) scores of 11.17 and 13.72 were estimated for *R. delavay* and *R. irroratum*, respectively, suggesting that the two newly assembled genomes reached the reference level.

By combining the evidence from homology searches and *ab initio* prediction, 51.04% (338.93 Mb) and 47.71% (334.77 Mb) repeat elements were annotated in *R. delavayi* and *R. irroratum*, respectively (Supplementary Table 8). Long terminal repeat retrotransposons (LTR-RTs) are the predominant type of repeat elements and occupied 38.15% (253.29 Mb) of the *R. delavayi* and 32.78% (229.97 Mb) of the *R. irroratum* genome. DNA transposons accounted for 2.45% (16.29 Mb) of the *R. delavayi* and 2.00% (14.06 Mb) of the *R. irroratum* genome. The protein coding gene models were annotated using both evidence-based and *ab initio* prediction. In total, 40,671 and 49,421 coding genes were predicted for *R. delavayi* and *R. irroratum*, respectively. The annotated gene number of the current *R. delavayi* genome exceeded that of previous genome obtained by Illumina short read sequencing (32,938). The predicted genes were characterized by a mean length of 4,707 bp for *R. delavayi* and of 4,323 bp for *R. irroratum*, and an average CDS length of 1,195 bp for *R. delavayi* and 1,096 bp for *R. irroratum* (Supplementary Table 9). Of the annotated coding genes, 37,586 (92.19%) in *R. delavayi* and 43,946 (88.92%) in *R. irroratum* could be functionally annotated (Supplementary Table 10). Additionally, 99.69% of the *R. delavayi* and 93.96% of the *R. irroratum* genes could be assigned to the 13 pseudo-chromosomes. The annotation quality was also estimated based on the BUSCO analysis and 90.9% complete BUSCOs in *R. delavayi* and 88.9% in *R. irroratum* were detected, respectively (Supplementary Table 11). The GC content, gene density, repeat density, LTR density, Gypsy density and Copia density throughout the genome were given in Figure 1B.

Phylogenetics and genome evolution

To elucidate genome evolutionary history in *Rhododendron* species, gene family (orthogroup) clustering were carried out using proteins of *R. delavayi*, *R. irroratum* and 17 other sequenced angiosperms. Totally, 46,033 orthogroups were identified, of which 5,213, 7,954, 10,437 orthogroups were shared by all investigated angiosperms, Ericales and *Rhododendron*, respectively (Supplementary Figure S3; Supplementary Table 12). The 10,437 *Rhododendron*-shared gene families can considerably signify the "core" proteomes of *Rhododendron* species (Figure 2B) and they were enriched in functional categories termed "biosynthetic, metabolic and catabolic process", and "response to fungus/bacterium/nematode" involved in metabolism and stress response based on GO enrichment analyses (Supplementary Figure S4). A phylogenetic tree was inferred for the 19 selected angiosperms, using 253 single-copy gene families (Figure 2A). The phylogenetic placement of main lineages agreed with previous studies (The Angiosperm Phylogeny Group et al., 2016). The *Rhododendron* genus was further divided into two clades: one clade comprising *R. delavayi*, *R. irroratum* and *R. williamsianum* (the evergreen rhododendrons) and the other containing *R. simsii* and *R. ovatum* (the azaleas). Based on calibration points, *Rhododendron* diverged from *Actinidia* (kiwifruit) at around 77.75 million years ago (Mya) and the divergence time of the two *Rhododendron* clades were estimated at around 27.76 Mya (Figure 2A). A total of 6,005 expanded genes in *R. delavayi* and 8,717 in *R. irroratum* belonged to 1,186 and 1,559 rapidly expanded gene families (EGFs), respectively (Supplementary Table 13, Figure 2A). KEGG enrichment analyses suggested the rapidly EGFs of two sequenced species in this study were both highly enriched in "plant-pathogen interaction", "cutin, suberine and wax biosynthesis", "terpenoid biosynthesis" and "anthocyanin biosynthesis" pathways (Supplementary Figure S5). Besides, GO enrichment for the rapidly EGFs revealed that there was a remarkable enrichment in GO terms involved in biotic stimuli and defense responses in the two species, for example, "cellular response to biotic stimulus", "defense response to insect/fungus", "immune response-activating signal transduction", "salicylic acid or jasmonic acid biosynthetic/metabolic processes" and "salicylic acid or jasmonic acid mediated signaling pathway" (Supplementary Table 14). Salicylic and jasmonic acid are plant hormones which can mediate defense responses against abiotic and biotic stresses (Browse, 2009; Kim et al., 2017). Therefore, the rapid expansion of stress-responsive gene families might underlie strong stress resistance of the *Rhododendron* species. To investigate potential WGD events in *Rhododendron* species, *Ks* was calculated for collinear paralogous genes in five *Rhododendron* species, as well as *A. chinensis* (kiwifruit) and *V. vinifera* (grape). The age distribution of *Ks* showed three peaks for kiwifruit, representing the three putative WGDs (Ad-α; Ad-β and At-γ), which supported the findings of previous reports (Huang et al., 2013; Wang et al., 2018; Wu et al., 2019). For the five *Rhododendron* species, two *Ks* peaks were observed, suggesting that the *Rhododendron* experienced only two rounds of polyploidy events (Figure 2C), of which the earliest event is shared by all core eudicots (γ triplication, WGT-γ). To further infer the polyploidy in *Rhododendron* species, we conducted collinearity and synteny analysis of *R. delavayi* genome (as a representative of *Rhododendron*) with grape and kiwifruit. Visualization of macrosynteny showed some ancestral regions in grape had two corresponding copy regions in *R. delavayi*, and that there were approximately two copies of each of these regions from *R. delavayi* in kiwifruit (Supplementary Figure S6). *Ks* distribution of collinear orthologs between kiwifruit (Actinidiaceae, *Actinidia*) and five studied *Rhododendron* plants revealed that *Rhododendron* and *Actinidia* diverged within the time frame of the Ad-β WGD in kiwifruit (Figure 2D). Previous studies suggested that *Actinidia* and

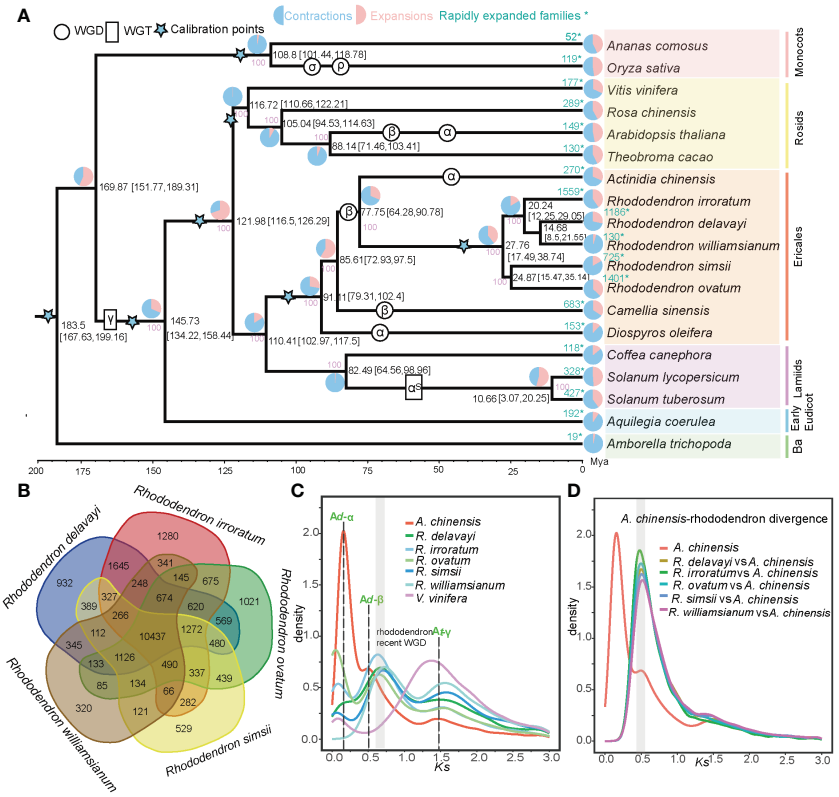

FIGURE 2
Evolutionary and comparative genomic analyses of *Rhododendron* species and representative angiosperms. **(A)** The phylogenetic tree of 19 angiosperms. All branches had 100% bootstrap support. Divergence time is displayed with 95% HPD in parentheses. Blue stars represented the calibration points. Estimated WGD/WGT events were placed according to results from our study and from those of previous investigations. The pink, light blue and light green numbers along particular branches indicated the number of orthogroups that undergone expansion, contraction and rapid evolution, respectively, * indicated the rapidly expanded families. **(B)** Shared and unique orthogroups among five *Rhododendron* species. **(C)** Synonymous substitution rate (*Ks*) distributions of syntenic paralogous genes for five.Rhododendron species, *V. vinifera* and *A chinensis* Ad-α; Ad-β and At-γ are illustrated with dotted lines. **(D)** *Ks* distributions of syntenic orthologous genes between *A chinensis* and five *Rhododendron* species.

Rhododendron shared the same WGD (Ad-β) that occurred near the Cretaceous-Tertiary (K-T) boundary (Wu et al., 2019). Here, we further explored whether the Actinidiaceae and Ericaceae shared the same β WGDs. The paralogous gene pairs of kiwifruit and *R. delavayi* (Ericaceae) from their β WGDs were extracted, respectively and orthologous relationships between them were established (Supplementary Table 15). Then, a phylogenetic tree was constructed for these genes, which revealed that the divergence of *R. delavayi* and kiwifriut paralogous genes was earlier than that of *R. delavayi*–kiwifriut orthologous genes (Supplementary Figure S7), implying that they shared the same Ad-β WGD, which is likely to have occurred before their lineages diverged. Consistent with previous research (Yang et al., 2020), the recent β WGD event in *Rhododendron* occurred ~78 Mya.

Gene duplication contributes differently to gene family expansion

Gene duplication has been considered as a key driving force of evolutionary novelties in plants (Flagel and Wendel, 2009; Panchy et al., 2016). Duplicated genes can be categorized into five types according to their origins (Qiao et al., 2019): whole-genome duplication (WGD), tandem duplication (TD), transposed duplication (TRD), proximal duplication (PD) and dispersed duplication (DSD). In addition to WGD, other patterns of duplications are always regarded as single-gene duplications (Qiao et al., 2019). Here, we identified 30,562 and 37,653 duplicated genes in *R. delavayi* and *R. irroratum* genomes, respectively and classified each as belonging to one of the five categories (Figure 3A; Supplementary Table 16). In both species, the DSD-derived genes accounted for the largest proportion. Two *Ks* peaks were found for gene pairs originating from WGDs, representing the two WGD events occurred in *Rhododendron*. Notably, gene pairs derived from TD and PD events exhibited only one *Ks* peak with lower *Ks* values that occurred after the WGD-β event, indicating the TD and PD duplications burst more recently (Figure 3B).

We analyzed the contribution of different duplicates to gene family expansion by searching for overlaps between each category of gene duplications and the rapid EGFs (Figure 3C). Of all the duplication types, DSD contributed the largest proportion to gene family expansion, followed by TD and PD, in both *R. delavayi* and *R. irroratum*. The KEGG enrichment analysis for genes overlapping

10.3389/fpls.2023.1123707

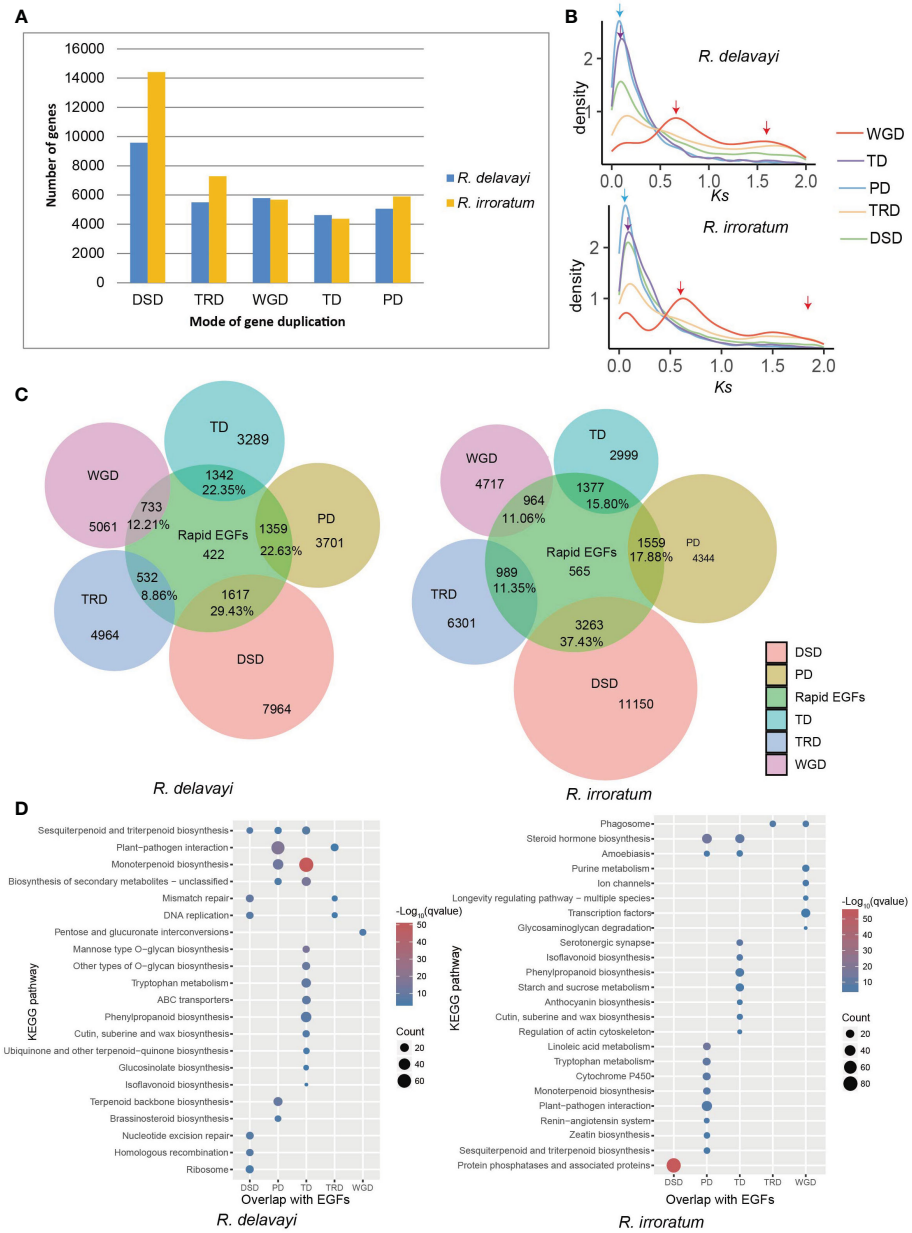

FIGURE 3
Gene duplication in *Rhododendron* and its contribution to gene family expansion. **(A)** Number of different kinds of duplicated genes in *R. delavayi* and *R. irrorarum*. WGD, whole-genome duplication; TD, tandem duplication; PD, proximal duplication; TRD, transposed duplication; DSD, dispersed duplication. **(B)** The *Ks* distributions of gene pairs derived from different modes of duplication. The red, purple, and blue arrows indicated the *Ks* peak of WGD-derived, TD-derived and PD-derived genes, respectively. **(C)** The relations between members of rapid expanded gene family (rapid EGFs) and members derived from different modes of duplication in *R. delavayi* and *R. irroratum*. **(D)** KEGG enrichment of genes in rapid EGFs and the different modes by which these genes have duplicated. The color and size of the circles represent statistical significance and the number of genes in a KEGG pathway, respectively; 'qvalue' is the correted P value.

between rapid EGFs and WGD-TD-PD-TRD-DSD indicated that different duplication-induced gene family expansions were enriched in divergent KEGG pathways (Figure 3D). For instance, the expansion of gene families relevant to "phenylpropanoid", "isoflavonoid" and "cutin, suberine and wax" biosynthesis were largely caused by TD, and expansion of "plant-pathogen interaction" genes was primarily attributed to PD events in both species. Moreover, in *R. delavayi*, the expansion of gene family participating in "terpenoid backbone biosynthesis" predominantly resulted from PD events, while the gene family expansions of "sesquiterpenoid and triterpenoid" and "monoterpenoid" biosynthesis were mainly induced by TD and PD events. In comparison, in *R. irroratum*, PD mainly contributed to the gene expansion of "monoterpenoid", "sesquiterpenoid and triterpenoid" biosynthesis and the expansion of "anthocyanin biosynthesis" gene families was largely due to TD events. WGD events largely accounted for the expansion of gene families involved in basic biological functions or transcription factors, e.g. pentose and glucuronate

interconversions in *R. delavayi* and purine metabolism, ion channels, transcription factors in *R. irroratum*. It is observed that the DSD events mainly contributed to expansion of gene families relevant to DNA replication, repair and homologous recombination in *R. delavayi*. Taken together, the contributions of different duplication modes to gene family expansion were different and the newly generated TD/PD might be important sources of expansion for gene families related to secondary metabolite biosynthesis and pathogen resistance. The flavonoids, terpenoids (isoprenoids) and 'cutin, suberine and wax' are substantially characterized as defense metabolites (barriers) because of their critical roles in plant abiotic or biotic stresses (Nakabayashi and Saito, 2015; Lepiniec et al., 2006; Pu et al., 2021; Lee and Suh, 2015). Thus, TD/PD duplications would enhance the ability of stress tolerance to various environmental stresses in the investigated species.

Evolution and comparative investigation of the R2R3-MYB subfamily

We searched the proteomes of four studied *Rhododendron* species to identify candidate MYB proteins using the HMM method. The proteins that contained two MYB DNA-binding repeats (Rs) were considered as the R2R3-MYB subfamily proteins. Finally, 102, 118, 151 and 133 *R2R3-MYB* genes were obtained for *R. delavayi*, *R. irroratum*, *R. ovatum*, *R. simsii*, respectively. Based on the genome annotations, 102 *R. delavayi*, 113 *R. irroratum*, 151 *R. ovatum* and 125 *R. simsii R2R3-MYB* genes were located on 13 chromosomes, respectively, which showed uneven distributions in each of the four *Rhododendron* species (Supplementary Figure S8). For example, chromosome 8 had the largest number of genes (14), while, chromosome 2 had only one gene in *R. delavayi*. The total number of *R2R3-MYB* genes of *R. delavayi* and *R. irroratum* was less than that of *R. ovatum* and *R. simsii*. To figure out the relatively small number of *R2R3-MYB* genes in *R. delavayi* and *R. irroratum* was due to gene loss or failure of gene annotations, CDS sequences of *R. irroratum* R2R3-MYBs were used as query to search the *R. delavayi* genome sequence with BLAT v.36x2 and vice versa. Finally, only 11 and 10 matched genomic regions were not structurally annotated in the *R. delavayi* and *R. irroratum* genomes, respectively. Therefore, the lower number of *R2R3-MYB* genes in *R. delavayi* and *R. irroratum* was likely caused by gene loss, pseudogenization and failure of annotation. To understand their group classification and evolutionary relationship, a phylogenetic tree was constructed using all R2R3-MYB proteins of *Arabidopsis* and four *Rhododendron* species. The R2R3-MYB members were separated into 42 subgroups (designed as R1 to R42) (Figure 4A; Supplementary Figure S9) based on the tree and subgroup classification of R2R3-MYB proteins of *Arabidopsis*. Compared to *Arabidopsis*, the R2R3-MYB members in 9 and 13 subgroups exhibited expansion and contraction for all four *Rhododendron* species, respectively. Specifically, the 9 expanded subgroups (R1, R2, R9, R12, R15, R24, R27, R29 and R32) consisted of *Rhododendron* specific members, implying their possible distinctive roles in *Rhododendron*. The *R. delavayi* gene *Rhdel07G0021400* was not categorized into any subgroup. The exon/intron patterns and conserved protein motifs of the *R2R3-MYB* genes suggested that most members in each subgroup were usually characterized by similar exon/intron patterns and motif distributions, whereas members belonging to different subgroups might have different exon/intron organizations and motif characteristics, e.g., the members belonging to subgroup R1 had two exons and one intron, while members of subgroup R3 had more than six exons and six introns (Supplementary Figure S9). Therefore, the gene structures and protein motif composition showed relatively higher conservation for *R2R3-MYB* genes of the same subgroup, which might be conserved in evolution. Considering that gene family members clustered together in the phylogeny usually tended to possess similar biological functions or control the identical metabolic pathway, it is possibility to perform function prediction of the *R2R3-MYB* genes in *Rhododendron*, based on their phylogenetic relationships with the *Arabidopsis* orthologs, of which the functions have been well studied. It has been reported that *AtMYB75/PAP1*, *AtMYB90/PAP2*, *AtMYB113* and *AtMYB114* (subgroup 6) were involved in the regulation of anthocyanin biosynthesis in *Arabidopsis* (Gonzalez et al., 2008). On the phylogenetic tree, these four genes were clustered with 15 *Rhododendron R2R3-MYB* genes, including one in *R. delavayi* (*Rhdel08G0093800*), four in *R. irroratum (Ri_282360*, *Ri_185830*, *Ri_282290*, *Ri_282390)*, three in *R. ovatum* (Ro_22285, Ro_22282, Ro_22319) and four in *R. simsii* (*Rhsim08G0087300*, *Rhsim08G0087200*, *Rhsim08G0087400*, *Rhsim10G0209900*, *Rhsim08G0087500*, *Rhsim08G0087600*, *Rhsim08G0087800*), which can be important candidate genes for exploration of their functions in anthocyanin synthesis and flower coloration in *Rhododendron* species. In addition to plant secondary metabolism, the *R2R3-MYB* proteins also participate in stress responses. For example, *Arabidopsis R2R3-MYB* genes such as *AtMYB60*, and *AtMYB96* (subgroup 1) have been proven to participate in drought response through mediating stomatal movement (Cominelli et al., 2005) and abscisic acid signaling pathway (Seo et al., 2009). In addition, *AtMYB96* can also be involved in ABA-mediated pathogen resistance (Seo and Park, 2010). These genes were clustered with 16 *Rhododendron R2R3-MYB* genes, including three in *R. delavayi* (*Rhdel10G0043900*, *Rhdel12G0197900*, *Rhdel08G0197900*), three in *R. irroratum* (*Ri_139470*, *Ri_195160*, *Ri_293310*), six in *R. ovatum* (*Ro_35465*, *Ro_09930*, *Ro_29255*, *Ro_14416*, *Ro_06615*, *Ro_31578*) and four in *R. simsii* (*Rhsim10G0039500*, Rhsim03G0045800, *Rhsim12G0156400*, *Rhsim08G0155100*), which can be further studied to illuminate their functions in stress resistance in *Rhododendron*.

The interspecific synteny of the *R2R3-MYB* genes among four *Rhododendron* species was investigated to explore their evolutionary conservation. First, the R2R3-MYB genes in syntenic blocks between any two genomes were determined based on the MCScanX analysis and 41 one-to-one orthologous genes among the four species were selected (Figure 4B; Supplementary Table 17), suggesting their conserved characteristics during the evolution. The duplicated mechanisms that generated the *R2R3-MYB* subfamily were also

10.3389/fpls.2023.1123707

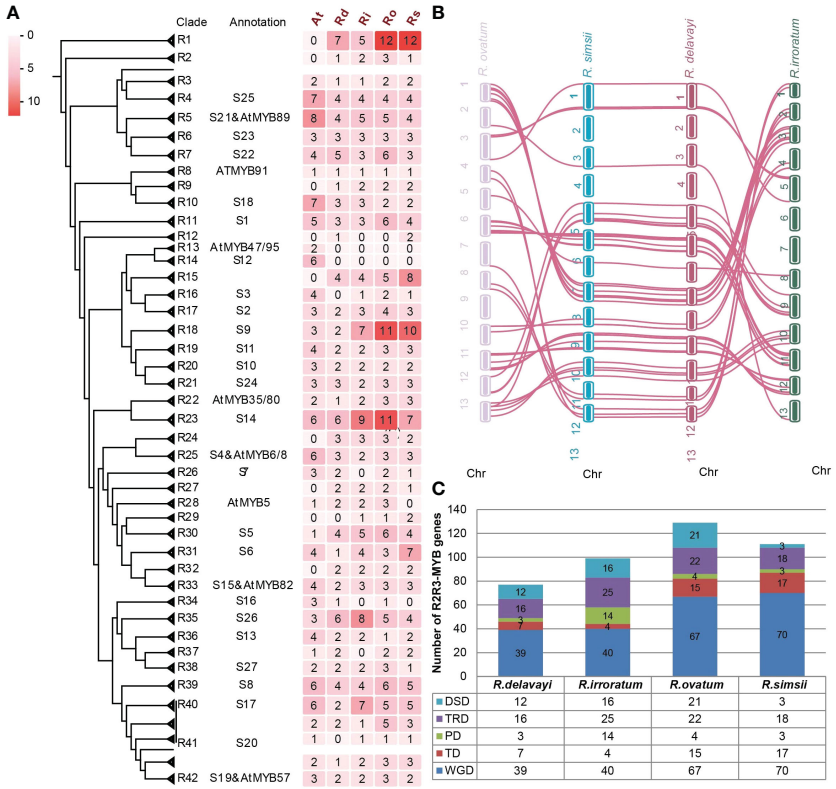

FIGURE 4
Comparative evolutionary analysis of the R2R3-MYB subfamily members in *Rhododendron*. **(A)** Phylogenetic tree of R2R3-MYB members from four *Rhododendron* species and *Arabidopsis*. The folding triangles represented the 42 defined subgroups (R1 to R42). One protein from *R. delavayi* (line) did not fit well into subgroups. Subgroups clustered with Arabidopsis R2R3-MYB proteins were annotated with designated *A thaliana* subgroups as previously reported (Dubos et al., 2010). Number of subgroup members for each species was indicated by the heatmap. The uncompressed tree was shown in Supplementary Figure S13. **(B)** Interspecies synteny of *R2R3-MYB* genes among four *Rhododendron* genomes; **(C)** Number of *R2R3-MYB* genes produced by distinct gene duplications.

analyzed and results showed that WGD events contributed to more than 50% of the overall gene number in each of the four R*hododendron* plants (but not in *R. irroratum*) (Figure 4C).

Evolution of the NBS-encoding genes

In this study, the NBS-encoding genes were identified and their evolutionary patterns were investigated during the speciation of four *Rhododendron* species. Using the HMM method, a total of 1,336 candidate genes were identified in four *Rhododendron* species: 332 from *R. delavayi*, 265 from *R. irroratum*, 527 from *R. ovatum*, and 211 from *R. simsii*. The NBS-encoding genes presented prominent expansion in *Rhododendron* genomes, compared with kiwifruit (102 genes). According to their N-terminal domains, these obtained genes were first classified into three subclasses: CNL, TNL and RNL. For each subclass, the genes were further categorized into different types based on their predicted protein domain compositions (Figure 5A). In addition, sequences that lacked the TIR, CC, or RPW8 domains were subjected to BLASTn analyses with the classified genes to determine their subclass classification based on sequence similarity. Cluster assignment suggested the majority of NBS-encoding genes were located in clusters on *Rhododendron* chromosomes and only a small number of genes existed as singletons (Supplementary Table 18). The domain compositions and cluster assignment of the 1,336 genes were shown in Supplementary Table 19.

An overall phylogeny was constructed for NBS-encoding genes from four *Rhododendron* species as well as kiwifruit, which revealed that the RNL clade was embedded in the TNL clade, and CNL formed an independent clade (Figure 5B). This phylogenetic tree was inconsistent with the topology in previous studies, in which the three subclasses (TNL, CNL and RNL) each formed a monophyletic clade (Shao et al., 2016). Here, two phylogenies for NBS-encoding genes (CNL and nCNL) were reconstructed and reconciled with the real species tree to restore ancestral gene duplication or loss events (Supplementary Figure S10, S11). According to the definitions, 63 ancient NBS-encoding gene lineages (55 CNL, 8 nCNL) were recovered in the common ancestor of kiwifruit and four *Rhododendron* species (Supplementary Tables 20, 21). Of these ancient gene lineages, 46 lineages (38 CNL, 8 nCNL) were inherited by the common ancestor of four *Rhododendron* species and further diverged into 493 gene lineages (411 CNL, 82 nCNL), which were defined as *Rhododendron* NBS-encoding gene lineages. Since divergence from kiwifruit, the 493 ancestral *Rhododendron* NBS-encoding gene lineages underwent different gene duplication/loss

10.3389/fpls.2023.1123707

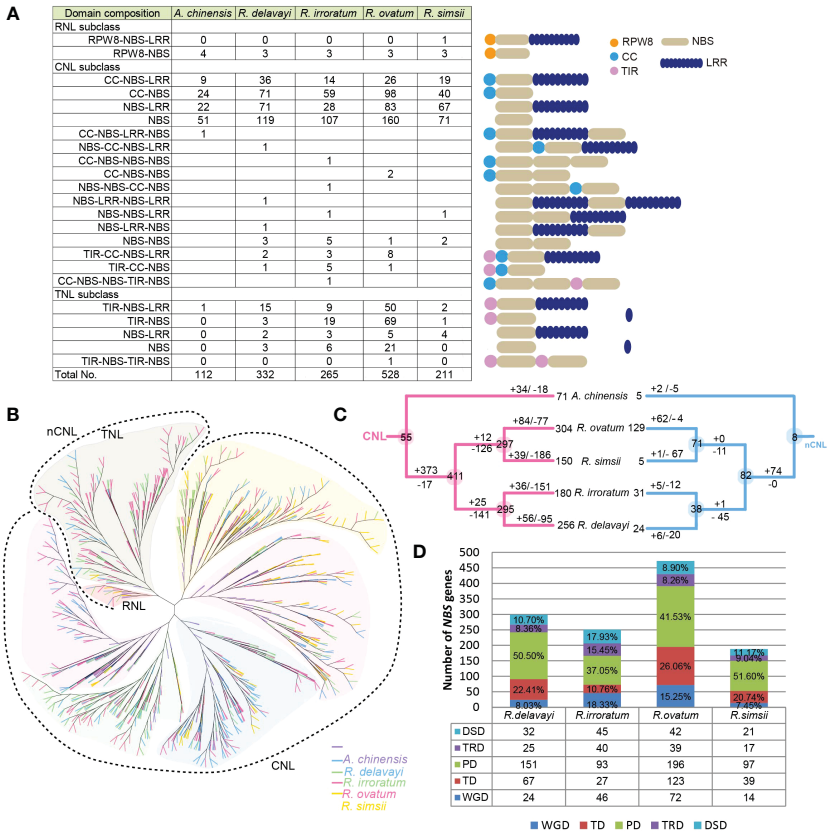

FIGURE 5
NBS-encoding gene evolution in *Rhododendron*. **(A)** Number of the NBS-encoding genes in each category in four *Rhododendron* species and kiwifruit. Twenty-three domain combinations were found. The numbers listed in table represented the number of genes for each domain combination. 'R', 'C', 'T' and 'N' were the abbreviations of RPW8, CC, TIR and NBS domains. Genes with both CC and TIR domains were clustered into CNL clade based on phylogenetic analysis. **(B)** The NBS-encoding gene phylogenetic tree for four *Rhododendron* species and kiwifruit; **(C)** Duplications/losses of the NBS-encoding genes during the evolutionary processes of *Rhododendron*. The number of ancestral genes for each node was shown in circles. Gene duplications or losses on each branch were illustrated by numbers with + or −symbols, respectively; **(D)** Numbers of NBS-encoding genes produced by different gene duplication events in four *Rhododendron* genomes.

events during the speciation of four *Rhododendron* species. It is found that 141 CNL lineages and 45 nCNL lineages were lost in the common ancestor of *R. delavayi* and *R. irroratum*, while 25 CNL and 1 nCNL lineages were gained, resulting in 295 CNL and and 38 nCNL lineages (Figure 5C). Then, the gene duplications occurred less frequently than gene losses in both *R. delavayi* and *R. irroratum*, causing a more reduced number of genes in the two species. Meanwhile, 126 CNL and 11 nCNL lineages were lost, while 12 CNL and 0 nCNL lineages were gained in the common ancestor of *R. simsii* and *R. ovatum*, resulting in 297 CNL genes and 71 nCNL genes (Figure 5C. However, *R. simsii* and *R. ovatum* underwent considerably different evolutionary processes after divergence. Specifically, *R. ovatum* duplicated 146 genes (84 CNL, 62 nCNL), and lost 82 genes (77 CNL, 4 nCNL), resulting in an increased number of NBS-encoding genes, especially for nCNL subclass. *R. simsii* duplicated 40 genes (39 CNL, 1 nCNL) and lost 253 genes (186 CNL, 67 nCNL), resulting in a decreased gene number in this genome. In general, the NBS-encoding genes appeared to experience abundant duplications in the common ancestor of four *Rhododendron* species. Then, more frequent gene losses than gains occurred during the speciation of *R. delavayi*, *R. irroratum*, and *R. simsii*. However, *R. ovatum* was the opposite.

The NBS gene duplications revealed that PD and TD events were responsible for larger proportion of the NBS-encoding genes in *R. delavayi* (73.91%), *R. ovatum* (67.58%) and *R. simsii* (72.34) genomes, whereas PD and WGD events produced more than of the genes (55.38%) in *R. irroratum* (Figure 5D). In particular, the PD-derived NBS-encoding genes occupied the largest proportion across all studied *Rhododendron* species (Figure 5D).

Discussion

In this study, the chromosome-level genome assemblies were generated for two alpine rhododendrons, *R. delavayi* and *R. irroratum*, which can be used as the parent to produce hybrid varieties for ornamental or landscape use. The newly assembled *R. delavayi* genome had higher accuracy and continuity than the

previously reported version (Zhang et al., 2017). In view of the relatively high heterozygosity (1.55%) of *R. irroratum* genome, Purge Haplotigs program were used to remove duplicated haploid contigs from the contig-level assembly. Through Hi-C scaffolding and manual correction, the genome sequences of *R. delavayi* and *R. irroratum* we can assemble into 13 pseudo-chromosomes with contigs/scaffolds accurately assigned, in accordance with the chromosome number mentioned by previous study (Zhang et al., 2007). Phylogenetic analysis placed five *Rhododendron* species into two sister clades, with that comprising *R. delavayi*, *R. irroratum* and *R. williamsianum* referred to as 'rhododendrons' and the other clade, containing *R. simsii* and *R. ovatum*, commonly considered as the 'azaleas' (De Riek et al., 2018). We dated a WGD event (Ad-β) in *Rhododendron* at around 78 Mya, prior to the divergence of *Rhododendron* and *Actinidia*. Although the *Rhododendron* genus contains the most diverse species of woody plants, there was no evidence for the occurrence of a recent WGD in the common ancestor of *Rhododendron* (Ericaceae). A study based on phylogenomic and ecological analyses shed light on the evolutionary radiations and species diversity of *Rhododendron* are driven by geographic, climatic factors and functional traits of leaves (Xia et al., 2022).

Gene duplications can provide important raw materials for evolutionary novelty and the duplicated genes were often biased retained for different modes of duplication (Freeling, 2009). Our analysis revealed that the duplication-involved expanded gene families were enriched in different KEGG pathways. TD/PD duplications substantially contributed to the expansion of gene families related to secondary metabolism and plant-pathogen interaction, while genes involved in basic biological process tended to be expanded by WGD event. The gene retentions after WGD versus TD/PD tend to be biased and might exhibit reciprocal relationship, which can be explained by balanced gene drive (Freeling, 2009). Different from whole-genome duplication, the tandem or proximal duplicates can emerge continuously (Qiao et al., 2019), to some extent, which can increase gene dosage efficiently (Conant and Wolfe, 2008) and play significant role in rapid environmental response (such as pathogen invasion) or rate-limiting steps in secondary metabolism.

The R2R3-MYB proteins have been proven to control various biological processes in plants, e.g. secondary metabolism, development and stresses responses (Dubos et al., 2010). Here, the R2R3-MYB members were comprehensively identified and characterized in four *Rhododendron* species. The number of *R2R3-MYB* genes in *R. ovatum* and *R. simsii* is more than that in *R. delavayi* and *R. irroratum*, which might be attributed to frequent gene losses, pseudogenization or failure of gene annotations. It was noteworthy that these *Rhododendron* genomes were annotated using different strategies, and there must be incomplete or error annotation to some extent. Thus, it is necessary to annotate these *Rhododendron* genomes in a uniform way. There were only 41 one-to-one syntenic orthologs found in the four *Rhododendron* species. The remaining genes either maintain collinearity relationships in two or three *Rhododendron* genomes or derived from other duplicated modes, such as tandem duplication, transposed duplication and so on. In addition, genes in the syntenic blocks might also experience loss or translocation.

Flower color is a critical ornamental characteristic of *Rhododendron* and is mainly determined by anthocyanins and carotenoids (Yang et al., 2020). Studies have demonstrated that the R2R3-MYB transcription factors are key regulators in the biosynthesis of anthocyanin and carotenoid in plants (Schwinn et al., 2006; Albert et al., 2014; Sagawa et al., 2016; Ampomah-Dwamena et al., 2019). Since the R2R3-MYB members within subgroups are conservative and functionally correlated, 15 *Rhododendron R2R3-MYB* genes clustered together with *Arabidopsis AtMYB75/AtMYB90/AtMYB113/AtMYB114* genes were supposed to be related to anthocyanin biosynthesis. These genes could be important candidates to further investigate their roles in flower coloration of *Rhododendron*. Moreover, we identified 16 *Rhododendron R2R3-MYB* genes clustered together with Arabidopsis MYB genes *AtMYB60* and AtMYB96, which might participate in drought stress response. Gene duplication is critical in gene family expansion and evolution. Here, the contributions of different gene duplications to *R2R3-MYB* gene expansions were dissected and the results revealed that WGD event played predominant roles in generating the *R2R3-MYB* genes for four investigated *Rhododendron* species, agreeing with previous findings that genes encoding transcription factors are preferentially retained after polyploidization (Tian et al., 2005).

Plants are threatened by various pathogens in their natural habitats. We found several gene families participating in biotic stimuli and defense responses such as "plant-pathogen interaction" and "immune response-activating signal transduction" were significantly enriched. It should be noted that NBS-encoding genes play crucial roles in plant disease resistance and immune response. Therefore, we identified this type of R genes and found they presented obvious expansion in *Rhododendron* genus compared with kiwifruit, suggesting the wild *Rhododendron* species may have strong resistance to disease erosion. *R. ovatum* had more identified NBS-encoding genes than other three *Rhododendron* species, perhaps because it is distributed in the subtropical zone at lower altitudes and in hotter environments where the plant may be more exposed to microorganisms. There were 493 ancestral *Rhododendron* NBS-encoding gene lineages, from which the genes detected currently in *Rhododendron* species should be derived. After divergence from common ancestor, the individual *Rhododendron* species experienced different numbers of gene gain/loss: *R. delavayi*, *R. irroratum* and *R. simsii* exhibited a "continuous contraction" pattern (for both CNLs and nCNLs), because these species lost more genes than they gained at two periods (before and after the split of the common ancestors of *R. delavayi*, and *R. irroratum*, and of *R. simsii* and *R. ovatum*) and *R. ovatum* showed a "contraction followed by expansion" pattern (for both CNLs and nCNLs) by losing more genes before the split with *R. simsii* but subsequently gaining more genes. Therefore, the different gene gains and losses resulted in different number of NBS-encoding genes in the current *Rhododendron* genomes. Consistent with the above findings that TD/PD duplications play critical roles in the expansion of gene families involved in plant-pathogen interaction, the TD/PD events largely contributed to the expansion of NBS-encoding genes, suggesting these duplicates are important for plant response to pathogenic organisms.

In summary, the two newly available *Rhododendron* genome sequences add to the increasing available genome information for *Rhododendron* and will facilitate the study on genome evolution of Ericaceae. The candidate *R2R3-MYB* genes predicted to be involved in anthocyanin biosynthesis and stress response will provide potential targets for further functional characterization, which can guide breeders to develop new ornamental varieties and enhance stress response of *Rhododendron* cultivars. Moreover, the comprehensive identification and evolutionary analysis of NBS-encoding genes could provide a reference for further investigation of this gene family in *Rhododendron* species.

Data availability statement

The raw sequence reads of R. delavayi/R. irroratum have been deposited in the NCBI under BioProject number PRJNA929713/PRJNA931779 and in the CNCB under BioProject number PRJCA011631/PRJCA011815, respectively. Genome assembly and annotation files are available at the Rhododendron Plant Genome Database (http://bioinfor.kib.ac.cn/RPGD/download_genome.html).

Author contributions

XWu, LZ, and RZ conducted the data analysis. XW, GJ and ZW assisted in the data analysis. HY and YH prepared the samples. JW, YM, CZ and XWu. wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was financially supported by Yunnan Province Agriculture joint special project (202101BD070001-010), the Science and Technology Talent and Platform Program of Yunnan Province (202105AC160017), the Construction of International Flower Technology Innovation Center and Industrialization of Achievements (2019ZG006), the Yunnan Young & Elite Talents Project (YNWR-QNBJ-2018-268), the ZIAT-Special Project (ZJKJT-2022-01) and Major Science and Technology Projects of Yunnan Province of China (2018BB014).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1123707/full#supplementary-material

References

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

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Ameline-Torregrosa, C., Wang, B., O'Bleness, M. S., Deshpande, S., Zhu, H., Roe, B., et al. (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant Physiol.* 146, 5–21. doi: 10.1104/pp.107.104588

Ampomah-Dwamena, C., Thrimawithana, A. H., Dejnoprat, S., Lewis, D., Espley, R. V., and Allan, A. C. (2019). A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytol.* 221, 309–325. doi: 10.1111/nph.15362

Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, W369–W373. doi: 10.1093/nar/gkl198

Bent, A. F., and Mackey, D. (2007). Elicitors, effectors, and r genes: the new paradigm and a lifetime supply of questions. *Annu. Rev. Phytopathol.* 45, 399–436. doi: 10.1146/annurev.phyto.45.062806.094427

Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinf.* 15, 1–9. doi: 10.1186/1471-2105-15-211

Browse, J. (2009). Jasmonate passes muster: a receptor and targets for the defense hormone. *Annu. Rev. Plant Biol.* 60, 183–205. doi: 10.1146/annurev.arplant.043008.092007

Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., and Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with arabidopsis. *BMC Genomics* 7, 327. doi: 10.1186/1471-2164-7-327

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

Chamberlain, D. F. (1982). A revision of *Rhododendron* II: Subgenus *Hymenanthes*. *Notes R. Botanic Garden Edinburgh* 39, 209–486.

Chamberlain, D., Hyam, R., Argent, G., Fairweather, G., and Walter, K. S. (1996). The genus *Rhododendron*: its classification and synonymy. *R. Botanic Garden Edinburgh*.

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

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, 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

Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi: 10.1038/nmeth.4035

Collinson, M. E., and Crane, P. R. (1978). *Rhododendron* seeds from the Palaeocene of southern England. *Bot. J. Linn. Soc* 76, 195–205. doi: 10.1111/j.1095-8339.1978.tb02361.x

Cominelli, E., Galbiati, M., Vavasseur, A., Conti, L., Sala, T., Vuylsteke, M., et al. (2005). A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Curr. Biol.* 15, 1196–1200. doi: 10.1016/j.cub.2005.05.048

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161

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

De Riek, J., De Keyser, E., Calsyn, E., Eeckhaut, T., Van Huylenbroeck, J., and Kobayashi, N. (2018). "Azalea," in *Ornamental crops*. Ed. J. Van Huylenbroeck (Cham: Springer), 237–271.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in arabidopsis. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005

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

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016a). Juicebox provides a visualization system for Hi-c contact maps with unlimited zoom. *Cell Syst.* 3, 99. doi: 10.1016/j.cels.2015.07.012

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

Fang, R., and Min, T. (1995). The floristic study on the genus *Rhododendron*. *Acta Botanica Yunnanica* 17, 359–379.

Flagel, L. E., and Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183, 557–564. doi: 10.1111/j.1469-8137.2009.02923.x

Freeling, M. (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60, 433–453. doi: 10.1146/annurev.arplant.043008.092122

Gonzalez, A., Zhao, M., Leavitt, J. M., and Lloyd, A. M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in arabidopsis seedlings. *Plant J.* 53, 814–827. doi: 10.1111/j.1365-313X.2007.03373.x

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

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7

Hoff, K. J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods Mol. Biol.* 1962, 65–95. doi: 10.1007/978-1-4939-9173-0_5

Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinf.* 12, 491. doi: 10.1186/1471-2105-12-491

Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36, 2253–2255. doi: 10.1093/bioinformatics/btz891

Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., et al. (2013). Draft genome of the kiwifruit *Actinidia chinensis*. *Nat. Commun.* 4, 2640. doi: 10.1038/ncomms3640

Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286

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

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf.* 19, 189. doi: 10.1186/s12859-018-2203-5

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* 44, e89–e89. doi: 10.1093/nar/gkw092

Kim, Y. S., An, C., Park, S., Gilmour, S. J., Wang, L., Renna, L., et al. (2017). CAMTA-mediated regulation of salicylic acid immunity pathway genes in *Arabidopsis* exposed to low temperature and pathogen infection. *Plant Cell* 29, 2465–2477. doi: 10.1105/tpc.16.00865

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly *via* adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lee, S. B., and Suh, M. C. (2015). Advances in the understanding of cuticular waxes in *Arabidopsis thaliana* and crop species. *Plant Cell Rep.* 34, 557–572. doi: 10.1007/s00299-015-1772-2

Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Yi, T., Gao, L., Ma, P., Zhang, T., Yang, J., et al. (2019). Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants* 5, 461–470. doi: 10.1038/s41477-019-0421-0

Liu, H., Wu, S., Li, A., and Ruan, J. (2020a). SMARTdenovo: A *de novo* assembler using long noisy reads. *Preprints*. doi: 10.20944/preprints202009.0207.v1

Lupas, A., Dyke, M. V., and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* 252, 1162–1164. doi: 10.1126/science.252.5009.1162

Ma, H., Liu, Y., Liu, D., Sun, W., Liu, X., Wan, Y., et al. (2021). Chromosome-level genome assembly and population genetic analysis of a critically endangered rhododendron provide insights into its conservation. *Plant J.* 107, 1533–1545. doi: 10.1111/tpj.15399

Marcais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of K-mers. *Bioinformatics* 27, 764–777. doi: 10.1093/bioinformatics/btr011

McDowell, J. M., and Simon, S. A. (2006). Recent insights into r gene evolution. *Mol. Plant Pathol.* 7, 437–448. doi: 10.1111/j.1364-3703.2006.00342.x

Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20, 317–332. doi: 10.1046/j.1365-313x.1999.t01-1-00606.x

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. (2003). Genome-wide analysis of *NBS-LRR*-encoding genes in arabidopsis. *Plant Cell* 15, 809–834. doi: 10.1105/tpc.009308

Nakabayashi, R., and Saito, K. (2015). Integrated metabolomics for abiotic stress responses in plants. *Curr. Opin. Plant Biol.* 24, 10–16. doi: 10.1016/j.pbi.2015.01.003

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

Nixon, K. C., and Crepet, W. L. (1993). Late Cretaceous fossil flowers of ericalean affinity. *Am. J. Bot.* 80, 616–623. doi: 10.1002/j.1537-2197.1993.tb15230.x

Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Res.* 46, e126. doi: 10.1093/nar/gky730

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256. doi: 10.1093/molbev/msn083

Pryszcz, L. P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44, e113–e113. doi: 10.1093/nar/gkw294

Pu, X., Dong, X., Li, Q., Chen, Z., and Liu, L. (2021). An update on the function and regulation of MEP and MVA pathways and their evolutionary dynamics. *J. Integr. Plant Biol.* 63, 1211–1226. doi: 10.1111/jipb.13076

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* 20, 38. doi: 10.1186/s13059-019-1650-2

Roach, M. J., Schmidt, S. A., and Borneman, A. R. (2018). Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinf.* 19, 460. doi: 10.1186/s12859-018-2485-7

Sagawa, J. M., Stanley, L. E., LaFountain, A. M., Frank, H. A., Liu, C., and Yuan, Y. (2016). An R2R3-MYB transcription factor regulates carotenoid pigmentation in *Mimulus lewisii* flowers. *New Phytol.* 209, 1049–1057. doi: 10.1111/nph.13647

Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., et al. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* 18, 831–851. doi: 10.1105/tpc.105.039255

Seo, P. J., and Park, C. M. (2010). MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in arabidopsis. *New Phytol.* 186, 471–483. doi: 10.1111/j.1469-8137.2010.03183.x

Seo, P. J., Xiang, F., Qiao, M., Park, J. Y., Lee, Y. N., Kim, S. G., et al. (2009). The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in arabidopsis. *Plant Physiol.* 151, 275–289. doi: 10.1104/pp.109.144220

Shao, Z., Xue, J., Wang, Q., Wang, B., and Chen, J. (2019). Revisiting the origin of plant NBS-LRR genes. *Trends Plant Sci.* 24, 9–12. doi: 10.1016/j.tplants.2018.10.015

Shao, Z., Xue, J., Wu, P., Zhang, Y., Wu, Y., Hang, Y., et al. (2016). Large-Scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol.* 170, 2095–2109. doi: 10.1104/pp.15.01487

Soza, V. L., Lindsley, D., Waalkes, A., Ramage, E., Patwardhan, R. P., Burton, J. N., et al. (2019). The *Rhododendron* genome and chromosomal organization provide insight into shared whole-genome duplications across the heath family (Ericaceae). *Genome Biol. Evol.* 11, 3353–3371. doi: 10.1093/gbe/evz245

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28, i409–i415. doi: 10.1093/bioinformatics/bts386

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–12. doi: 10.1093/nar/gkl315

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

The Angiosperm Phylogeny Group, Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., et al. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc* 181, 1–20. doi: 10.1111/boj.12385

Teh, B. T., Lim, K., Yong, C. H., Ng, C. C.Y., Rao, S. R., Rajasegaran, V., et al. (2017). The draft genome of tropical fruit durian (*Durio zibethinus*). *Nat. Genet.* 49, 1633–1641. doi: 10.1038/ng.3972

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Tian, X., Chang, Y., Neilsen, J., Wang, S., and Ma, Y. (2019). A new species of *Rhododendron* (Ericaceae) from northeastern yunnan, China. *Phytotaxa* 395, 7. doi: 10.11646/phytotaxa.395.2.2

Tian, C., Xiong, Y., Liu, T., Sun, S., Chan, L., and Chen, M. (2005). Evidence for an ancient whole-genome duplication event in rice and other cereals. *Yi Chuan Xue Bao* 32, 519–527.

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., et al. (2010). Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* 6, 1869. doi: 10.3791/1869

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33, 2202–2204. doi: 10.1093/bioinformatics/btx153

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9, e112963. doi: 10.1371/journal.pone.0112963

Wang, X., Gao, Y., Wu, X., Wen, X., Li, D., Zhou, H., et al. (2021). High-quality evergreen azalea genome reveals tandem duplication-facilitated low-altitude adaptability and floral scent evolution. *Plant Biotechnol. J.* 19, 2544–2560. doi: 10.1111/pbi.13680

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293

Wang, J., Yu, J., Li, J., Sun, P., Wang, L., Yuan, J., et al. (2018). Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the actinidiaceae family. *iScience* 7, 230–240. doi: 10.1016/j.isci.2018.08.003

Wang, D. P., Zhang, Y. B., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genom. Proteom. Bioinf.* 8, 77–80. doi: 10.1016/S1672-0229(10)60008-3

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

Wu, H., Ma, T., Kang, M., Ai, F., Zhang, J., Dong, G., et al. (2019). A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic. Res.* 6, 117. doi: 10.1038/s41438-019-0202-y

Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. doi: 10.1093/bioinformatics/bti310

Xia, X., Yang, M., Li, C., Huang, S., Jin, W., Shen, T., et al. (2022). Spatiotemporal evolution of the global species diversity of *Rhododendron*. *Mol. Biol. Evol.* 7,39, msab314. doi: 10.1093/molbev/msab314

Xiao, S., Ellwood, S., Calis, O., Patrick, E., Li, T., Coleman, M., et al. (2001). Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by *RPW8*. *Science* 291, 118–120. doi: 10.1126/science.291.5501.118

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

Xu, G., Xu, T., Zhu, R., Zhang, Y., Li, S., Wang, H., et al. (2019). LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *Gigascience* 8, giy157. doi: 10.1093/gigascience/giy157

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13, 555–556. doi: 10.1093/bioinformatics/13.5.555

Yang, F., Nie, S., Liu, H., Shi, T., Tian, X., Zhou, S., et al. (2020). Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat. Commun.* 11, 5269. doi: 10.1038/s41467-020-18771-4

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X., et al. (2012). ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* 419, 779–781. doi: 10.1016/j.bbrc.2012.02.101

Zhang, L., Xu, P., Cai, Y., Ma, L., Li, S., Li, S., et al. (2017). The draft genome assembly of *Rhododendron delavayi* franch. var. *delavayi*. *Gigascience* 6, 1–11. doi: 10.1093/gigascience/gix076

Zhang, J., Zhang, C., Gao, L., Yang, J., and Li, H. (2007). Natural hybridization origin of *Rhododendron agastum* (Ericaceae) in yunnan, China: inferred from morphological and molecular evidence. *J. Plant Res.* 120, 457–463. doi: 10.1007/s10265-007-0076-1

Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-c data. *Nat. Plants* 5, 833–845. doi: 10.1038/s41477-019-0487-8

Zhou, X., Li, J., Wang, H., Han, J., Zhang, K., Dong, S., et al. (2022). The chromosome-scale genome assembly, annotation and evolution of *Rhododendron henanense* subsp. *lingbaoense*. *Mol. Ecol. Resour.* 22, 988–1001. doi: 10.1111/1755-0998.13529