Genomics 115 (2023) 110617

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Insights into identifying resistance genes for cold and disease stresses through chromosome-level reference genome analyses of *Poncirus polyandra*

Shanshan Zhang [a], Jian Chen [a], Chuanguang Zhang [a], Siqi Zhang [a], Xiaolei Zhang [b], Lanjing Gao [c], Wenzhong Yang [a,*]

[a] Key Laboratory of Rare and Endangered Forest Plants of State Forestry and Grassland Administration, Yunnan Academy of Forestry and Grassland, Kunming 650201, China
[b] Kunming Endangered Animal and Plant Asylum Save Center, Kunming 650224, China
[c] College of Forestry, Beijing Forestry University, Beijing 100083, China

ARTICLE INFO

Keywords:
Poncirus
Chromosome-level genome
Resistance genes
Evolutionary and comparative genomics

ABSTRACT

Poncirus polyandra, a plant species with extremely small populations in China, has become extinct in the wild. This study aimed to identify functional genes that improve tolerance to abiotic and biotic stresses. Here, we present a high-quality chromosome-scale reference genome of *P. polyandra*. The reference genome is 315.78 Mb in size, with an N50 scaffold size of 32.07 Mb, and contains nine chromosomes with 20,815 protein-coding genes, covering 97.82% of the estimated gene space. We identified 17 rapidly evolving nucleotide-binding-site (NBS) genes, three C-repeat-binding factors (CBF) genes, 19 citrus greening disease (Huanglongbing, HLB) tolerance genes, 11 citrus tristeza virus (CTV) genes, and one citrus nematode resistance gene. A divergence time of 1.96 million years ago was estimated between *P. polyandra* and *P. trifoliata*. This is the first genome-scale assembly and annotation of *P. polyandra*, which will be useful for genetic, genomic, and molecular research and provide guidance for the development of conservation strategies.

1. Introduction

Poncirus Raf., also known as *Citrus*, belongs to the citrus subfamily *Rutaceae* and comprises two species, *Poncirus polyandra* and *P. trifoliata* [19,23]. Originated in China, *Poncirus* differs from Citrus, Kumquats, and other genera [29,35]. It is mainly distributed in the provinces on both sides of the middle reaches of the Yangtze River and Huaihe River Basin, which is wider than that of the *Citrus* species. It was recorded in Li Shizhen's Compendium of Materia Medica that the *Poncirus* genus had been used in traditional Chinese medicine for an extended period of time [38,39,69]. Plants of the *Poncirus* genus have been widely utilized as germplasm resources for rootstocks in citrus production owing to several advantageous features, including efficient absorption and utilization of nutrients, cold hardiness, and resistance to several major diseases and pests [4,21,31,37].

P. polyandra ($2n = 2\times = 18$), a new species published by Ding et al. (1984) [19], belongs to the *Poncirus* genus of the *Rutaceae* family [59]. It is a key national protected species under grade II protection in China, and it is endemic to Fumin County, Yunnan Province (http://rep.iplant.cn/protlist). It is sexually compatible with *Citrus* species but distinct from *Citrus* and *P. polyandra* in many characteristics, including flower size, petals, and stamen number [22,30,87,98]. It is evergreen (*versus* deciduous in *P. polyandra*), has three leaflets on each leaf, has large thorns on the shoots, and produces bitter, inedible fruits [88,100]. *P. polyandra* has potential as a rootstock material for *Citrus* species and may provide useful information on the nature of the ancient divergence of *Citrus* and *Poncirus*. It is a valuable medicinal resource that inhibits angiogenesis and locally reduces the expression of tumor cells [40]. Owing to the influence of human disturbance, overgrazing of the original land, and massive excavation as an evergreen ornamental plant, *P. polyandra* has become extinct in the wild [27,51,89]. It is listed as a plant species with extremely small populations endemic to Yunnan Province; therefore, it is necessary to protect *P. polyandra* germplasm resources while still in existence.

The C-repeat-binding factor (CBF) pathway [10,58,73,82], nucleotide-binding site (NBS)-leucine-rich repeat (LRR) genes [4,18], and transcription factors (TFs) [14,28,42,63] are known to play very important roles in the response to stresses. *Poncirus trifoliata*, the only

* Corresponding author at: Yunnan Academy of Forestry and Grassland, Kunming 650201, China.
E-mail addresses: zhangshanshan@yafg.ac.cn (S. Zhang), chenjian@yafg.ac.cn (J. Chen), zhangchuanguang@yafg.ac.cn (C. Zhang), yangwenzhong@yafg.ac.cn (W. Yang).

https://doi.org/10.1016/j.ygeno.2023.110617
Received 22 December 2022; Received in revised form 22 March 2023; Accepted 27 March 2023
Available online 30 March 2023

other species of *Poncirus*, has been widely used as a rootstock for citrus production because of its resistance genes associated with biotic [41,68] and abiotic stresses [20,64,83,85]. For example, many candidate genes involved in disease tolerance and cold stress have been well characterized, including CBF genes, NBS genes, TFs, and other disease resistance genes in the *P. trifoliata* genome [68,81]. Thus, we hypothesized that *P. polyandra*, the species most closely related to *P. trifoliata*, may also have favorable traits that confer resistance to several diseases and cold.

The objective of this study was to discover genes with functional properties that enhance the capacity to withstand abiotic and biotic stresses. To explore the striking features of *P. polyandra*, we performed genome sequencing and *de novo* assembly to obtain a high-quality chromosome-scale reference genome for this important species. We then investigated gene family evolution and positive selection in *P. polyandra* by performing a comparative genomic analysis of the *P. trifoliata* genome and other publicly available *Citrus* genomes. Furthermore, we identified candidate genes in *P. polyandra* using previously identified genes for cold-signaling-related Huanglongbing (HLB), nematodes, and citrus tristeza virus (CTV) resistance. High-quality genome assemblies can serve as valuable resources for genetic, genomic, and molecular research.

2. Materials and methods

2.1. Plant materials

Samples of *P. polyandra* were collected from Fumin County, Kunming City, Yunnan Province, China. The young leaf tissues were immediately placed in liquid nitrogen.

2.2. Genomic DNA extraction and sequencing

Genomic DNA was isolated from the leaves of each plant using a Plant Genomic DNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions [12]. DNA samples were purified using a Zymo Genomic DNA Clean Kit (Zymo Research, Irvine, CA, USA). The purified DNA was detected using a NanoDrop™ One UV–Vis spectrophotometer (Thermo Fisher Scientific, USA). The DNA concentration was further measured by Qubit® 3.0 Fluorometer (Invitrogen, USA).

A total amount of 2 μg DNA per sample was used as input material for the ONT (Oxford Nanopore Technologies, UK) library preparations. After the sample was qualified, the size selection of long DNA fragments was performed using the BluePippin system (Sage Science, USA). Next, the ends of the DNA fragments were repaired, and the A-ligation reaction was conducted with the NEBNext Ultra II End Repair/dA-tailing Kit (New England Biolabs, UK). The adapter in the SQK-LSK109 Kit (Nanopore, UK) was used for further ligation reaction, and a Qubit® 3.0 Fluorometer (Invitrogen, USA) was used to quantify the size of library fragments. The genome of *P. polyandra* was sequenced using an Oxford Nanopore PromethION at Nextomics Biosciences (Wuhan, China). In addition, a total of 8 μg genomic DNA was randomly fragmented by g-TUBE devices (Covaris, MA, United States). Then a separate paired-end DNA library with an insert size of 150 bp was constructed and sequenced on MGISEQ-T7 platform.

2.3. Data quality control

Nanopore sequencers output FAST5 files containing signal data, and base calling was performed to convert the FAST5 files to FASTQ format using Guppy v3.2.2 [86]. Raw reads in fastq format with mean_qscore_template <7 were then filtered, resulting in pass reads [8]. Unpaired, short, and low-quality reads were removed, and clean reads were used.

2.4. Genome assembly

For *de novo* genome assembly, an ONT assembly was constructed with NextDenovo v2.3.1 (https://github.com/Nextomics/NextDenovo) [56,65]. Considering the high error rate of ONT raw reads, the original subreads were first self-corrected using NextCorrect to obtain consensus sequence (CNS) reads. The CNS reads were then compared with the NextGraph module to capture their correlations. A preliminary genome was assembled based on the correlation with CNS reads. To improve the accuracy of the assembly, the contigs were refined and revised with Racon v1.3.1 using ONT long reads and NextPolish v1.3.0 using MGISEQ-T7 short reads with default parameters [36]. To discard possibly redundant contigs and generate a final assembly, similarity searches were performed with the parameters "identity 0.8–overlap 0.8."

The completeness of the genome assembly was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO v4.0.5) [74] and Core Eukaryotic Gene Mapping Approach (CEGMA v2) [67]. To evaluate the accuracy of the assembly, all MGISEQ-T7 paired-end reads were mapped to the assembled genome using the Burrows-Wheeler Aligner (BWA v0.7.12) [52], and the mapping rate and genome coverage of the sequencing reads were assessed using SAMtools v1.4 [53]. In addition, the base accuracy of the assembly was calculated using bcftools v1.8.0 [15].

2.5. Chromosome assembly using Hi-C data

To anchor the hybrid scaffolds to the chromosome, genomic DNA was extracted for the Hi-C library of *P. polyandra* [57]. We then constructed a Hi-C library and obtained sequencing data using MGISEQ-T7 platform.

In total, 382 million paired-end reads were generated from these libraries. Quality control of the Hi-C raw data was performed using HiC-Pro v2.8.1 (HiC-Pro, RRID:SCR 017643) [72]. First, low-quality sequences (quality scores <20), adaptor sequences, and sequences shorter than 30 bp were filtered out using fastp v0.19.4 [11], and then the clean paired-end reads were mapped to the draft assembled sequence using bowtie2 v2.2.5 (Bowtie, RRID: SCR 005476) [50] to obtain unique mapped paired-end reads. Valid interaction paired reads were identified and retained from uniquely mapped paired-end reads using HiC-Pro (v2.8.1) [72] for further analysis. Invalid read pairs, including dangling ends, self-cycles, relegation, and dumped products, were filtered using HiC-Pro (v2.8.1). The scaffolds were further clustered, ordered, and oriented onto chromosomes using LACHESIS (https://github.com/shendurelab/LACHESIS) [6] with the recommended parameters. Finally, placement and orientation errors exhibiting obvious discrete chromatin interaction patterns were manually adjusted.

2.6. Genome annotation

To identify known transposable elements (TEs) in the *P. polyandra* genome, RepeatMasker v1.331 [78] was used to screen the assembled genome against the Rep-base (v22.11) [3] and Mips-REdat libraries [66]. In addition, *de novo* repeat annotation was performed using RepeatModeler v1.0.11 [78].

Homology-based non-coding RNA annotation was performed by mapping plant ribosomal RNA (rRNA), microRNA, and small nuclear RNA genes from the Rfam database (release 13.0) [43] to the *P. polyandra* genome using BLASTN (BLASTN, RRID:SCR 001598) (*E*-value ≤1e–5) [9]. tRNAscan-SE v2.0 [60] was used to annotate the transfer RNAs (tRNAs) using the default parameters for eukaryote annotation. RNAmmer v1.2 [49] was used to predict rRNAs and their subunits. To annotate genes in the *P. polyandra* genome, gene prediction was performed using homology-based, *de novo*, and transcriptome sequencing-based methods. Finally, EVidenceModeler v1.1.1 [33] was used to integrate the predicted genes and generate a consensus gene set.

Genes with TEs were discarded using the TransposonPSI package [79]. Low-quality genes consisting of <50 amino acids and exhibiting premature termination (by aligning codons 1 by 1, with fragments with termination codons in the middle) were also removed from the gene set.

Functional annotations of the predicted genes were mainly based on homology to known annotated genes within different databases using Basic Local Alignment Search Tool (BLAST v2.9). To achieve their corresponding annotations, we aligned protein models in SwissProt, TrEMBL, the National Center for Biotechnology Information (NCBI) nonredundant (NR) database, InterPro, Gene Ontology (GO) [1], NCBI Clusters of Orthologous Groups of Proteins (COG) [26], and Kyoto Encyclopedia of Genes and Genomes (KEGG) [45]. The completeness of the predicted gene set was evaluated in the BUSCO OrthoDB v10.0 embryophyta dataset [74].

2.7. TF gene analysis

To identify the genes encoding TFs in *P. polyandra*, the protein sequences of the predicted gene models were searched against the Plant Transcription Factor Database v5.0 (http://planttfdb.cbi. pku.edu.cn) [68].

2.8. Analysis of NBS-LRR genes

To identify NBS-containing genes in the *P. polyandra* genome, the protein sequences were first scanned using the NBS Hidden Markov Model (HMM) profile (PF00931) and 'hmmsearch' in Hmmer under an *E*-value of 1e–03 [24]. In addition, a *P. polyandra* specific NBS domain HMM profile was made by using the high-quality hits from 'hmmsearch' (E-value 1e–05) and 'hmmbuild,' which was also used to scan the protein sequences. The presence of the NBS domain was confirmed using PfamScan [25]. The toll/interleukin-1 receptor (TIR) and LRR domains were identified using the NCBI conserved domain tool. The coiled coil (CC) domain was identified using Marcoil's test (probability >90%) [17].

2.9. Identification of candidate genes associated with disease resistance and cold signaling pathways

Candidate disease resistance genes related to HLB, CTV, and citrus nematodes were retrieved according to the annotation file of the whole genome sequencing results of *P. polyandra*, and candidate genes related to cold signals were retrieved using the same method [68]. TBtools software was used to determine the distribution of these genes on the chromosomes and to cross them with the NBS gene to obtain candidate HLB, CTV, and citrus nematode-related disease resistance genes containing the NB-ARC domain.

2.10. Gene family analysis

OrthoMCL v2.0.9 [54] software was used for gene family identification. The basic algorithm process was to first use BLASTP v2.7.1 software to compare the protein sequence data sets of all the studied species with each other (*E*-value ≤1e−5) and obtain the similarity information between protein sequences; then, we used the MCL algorithm to cluster genes to obtain orthologs, paralogs, and single-copy homologous genes (each species only contains one copy of this gene). Finally, the identified single-copy orthologous gene was used as the input for downstream analysis.

2.11. Phylogenetic tree construction and divergence time estimation

The detected *P. polyandra* genes were clustered into families using the OrthoMCL v2.0.9 pipeline [46], with an E-value cut-off of 1e−5, and Markov chain clustering with a default inflation parameter in an all-to-all BLASTP analysis of entries for six species, including *Arabidopsis thaliana* (downloaded from NCBI), *Citrus clementina* (downloaded from NCBI), *C. sinensis* (downloaded from NCBI), *C. unshiu* (downloaded from NCBI), *P. trifoliata* (downloaded from Phytozome), and *P. polyandra* (from this study).

We codon-aligned each gene family using Mafft v7.313 and curated the alignments using Gblocks v0.91b [77]. Phylogenetic analysis was performed using RAxML v8.2.11 [75] with the GTRGAMMA model and 1000 bootstrap replicates. We used MCMCTREE, as implemented in PAML v4.9e [92], to estimate the divergence times of *P. polyandra* from other plants.

2.12. Positive selection analysis

According to the neutral theory of molecular evolution, the ratio of the non-synonymous substitution rate (*Ka*) to the synonymous substitution rate (*Ks*) of protein-coding genes can be used to identify genes that show signatures of natural selection. Thus, we calculated the average *Ka/Ks* values and conducted a branch-site likelihood ratio test using Codeml implemented in the PAML [92] package to identify positively selected genes in the *P. polyandra* lineage. Genes with a *p*-value <0.05 under the branch-site model were considered positively selected genes.

2.13. Gene family expansion and contraction analysis

To understand the relationships between the *P. polyandra* gene families and those of other plants, we performed a systematic comparison of genes among different species. To gain greater insight into the evolutionary dynamics of the genes, we determined the expansion and contraction of orthologous gene clusters using CAFE v4.2.1 software [16].

Furthermore, the enrichment pipeline software clusterProfiler [93] was used to test the statistical enrichment of expanded and contracted gene families in the KEGG and GO pathway analyses. Pathways with a Q-value <0.05 [76] were considered to be significantly enriched.

2.14. Whole-genome duplication (WGD) analysis and estimation of divergence

The synonymous substitution rate (*Ks*) estimation was used to detect WGD events in the *P. polyandra* genome [48]. First, protein sequences were extracted, and all-*vs*-all paralog analyses were performed using the best hits from primary protein sequences by self-BLASTP in plants. After filtering by identity and coverage, the BLASTP v2.6.0 results were subjected to MCScanX [84], and the respective collinear blocks were identified. Finally, *Ks* was calculated for the syntenic block gene pairs using KaKs_Calculator 2.0 [80], and the potential WGD events in each genome were evaluated based on their *Ks* distribution.

3. Results

3.1. Genome sequencing and assembly

A total of 21.61 Gb of raw data was obtained in this study. After filtering the original data, approximately 19.68 Gb of clean data were obtained. The estimated genome size was approximately 341.36 Mb, and the heterozygosity was approximately 3.40%. After assembly and calibration, the corrected genome was 315.78 Mb in size, with a scaffold N50 length of 32.07 Mb (Table 1). We used Hi-C chromosome conformation capture data to achieve chromosome-scale scaffolding (Fig. S1), and the final valid data accounted for 34.97% of the total HiC sequencing data. After LACHESIS clustering, the sequence was located on nine chromosomes, accounting for 99.33% of the total length, with the largest scaffold of 46.06 Mb on Chr1 (Table S1).

Table 1
Summary statistics of the *P. polyandra* and closely related species genome assembly and annotation.

Parameter	*Poncirus polyandra*	*Poncirus trifoliata*	*Citrus clementina*	*Citrus sinensis*
Total size of genome assembly	315.78 Mb	265 Mb	301.4 Mb	327.9 Mb
Chromosome number (2n)	18	18	18	18
Number of chromosomal pseudomolecules	9	9	9	9
Number of scaffolds	46	152	1398	4994
Longest scaffold	46.06 Mb	42.7 Mb	51.1 Mb	88.95 Mb
Scaffold N50 length	32.07 Mb	27.7 Mb	31.4 Mb	30.8 Mb
Number of contigs	112	809	8962	17,382
Contig N50 length	7.57 Mb	842.8 kb	109.9 kb	49.9 kb
GC content	40%	33.9%	34.5%	34%
Number of gene models	21,278	25,538	24,533	29,406
Mean transcript length	3717.79 bp	1667 bp	1814 bp	1882 bp
Mean coding sequence length	1322.71 bp	1268 bp	1480 bp	1454 bp
Mean exon length	223.95 bp	296 bp	317 bp	336 bp
Average exons number per gene	5.91		5.72	5.86
Average intron length	488.18 bp		430 bp	548 bp
Percentage of TEs	46.00%	42.6%		
TE, transposable element	346,948	272,734		

3.2. Repeat sequences

The combined results of the homology-based and *de novo* predictions indicated that the total amount of repeated sequences was 145.27 Mb, accounting for 52.23% of the *P. polyandra* genome assembly. A large proportion (46%) of the genome contained TEs (Table S2). Long terminal repeats (LTRs) were predominant (29.05%), including LTR/gypsy (11.34%) and LTR/copia (10.38%). Other types of repeats accounted for smaller proportions of the genome, including DNA transposons (11.57%), LINE (3.72%), and SINE (0.29%).

3.3. Genome annotation

A total of 21,278 genes were predicted from the genome, with an average gene length of 3717.79 bp and an average CDS length of 1322.71 bp. The average number of exons in each gene was 5.91; the average exon length was 223.95 bp, and the average intron length was 488.18 bp (Table 1).

For RNA, 655 rRNAs, 557 small RNAs, nine regulatory RNAs, and 32 tRNAs were predicted (Table S3). The annotation results of the KOG (Fig. 1), GO (Fig. 2), and KEGG (Fig. 3) protein databases showed that 20,815 genes could be annotated in the functional database, accounting for 97.82% of all genes (Table S4).

According to the NCBI KOG mapping, 11,485 proteins (53.98%) were assigned to the KOG categories. The “General function prediction only” group had the highest number of genes (2436). This was followed by “Posttranslational modification, protein turnover, chaperones” (1360), “Signal transduction mechanisms” (1194), “Transcription” (790), “Carbohydrate transport and metabolism” (724), and “Function unknown” (693) as the most gene-rich classes in the KOG groupings. These results show that *P. polyandra* has a variety of protein and energy metabolic functions that can better absorb and transform nutrients.

According to the GO database, 12,905 predicted proteins that accounted for 60.65% of the entire homologous protein were mainly distributed in five functional entries: “Metabolic process,” “Cellular process,” “Binding,” “Single-organism process,” and “Catalytic activity.”

To further understand the gene functions of *P. polyandra*, we assigned 8150 putative proteins to homologs in the KEGG database, the functional classification of which is shown in Fig. 3. Some metabolic and biosynthesis categories in KEGG were highly rich, which is similar to the KOG annotation.

The BUSCO evaluation revealed that approximately 96.84% of the complete gene elements could be found in the annotated gene set, indicating that the prediction of most conserved genes was relatively complete, which reflected the high reliability of the prediction results.

3.4. Identification of cold signaling-related genes, NBS-LRR genes, and TFs

According to the analysis using existing annotations of the *P. polyandra* genome, the homologs of the *A. thaliana* cold-signaling pathway genes were assigned to 41 gene families containing 83 genes in *P. polyandra* (Figs. 4 and 6; Table S5).

A total of 89 NBS genes, accounting for only 0.43% of the 20,815 annotated genes, were identified in the *P. polyandra* genome, which was much fewer than those in the *P. trifoliata* genome (Table S6). These NBS genes were further classified into different classes based on the presence of the TIR, CC, and LRR domains [2]. The two largest classes with at least one of these domains were the ‘CC-NBS-LRR’ (17; 19.1%) and ‘TIR-NBS-LRR’ (13; 14.6%) types. Compared with other domains, the ‘RPW8-NBS-LRR’ (3; 3.4%) type was the least common, and there was no ‘TIR-CC-NBS’ type in *P. polyandra*, but it was found in the *P. trifoliata* genome (Table S7). Through phylogenetic analysis, the NBS genes were classified into four major groups (Fig. 5). TF analysis revealed 1281 putative TF genes in 58 gene families, representing 6.02% of the predicted gene models in the *P. polyandra* genome (Table S8). The number of TF genes in *P. polyandra* was lower than that in *P. trifoliata*. The predominant TF family in *P. polyandra* was bHLH (117), followed by MYB (110), NAC (106), and ERF (86).

3.5. Candidate genes associated with cold signaling, HLB tolerance, and resistance to CTV and citrus nematodes

We screened three candidate genes related to the CBF pathway through sequence alignment, named *CBF2/DREB1B*, *CBF3/DREB1C-1*, and *CBF3/DREB1C-2*, according to the alignment results (Fig. 6). Using the same method, we screened 19 HLB tolerance genes, 11 CTV, and one citrus nematode resistance gene, and named these candidate genes in order (Table S9).

3.6. Gene family analysis and evolutionary dynamics among Citrus-related species

We used existing annotations of the genomes of closely related species (Table S10) to allocate gene families and conduct a global view. Gene family analysis and statistical analysis of the genes contained in each category were subsequently performed (Fig. 7a). As the target species of this analysis, *P. polyandra* had 6114 single-copy orthologs, 363 unique paralogs, 1946 unclustered genes, and 2309 species-specific genes combined with unclustered and unique genes (Table S11). To obtain more characteristics of the gene families, we performed statistics in different forms for the analysis results of the gene families (Table S12). The statistical results showed that 21,278 genes of the target species *P. polyandra* were clustered into 15,162 gene families, 111 of which were unique to *P. polyandra*. The number of *P. polyandra* genes that could be classified into different gene families was 19,332, and the number of unclustered genes was 1946, with an average of 1.28 genes per family. Through GO term enrichment analysis of *P. polyandra*-specific genes, we found that these genes were mainly enriched in five categories of molecular functions: iron ion binding (GO: 0005506, 44), oxidoreductase activity (acting on paired donors, with incorporation or reduction of molecular oxygen, GO: 0016705, 38), heme binding (GO: 0020037, 45), methyltransferase activity (GO: 0008168, 28), and *O*-methyltransferase activity (GO: 0008171, 9). To explore the phylogenetic relationships among the six species, a highly confident

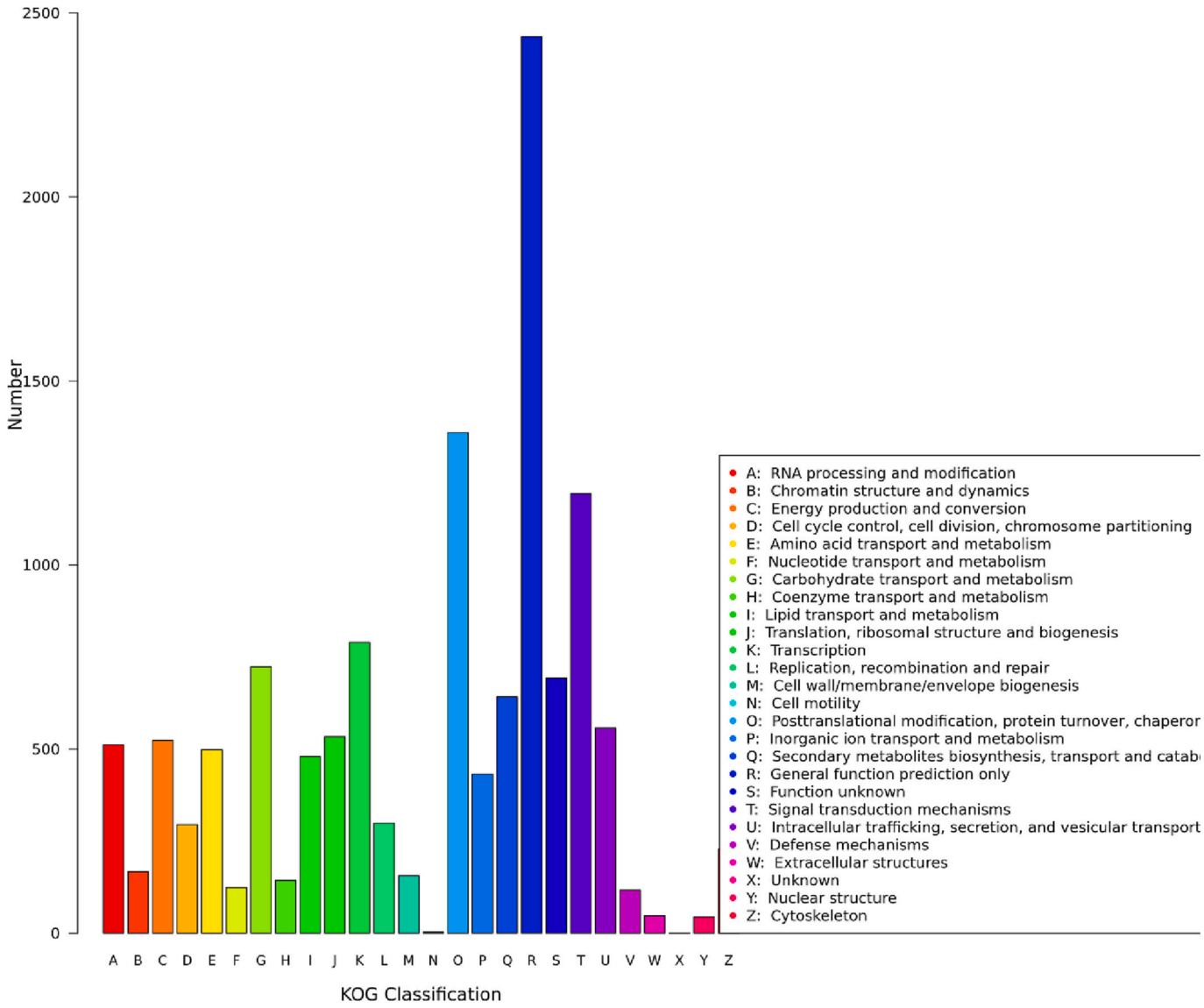

Fig. 1. The Eukaryotic Orthologous Groups of protein (KOG) function annotation of *P. polyandra*.

phylogenetic tree was constructed using the protein sequences of 6114 single-copy orthologs (Fig. 7b). Phylogenetic analysis showed that *P. polyandra*, *P. trifoliata*, and *Citrus* clades diverged from a common ancestor at ~9.53 Mya. The estimated divergence time for *P. polyandra* and *P. trifoliata* was ~1.96 Mya.

According to the results of the above gene families and the phylogenetic tree, we also found the evolution of gene families of different species in each evolutionary branch (Fig. 7c). Among these, there were 519 expanded groups and 2889 contracted groups in *P. polyandra*. Through gene family expansion and contraction analysis using existing annotations, 53 rapidly evolving gene families were significantly expanded (31) or contracted (22) across the species tree in *P. polyandra* (p-value $\leq$0.01) (Table S13). To further investigate gene function, GO enrichment analysis was performed on the 53 rapidly evolving gene families of *P. polyandra* (Fig. 8). Five GO terms were significantly enriched for rapidly evolving genes: ADP binding (GO: 0043531), oxidoreductase activity (acting on paired donors with incorporation or reduction of molecular oxygen, GO: 0016705), iron ion binding (GO: 0005506), heme binding (GO: 0020037), and zinc ion binding (GO: 0008270).

3.7. WGD

To investigate the evolutionary history of *P. polyandra* and other *Citrus* genomes, *Ks* was estimated for the paralogous gene groups in each species. *P. polyandra* and *P. trifoliata* had similar peak patterns, which proved that their genetic relationship was the closest (Fig. S2). Peng et al. (2020) previously observed similar results in *P. trifoliata* and confirmed that the *Citrus-Poncirus* lineage shared an ancient WGD event, which is also shared with other eudicots [5]. These results demonstrate that from an evolutionary perspective, the *P. polyandra/P. trifoliata* split is one of the most recent divergences among *Citrus*-related species.

3.8. Genes under positive selection

To evaluate adaptive evolution in the *P. polyandra* genome, we identified genes under positive selection. Finally, 471 positive selection characteristic genes were identified, 392 of which had potential functions in the SwissProt database (Table S14). GO term enrichment analysis of *P. polyandra* positively selected genes revealed that they were significantly enriched in DNA-binding transcription factor activity (GO:0003700), zinc ion binding (GO: 0008270), nucleus (GO: 0005634), and sequence-specific DNA binding (GO: 0043565) (Fig. S3).

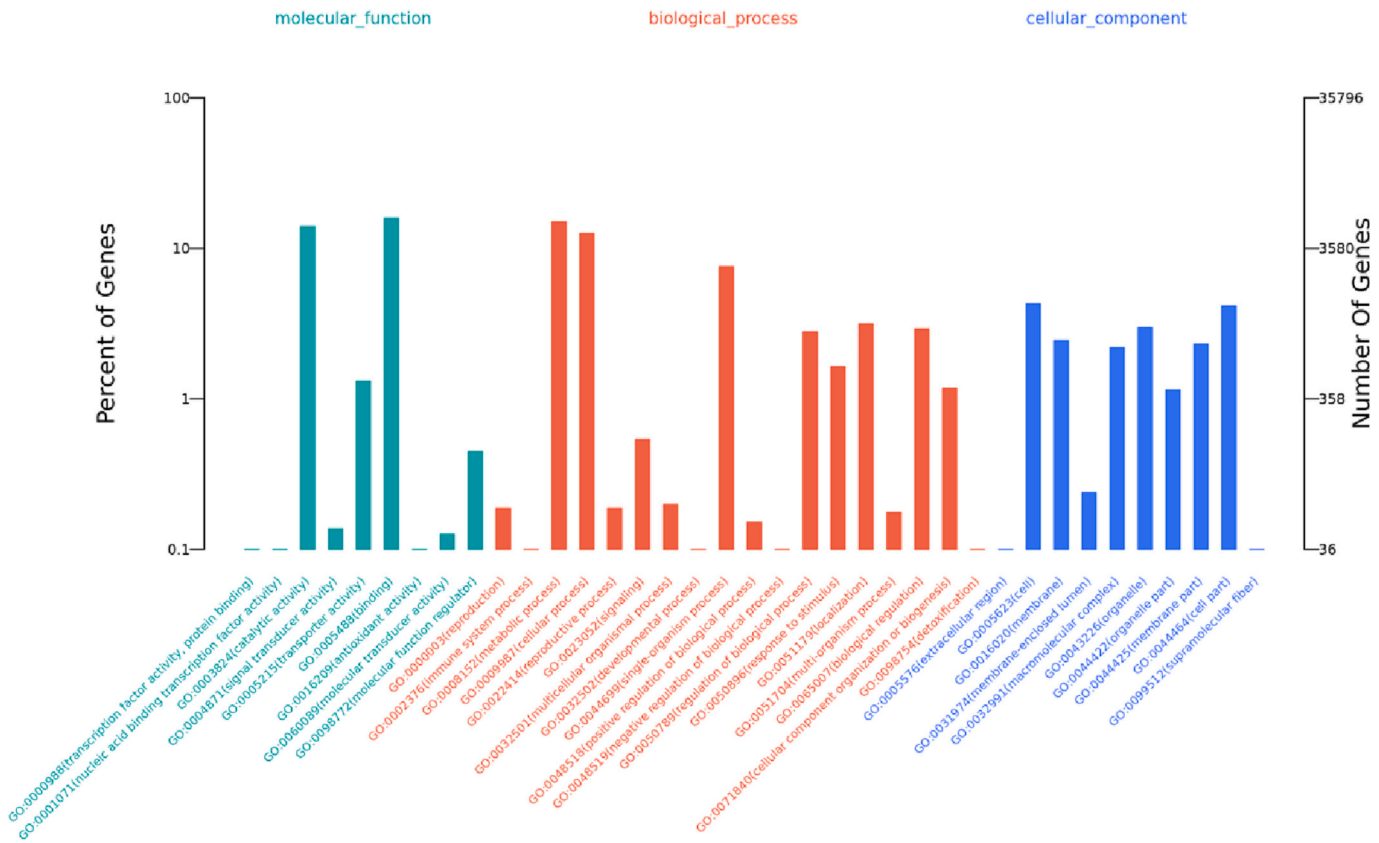

Fig. 2. Gene Ontology (GO) function annotation of *P. polyandra*.

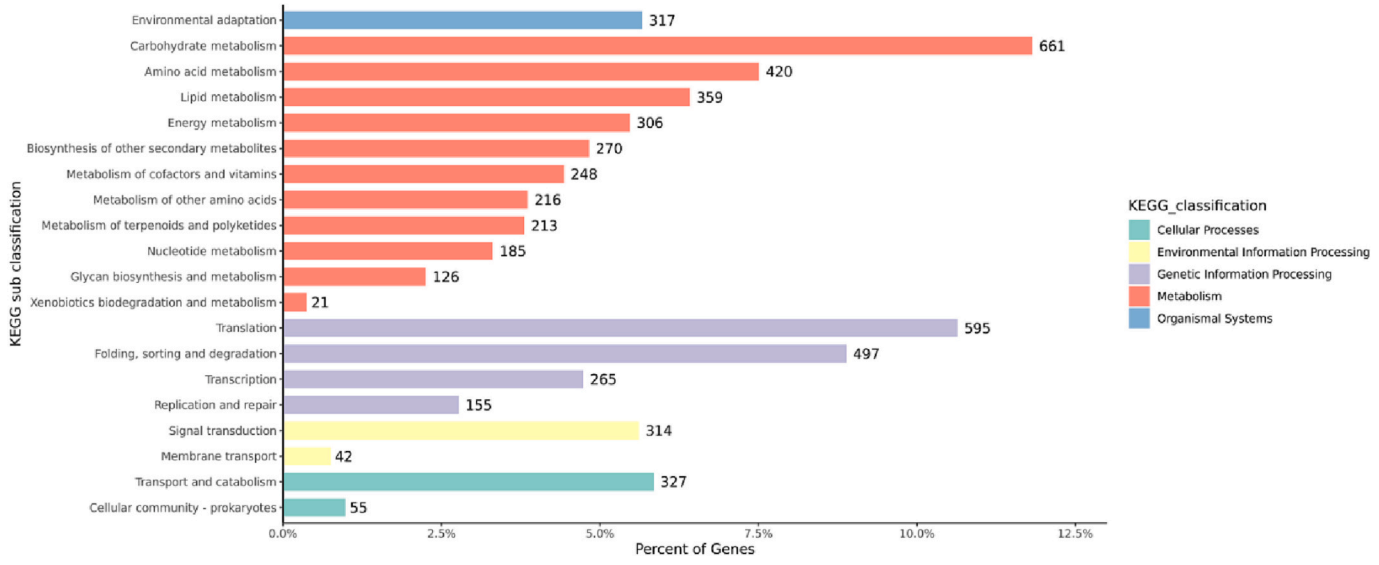

Fig. 3. The Kyoto Encyclopedia of Genes and Genomes (KEGG) function annotation of *P. polyandra*.

4. Discussion

A high-quality reference genome is a prerequisite for the effective and efficient mining of crop genomes. To date, only the chloroplast genome of *P. polyandra* has been sequenced and analyzed [55,91], and a *P. polyandra* genome assembly is not yet available. Therefore, there is an urgent need to obtain a high-quality chromosome-scale reference genome for *P. polyandra* for further research on its genetic resources. Our study fills this gap and lays a solid foundation for future genetic, genomic, and molecular studies of this important species. The Hi-C data and the resulting scaffolding improved the assembly N50 length by approximately 32-fold. To the best of our knowledge, this is the first assembly of the *P. polyandra* genome to yield nine chromosome-level scaffolds, reflecting an important milestone for future genetic studies on *P. polyandra.* Compared with the aforementioned closely related species (*C. clementina*, *C. sinensis*, *C. unshiu*, and *P. trifoliata*), this genome is the third reference genome assembly at the chromosome scale in the family *Rutaceae*, following *C. sinensis* [90] and *C. maxima* [68,81]. The

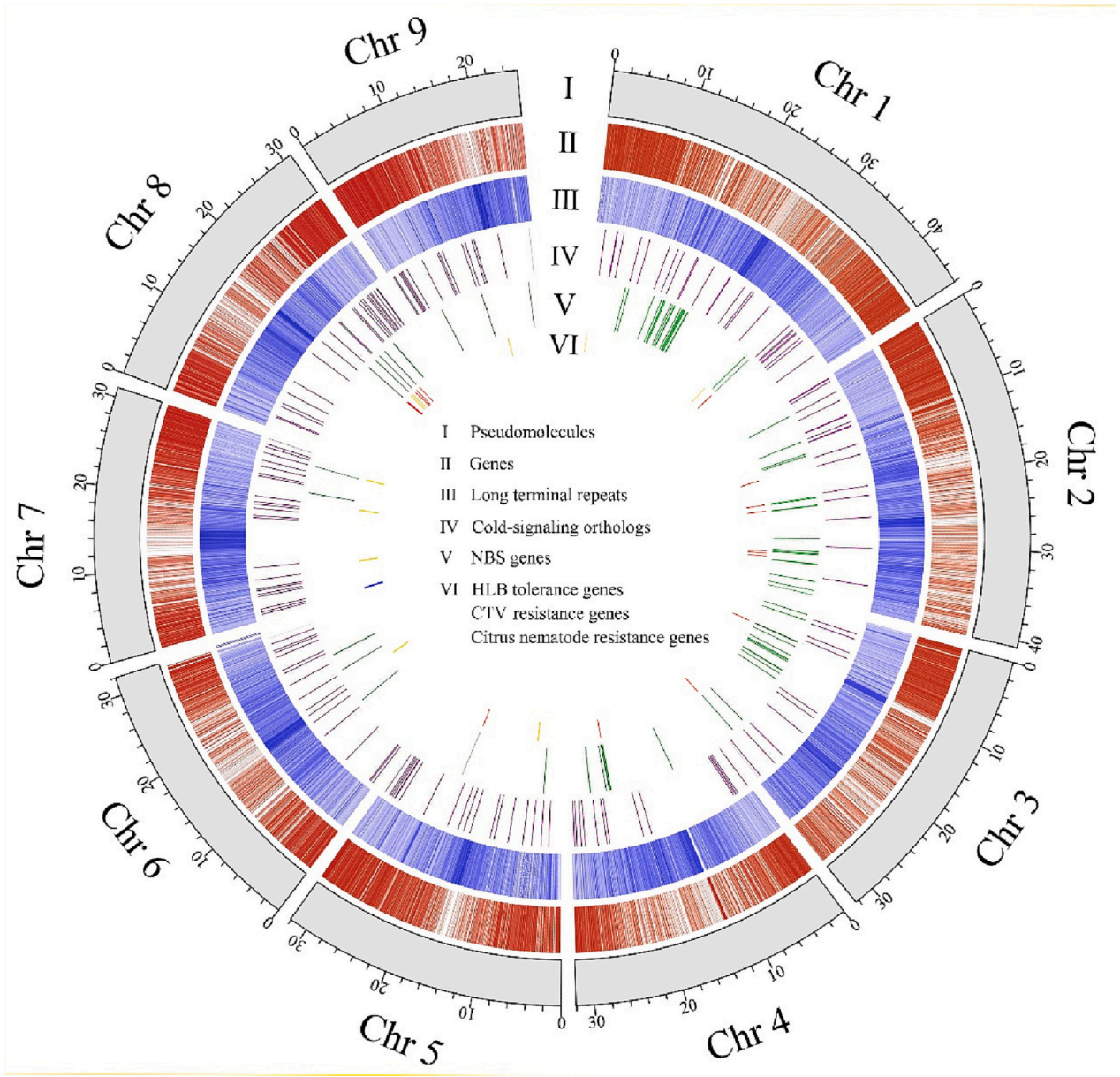

Fig. 4. Characterization of the *Poncirus polyandra* genome.
I. The nine assembled pseudomolecules (in Mb) corresponding to the nine chromosomes (Chr1–Chr9) of *P. polyandra*.
II. Locations of predicted gene models.
III. Locations of predicted long terminal repeat (LTR) transposable elements (TEs).
IV. Locations of predicted homologous cold-signaling genes.
V. Locations of predicted nucleotide-binding site (NBS)-containing genes.
VI. Red: HLB-tolerance genes. Yellow: citrus tristeza virus (*Ctv*) resistance genes. Blue: citrus nematode resistance genes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

completeness of this genome was 98.82% based on the BUSCO score, which was higher than *P. trifoliata* (97.20%) [68], *C. reticulata* (96.00%) [81], and *F. hindsii* (93.10%) [101]. High completeness and N50 length supported the quality of the genome assembly.

TEs are sources of genetic variation in animals and plants and contribute to their evolution and genetic diversity [5,7,47]. In *P. polyandra*, we identified 346,948 copies of *P. polyandra* belonging to six TE classifications. A considerable proportion of the *P. polyandra* genome comprised TEs (46.00%), and the most predominant repeats were LTRs (29.05%), similar to other *Citrus*-related genomes [68]. The number of TEs was higher (46%) than that of *P. trifoliata* (42.60%), indicating a greater possibility of evolution and variation in the genome of *P. polyandra*.

Resistance genes have been identified in several model plant species [70,94]. *Poncirus polyandra*, as in other species of *Poncirus* (*P. trifoliata*), is supposed to contain resistance genes that help exert many desirable traits, including resistance to HLB, CTV, nematodes, and cold stress [13,68,99]. Initially, we utilized already available labels to obtain an overall understanding of how genes are categorized into families. This allowed us to identify a preliminary set of genes or gene families of interest, including candidate genes associated with cold signaling, resistance to HLB, CTV, and citrus nematodes, rapidly evolving gene families, *P. polyandra*-specific genes, and single-copy genes, by matching with *P. trifoliata* or *A. thaliana*.

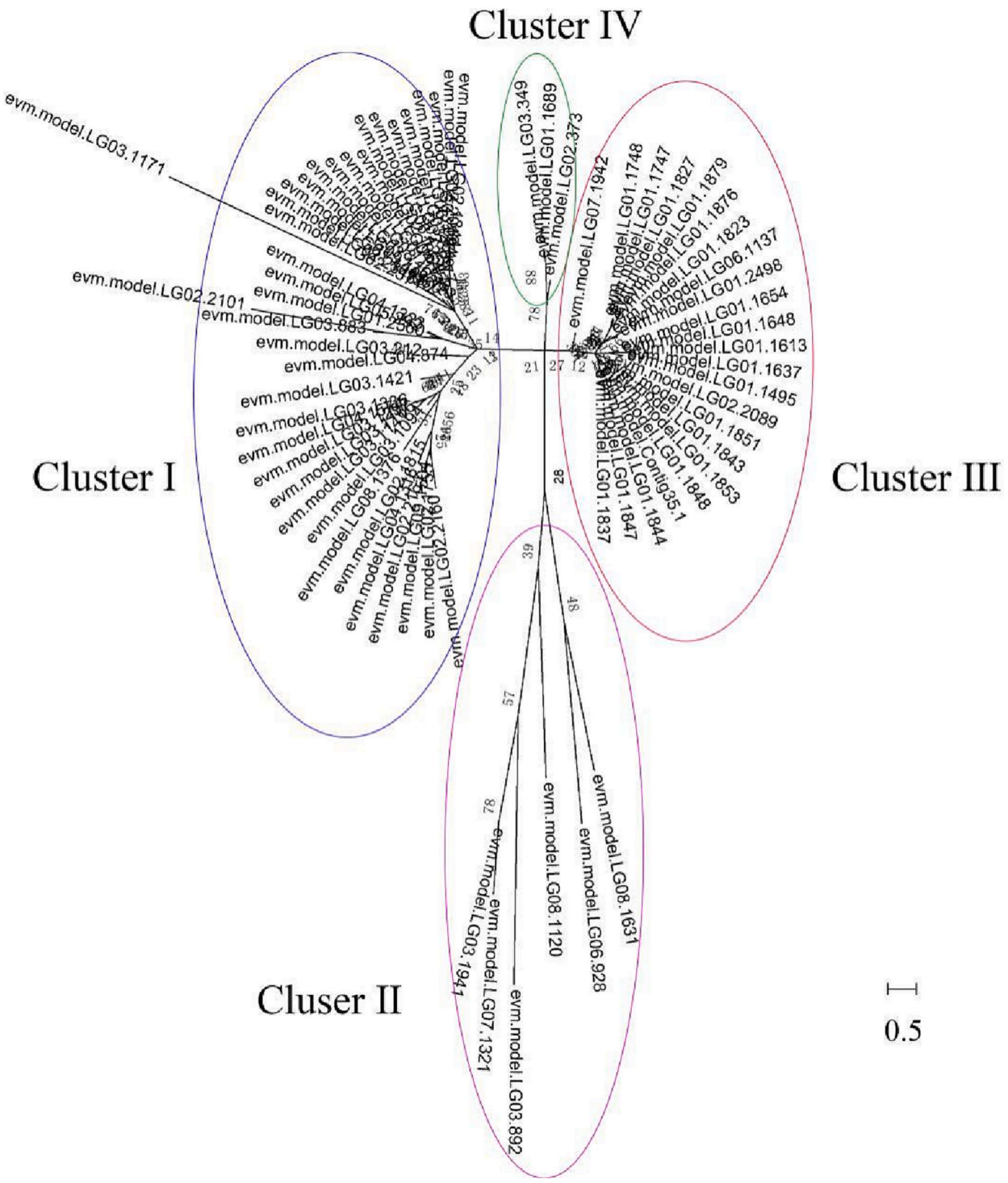

Fig. 5. Phylogenetic tree of identified NBS genes in *P. polyandra*. Four major clusters were identified.

The NBS is a widely occurring domain that is present in the protein products of resistance genes, which are involved in encoding signal transduction pathways [2,34]. In particular, NBS-LRR genes, the most common type of resistance gene in many plant species, encode proteins with variable N-terminal domains connected by a predicted NBS domain and a short LRR motif [44,61,71]. In the present study, 89 NBS genes were classified into four major groups, of which we identified 17 rapidly evolving NBS-type LRR genes in *P. polyandra.* The results showed abundant variation in the domains of *P. polyandra* NBS genes, although the number of NBS genes was lower than that in the *P. trifoliata* genome.

Previous studies on the genome of the model plant *A. thaliana* have shown that CBF TFs play crucial roles in plants under cold environmental conditions by regulating numerous cold-regulated genes and activating downstream target cold-regulated genes or CBF regulons. CBF1–CBF3, also known as dehydration-responsive element binding factor 1 (DREB1) proteins, consist of three members of this family [28,62,64]. Based on the hypothesis that *P. polyandra* has cold tolerance-related genes that may be homologous to *Arabidopsis* cold signaling pathway genes, we identified three candidate genes associated with the cold signaling pathway in *P. polyandra* using sequence alignment [32].

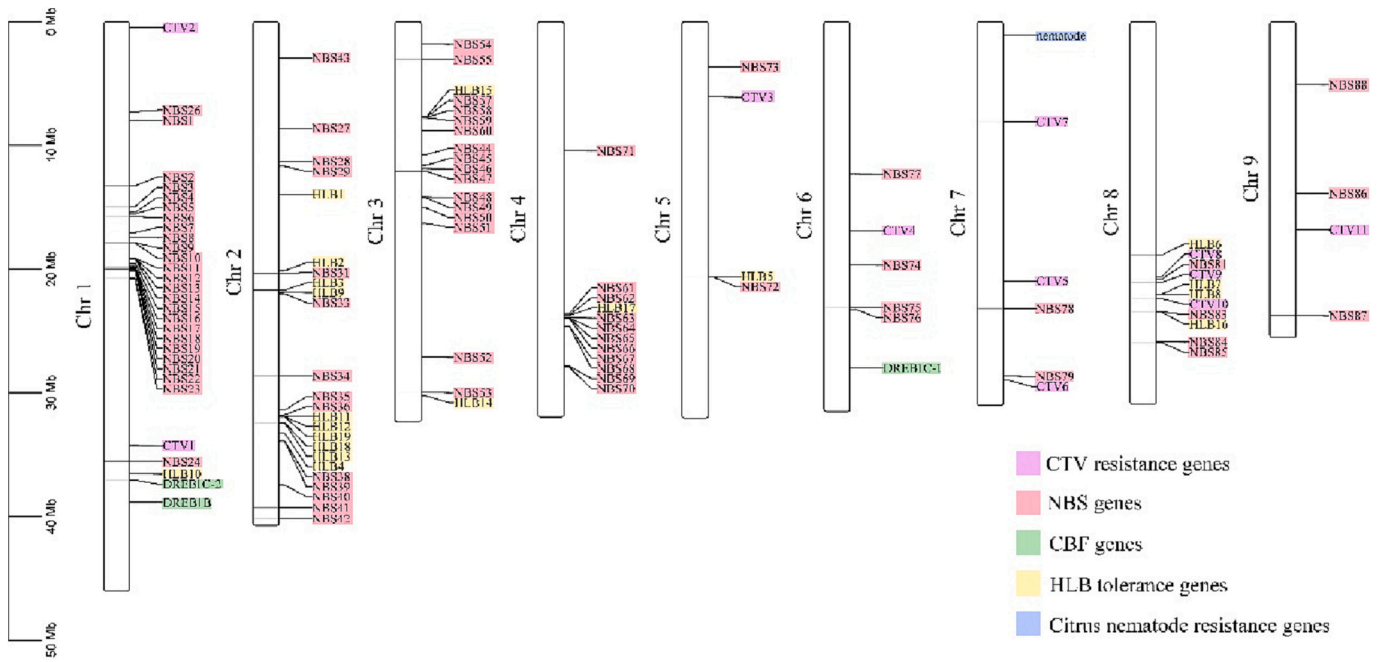

Fig. 6. Candidate genes associated with cold tolerance and disease resistance.
A graphic map showing candidate genes in the *P. polyandra* genome associated with tolerance to Huanglongbing (HLB) disease, citrus tristeza virus (CTV), and nematode resistance, as well as candidate genes associated with CBF and NBS.

With the availability of the *P. polyandr* genome, other candidate genes were identified. In total, 41 gene families containing 83 genes related to the *A. thaliana* cold signaling pathway, 1281 putative TF genes in 58 gene families, 19 HLB tolerance genes, 11 CTV genes, and one citrus nematode resistance gene were identified in *P. polyandra* genome, in addition to the NBS and CBF genes mentioned above. The physical locations of the identified NBS, CBF, HLB, CTV, and citrus nematode candidate genes were mapped onto nine assembled chromosomes of the *P. polyandra* genome using TBtools. Strikingly, there were seven HLB tolerance candidate genes with the NBS domain. Overall, these genes could serve as candidates for further investigation of the genetic factors conferring HLB tolerance in *P. polyandra*. In addition, we found that NBS genes were mainly located on Chr1–Chr4, and that there were gene clusters enriched with NBS on these four chromosomes. HLB tolerance genes were mainly located on Chr2 and Chr8, whereas CTV resistance genes were mainly located on Chr7 and Chr8. Therefore, we speculated that chromosomes 2, 7, and 8 of *P. polyandra* might play an important role in the resistance and tolerance to HLB, CTV, and citrus nematodes.

Our analysis using existing annotations also revealed that *P. polyandra* has 2309 species-specific genes combined with unclustered and unique genes that may represent the diversity or uniqueness of *P. polyandra*. In *P. polyandra*, 53 rapidly evolving gene families have been significantly expanded (31) or contracted (22). GO enrichment analysis showed that five GO terms belonging to molecular functions were significantly enriched in rapidly evolving genes. Notably, we found that the rapidly evolving specific genes were mainly classified into biological process and molecular function categories in *P. polyandra*.

The use of highly conserved and low-copy nuclear genes to construct a phylogenetic tree is helpful in understanding the evolutionary relationships among species [68]. With the availability of six related genomes, we constructed a highly reliable phylogenetic tree using single-copy protein orthologs. In this analysis, *P. polyandra* was proposed to have the closest relationship with *P. trifoliata*, with a divergence time of 1.96 Mya, consistent with previous reports [68,91,102]. This implies that the *P. polyandra* genome will be useful for citrus phylogenetic studies. Overall, the enriched functions of lineage-specific and rapidly evolving genes in *P. polyandra* may be essential for its adaptive evolution and may contribute to the characteristics specific to this species. From the results of WGD, *P. polyandra* and *P. trifoliata* had the closest genetic relationship, which was consistent with the results previously observed in *P. trifoliata* [68] and confirmed that the *Citrus-Poncirus* lineage shared an ancient WGD event, which was also shared with other eudicots. These results confirm that the *P. polyandra*/*P. trifoliata* split is one of the most recent divergences among *Citrus*-related species from an evolutionary perspective. Previous studies have shown that *P. polyandra* is very important for the breeding and genetic improvement of citrus with a high genetic diversity in China [68,95–97]. Our study further proves that the exploration and preservation of natural genetic variation of *P. trifoliata* in China should be regarded as an important source of citrus germplasm resources protection.

5. Conclusion

To our knowledge, this is the first study to obtain a high-quality chromosome-level reference genome sequence of *P. polyandra* using multiple types of sequencing data and assembly technologies. Our study provides not only a valuable genetic resource for *P. polyandra* with resistance to disease and tolerance to cold, but also a better understanding for further research on gene function mining of *P. polyandra*, a plant species with extremely small populations in China. However, future studies should foucus on verifying whether these identified resistance genes have corresponding functions in transcriptome research.

Authors contributions

YW designed and supervised the manuscript. ZS analyzed the data. CJ performed genome assembly. ZC and ZS performed genome annotation. ZS performed comparative genomics and evolutionary analyses. ZX and GL prepared and collected plant materials, isolated DNA, and coordinated sequencing. CJ and ZS participated in functional genes identification. ZS wrote the original manuscript. CJ, ZC, ZS, ZX, and GL critically revised the manuscript. All authors have read and agreed to the published version of the manuscript.

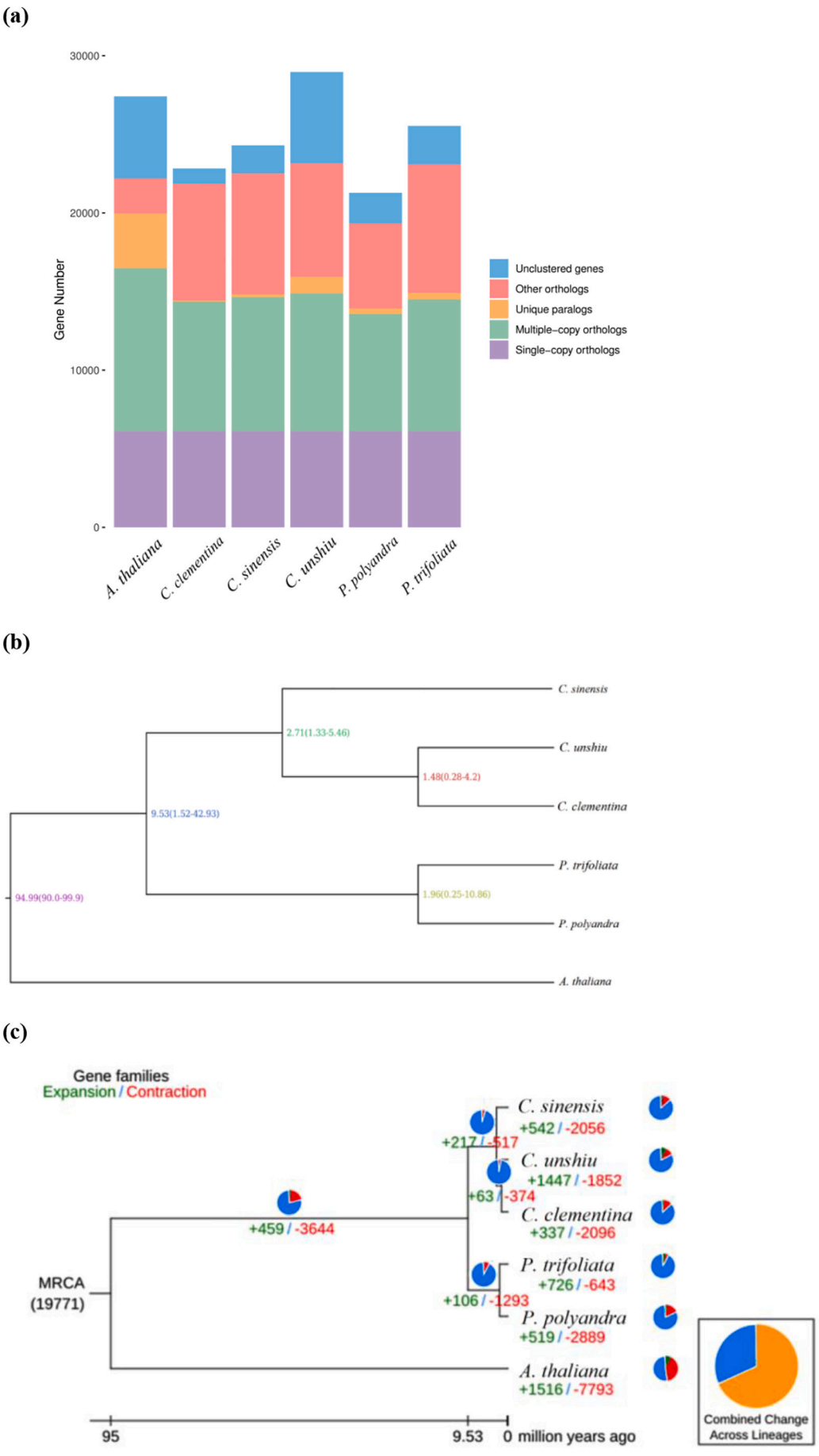

Fig. 7. Comparative genomics of the six related species analyzed in this study.
(a) Gene numbers for different types of orthologous/paralogous gene family relationships in six species.
(b) A phylogenetic tree containing six analyzed species.
(c) Gene family expansions and contractions in *P. polyandra* and five other plants.

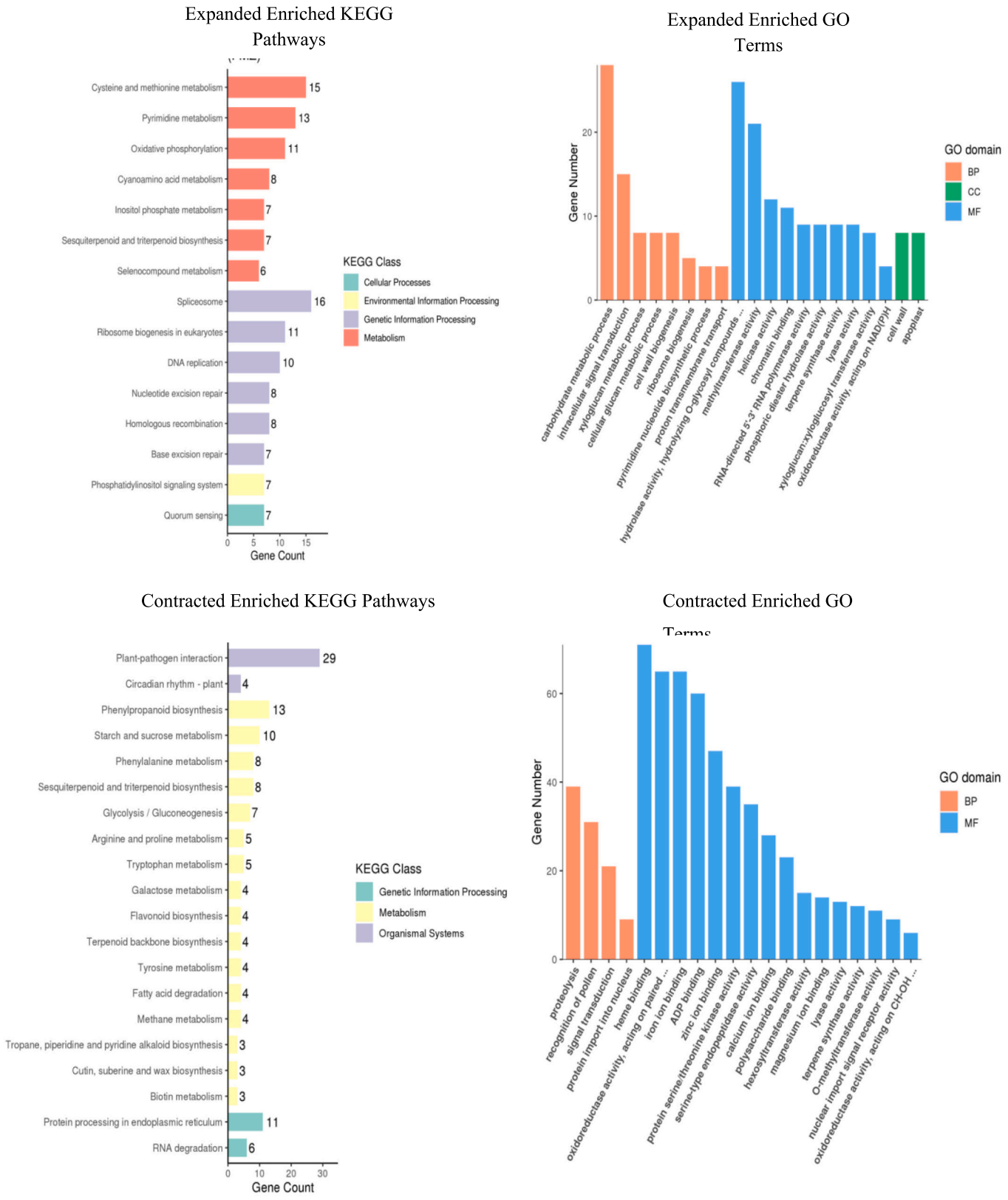

Fig. 8. KEGG and GO pathway analysis of expanded and contracted gene families.

Funding sources

This research was supported by the National Natural Science Foundation of China (31960278).

Declaration of Competing Interest

The authors declare no competing interests.

Data availability

All relevant data are within the manuscript and its Supporting data files. The whole-genome sequencing data from this study have been deposited in the Sequence Read Archives (SRA) of the National Center Biotechnology Information (NCBI) at SRA (BioProject: PRJNA940198, SRA accession number: SRR23760602-SRR23760608, SRR237606010-SRR237606011, SRR23825300, and SRR23825301). The raw and clean fastq files were deposited at SRA (BioProject: PRJNA940584, SRA accession number: SRR23692455). The genome assembly data are available in SRA under SRA accession of SRR23760599 (BioProject: PRJNA940612).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110617.

References

[1] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A. P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L.A. Issel-Tarver, S. Kasarskis, J.C. Lewis, J.E. Matese, M. Richardson, G.M. Ringwald, G. Sherlock Rubin, Gene ontology: tool for the unification of biology. The gene ontology consortium, Nat. Genet. 25 (2000) 25–28.

[2] B. Baker, P. Zambryski, B. Staskawicz, S.P. Dinesh-Kumar, Signaling in plant-microbe interactions, Science 276 (1997) 726–733.

[3] W. Bao, K.K. Kojima, O. Kohany, Repbase update, a database of repetitive elements in eukaryotic genomes, Mob. DNA 6 (2015) 11.

[4] L.P. Boava, M. Cristofani-Yaly, V.S. Mafra, K. Kubo, L.T. Kishi, M.A. Takita, M. Ribeiro-Alves, M.A. Machado, Global gene expression of *Poncirus trifoliata*, *Citrus sunki* and their hybrids under infection of *Phytophthora parasitica*, BMC Genomics 12 (2011) 39.

[5] C. Borredá, E. Pérez-Román, V. Ibanez, J. Terol, M. Talon, Reprogramming of retrotransposon activity during speciation of the genus *Citrus*, Genome Biol. Evol. 11 (2019) 3478–3495.

[6] J.N. Burton, A. Adey, R.P. Patwardhan, R. Qiu, J.O. Kitzman, J. Shendure, Chromosome-scale contiging of *de novo* genome assemblies based on chromatin interactions, Nat. Biotechnol. 31 (2013) 1119–1125.18.

[7] E. Butelli, C. Licciardello, Y. Zhang, J. Liu, S. Mackay, P. Bailey, G. Reforgiato-Recupero, C. Martin, Retrotransposons control fruitspecific, cold-dependent accumulation of anthocyanins in blood oranges, Plant Cell 24 (2012) 1242–1255.

[8] D.S. Cali, J.S. Kim, S. Ghose, C. Alkan, O. Mutlu, Nanopore sequencing technology and tools for genome assembly: computational analysis of the current state, bottlenecks and future directions, Brief. Bioinform. 20 (2019) 1542–1559.

[9] C. Camacho, G. Coulouris, V. Avagyan, M. Ning, T.L. Madden, BLAST+: architecture and applications, BMC Bioinform. 10 (2009) 421.

[10] Z. Chan, Y. Wang, M. Cao, Y. Gong, Z. Mu, H. Wang, Y. Hu, X. Deng, X.J. He, J. K. Zhu, RDM4 modulates cold stress resistance in *Arabidopsis* partially through the CBF-mediated pathway, New Phytol. 209 (2016) 1527–1539.

[11] S.F. Chen, Y.Q. Zhou, Y.R. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (2018) 884–890.

[12] Y. Cheng, W. Guo, H. Yi, X. Pang, X. Deng, An efficient protocol for genomic DNA extraction from *Citrus* species, Plant Mol. Biol. Report. 21 (2003) 177–178.

[13] B. Dahro, Y. Wang, A. Alhag, C.L. Li, D.Y. Guo, J.H. Liu, Genome-wide identifcation and expression profling of invertase gene family for abiotic stresses tolerance in *Poncirus trifoliat*a, BMC Plant Biol. 21 (2021) 1–18.

[14] W.S. Dai, M. Wang, X.Q. Gong, J.H. Liu, The transcription factor FcWRKY40 of *Fortunella crassifolia* functions positively in salt tolerance through modulation of ion homeostasis and proline biosynthesis by directly regulating *SOS2* and *P5CS1* homologues, New Phytol. 219 (2018) 972–989.

[15] P. Danecek, S.A. McCarthy, BCFtools/csq: haplotype-aware variant consequences, Bioinformatics 33 (2017) 2037–2039.

[16] T.J. De Bie, N. Cristianini, J.P. Demuth, M.W. Hahn, CAFE: a computational tool for the study of gene family evolution, Bioinformatics 22 (2006) 1269–1271.

[17] M. Delorenzi, T. Speed, An HMM model for coiled-coil domains and a comparison with PSSM-based predictions, Bioinformatics 18 (2002) 617–625.

[18] Z. Deng, S. Huang, P. Ling, C. Chen, C. Yu, C.A. Weber, G.A. Moore, F.G. Gmitter, Cloning and characterization of NBS-LRR class resistance-gene candidate sequences in citrus, Theor. Appl. Genet. 101 (2000) 814–822.

[19] S.Q. Ding, X.N. Zhang, Z.R. Bao, M.Q. Liang, A new species of *Poncirus* from China, Acta Bot. Yunnanica 6 (1984) 292–923.

[20] Y. Ding, Y. Shi, S. Yang, Molecular regulation of plant responses to environmental temperatures, Mol. Plant 13 (2020) 544–564.

[21] E.P. Ducharme, Resistance of *Poncirus trifoliata* rootstock to nematode infestation in Argentina, Citrus Ind. 29 (1948) 9–15.

[22] M.T. Fan, M.Q. Liang, W.Q. Pu, Studies on the pollen morphology and taxonomic position of *P. trifoliate* and *P. polyandra*, J. Yunnan Agric. Univ. 13 (1998) 298–300.

[23] D.Q. Fang, Intra and intergeneric relationships of *Poncirus polyandra*: investigation by leaf isozymes, J. Wuhan Bot. Res. 11 (1993) 34–40.

[24] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, Nucleic Acids Res. 39 (2011) W29–W37.

[25] R.D. Finn, P. Coggill, R.Y. Eberhardt, S.R. Eddy, J. Mistry, A.L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, The Pfam protein families database: towards a more sustainable future, Nucleic Acids Res. 44 (2016) D279–D285.

[26] M.Y. Galperin, K.S. Makarova, Y.I. Wolf, E.V. Koonin, Expanded microbial genome coverage and improved protein family annotation in the COG database, Nucleic Acids Res. 43 (2015) D261–D269.

[27] X.S. Gao, Y.P. Zhang, J.W. Tang, W.H. Gu, Discussion on survival status, protection and development of *Poncirus polyandra*, Prot. Forest Sci. Technol. 12 (2017) 83–85.

[28] J.J. Geng, T.L. Wei, Y. Wang, X.S. Huang, J.H. Liu, Overexpression of PtrbHLH, a basic helix-loop-helix transcription factor from *Poncirus trifoliata*, confers enhanced cold tolerance in pummelo (*Citrus grandis*) by modulation of H_2O_2 level via regulating a CAT gene, Tree Physiol. 39 (2019) 2045–2054.

[29] J. George, S.L. Lapointe, Host-plant resistance associated with *Poncirus trifoliata* influence oviposition, development and adult emergence of *Diaphorina citri* (Hemiptera: Liviidae), Pest Mang. Sci. 75 (2019) 279–285.

[30] F.G. Gmitter Jr., X.L. Hu, The possible role of Yunnan, China, in the origin of contemporary *Citrus* species (Rutaceae), Econ. Bot. 44 (1990) 267–277.

[31] X.Q. Gong, J. Liu, Genetic transformation and genes for resistance to abiotic and biotic stresses in *Citrus* and its related genera, Plant Cell Tissue Organ Cult. 113 (2013) 137–147.

[32] X. Guo, D. Liu, K. Chong, Cold signaling in plants: insights into mechanisms and regulation, J. Integr. Plant Biol. 60 (2018) 745–756.

[33] B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, O. White, C. R. Buell, J.R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments, Genome Biol. 9 (2008) R7.

[34] K.E. Hammond-Kosack, J.D.G. Jones, Plant disease resistance genes, Annu. Rev. Plant Physiol. Plant Mol. Biol. 48 (1997) 575–607.

[35] Y. Heo, Y. Cho, K.S. Ju, H.S. Cho, K.H. Park, H. Choi, J.K. Yoon, C. Moon, Y. B. Kim, Antiviral activity of *Poncirus trifoliata* seed extract against oseltamivir-resistant influenza virus, J. Microbiol. 56 (2018) 586–592.

[36] J. Hu, J.P. Fan, Z.Y. Sun, S.L. Liu, NextPolish: a fast and efficient genome polishing tool for long-read assembly, Bioinformatics 36 (2020) 2253–2255.

[37] M. Hu, X. Lan, Y.G. He, C.S. Zheng, J.X. Li, C.X. Hu, Q.L. Tan, Effects of different rootstocks on the contents of nutrient in leaves of Yunning no.1 *Citrus*×*Limon*, Hubei Agr. Sci. 53 (2014) 3066–3073.

[38] Y.X. Hu, H.F. Chen, Y.P. Song, S.S. Tan, D. Pi, Determination of flavonoids in Fructus *Poncirus trifoliata* and Fructus *Aurantii* by HPLC, J. Pharm. Anal. 37 (2017) 2162–2166.

[39] Y.X. Hu, Y.P. Song, S.S. Tan, W.L. Yang, Extraction process of *Poncirus* in Fructus *Poncirus trifoliata*, J. Jinggangshan Univ. 38 (2017) 83–86.

[40] Y.X. Hu, H.Q. Zhu, S.S. Tan, D. Pi, W.L. Yang, Isolation and purification of Poncirin form Fructus *Poncirus trifoliata* by multi-solvent partition, J. Jinggangshan Univ. 38 (2017) 85–88.

[41] M. Huang, M.L. Roose, Q. Yu, D. Du, Y. Yu, Y. Zhang, Z. Deng, E. Stover, F. G. Gmitter Jr., Construction of high-density genetic maps and detection of QTLs associated with Huanglongbing tolerance in citrus, Front. Plant Sci. 9 (2018) 1694.

[42] X.S. Huang, W. Wang, Q. Zhang, J.H. Liu, A basic helix-loop-helix transcription factor, PtrbHLH, of *Poncirus trifoliata* confers cold tolerance and modulates peroxidase-mediated scavenging of hydrogen peroxide, Plant Physiol. 162 (2013) 1178–1194.

[43] I. Kalvari, J. Argasinska, N. Quinones-Olvera, E.P. Nawrocki, E. Rivas, S.R. Eddy, A. Bateman, R.D. Finn, A.I. Petrov, Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families, Nucleic Acids Res. 46 (2018) D335–D342.

[44] V. Kanazin, L.F. Marek, R.C. Shoemaker, Resistance gene analogs are conserved and clustered in soybean, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 1746–1750.

[45] M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.

[46] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780.

[47] M.G. Kidwell, D. Lisch, Transposable elements as sources of variation in animals and plants, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 7704–7711.

[48] Y. Mao, GenoDup pipeline: a tool to detect genome duplication using the dS-based method, Peer. J. 7 (2019), e6303.

[49] K. Lagesen, P. Hallin, E.A. Rødland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res. 35 (2007) 3100–3108.

[50] B. Langmead, S. Salzberg, Fast gapped-read alignment with bowtie 2, Nat. Methods 9 (2012) 357–359.
[51] G. Li, Z.Q. Ouyang, Y. Qi, L.J. Lu, Protection measures on rare and endangered plants in Kunming region, Environ. Sci. Survey 29 (2010) 27–29.
[52] H. Li, R. Durbin, Fast and accurate long-read alignment with burrows-wheeler transform, Bioinformatics 26 (2010) 589–595.
[53] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079.
[54] L. Li, C.J. Stoeckert Jr., D.S. Roos, OrthoMCL: identification of ortholog groups for eukaryotic genomes, Genome Res. 13 (2003) 2178–2189.
[55] S. Li, D. Zong, A. Zhou, C. He, The complete chloroplast genome sequence of *Poncirus polyandra* (Rutaceae), an endangered species endemic to Yunnan Province, China, Mitochondrial. DNA B 4 (2019) 766–768.
[56] Z.Y. Li, Y.X. Chen, D.S. Mu, J.Y. Yuan, Y.J. Shi, H. Zhang, J. Gan, I. Ln, X.S. Hu, B. H. Liu, B.C. Yang, W. Fan, Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph, Brief Funct. Genom. 11 (2012) 25–37.
[57] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E.S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, Science 326 (2009) 289–293.
[58] J. Liu, Y. Shi, S. Yang, Insights into the regulation of C-repeat binding factors in plant cold signaling, J. Integr. Plant Biol. 60 (2018) 780–795.
[59] L.Q. Liu, J. Yang, Z.J. Gu, Karyomorphology and taxonomic position of *Poncirus polyandra* (Rutaceae), Acta Bot. Yunnanica 29 (2007) 198–200.
[60] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964.
[61] L. McHale, X. Tan, P. Koehl, R.W. Michelmore, Plant NBS-LRR proteins: adaptable guards, Genome Biol. 7 (2006) 1–11.
[62] J. Medina, R. Catalá, J. Salinas, The CBFs: three arabidopsis transcription factors to cold acclimate, Plant Sci. 180 (2011) 3–11.
[63] T. Meshi, M. Iwabuchi, Plant transcription factors, Plant Cell Physiol. 36 (1995) 1405–1420.
[64] R. Ming, Y. Zhang, Y. Wang, M. Khan, B. Dahro, J.H. Liu, The JA-responsive MYC2-BADH-like transcriptional regulatory module in *Poncirus trifoliata* contributes to cold tolerance by modulation of glycine betaine biosynthesis, New Phytol. 229 (2021) 2730–2750.
[65] E.W. Myers, The fragment assembly string graph, Bioinformatics 21 (2005) 79–85.
[66] T. Nussbaumer, M.M. Martis, S.K. Roessner, M. Pfeifer, K.C. Bader, S. Sharma, H. Gundlach, M. Spannagl, M.I.P.S. PlantsDB, A database framework for comparative plant genome research, Nucleic Acids Res. 41 (2013) 1144–1151.
[67] G. Parra, K. Bradnam, I. Korf, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes, Bioinformatics 23 (2007) 1061–1067.
[68] Z. Peng, J.V. Bredeson, G.H.A. Wu, S.H. Shu, N.H. Rawat, D.L. Du, S. Parajuli, Q. B. Yu, Q. You, D.S. Rokhsar, F.G.G. Jr, Z.N. Deng, A chromosome-scale reference genome of trifoliate orange (*Poncirus trifoliata*) provides insights into disease resistance, cold tolerance and genome evolution in *Citrus*, Plant J. 104 (2020) 1215–1232.
[69] A. Rahman, S.A. Siddiqui, R. Jakhar, S.C. Kang, Growth inhibition of various human cancer cell lines by imperatorin and limonin from *Poncirus trifoliata* Raf. Seeds, Anti Cancer Agents Med. Chem. 15 (2015) 236–241.
[70] N. Rawat, B. Kumar, U. Albrecht, D. Du, M. Huang, Q. Yu, Y. Zhang, Y.P. Duan, K. D. Bowman, F.G. Gmitter Jr., Z. Deng, Genome resequencing and transcriptome profiling reveal structural diversity and expression patterns of constitutive disease resistance genes in Huanglongbing-tolerant *Poncirus trifoliata* and its hybrids, Hortic. Res. 4 (2017) 77–84.
[71] S. Seah, K. Sivasithamparam, A. Karakousis, E.S. Lagudah, Cloning and characterisation of a family of disease resistance gene analogs from wheat and barley, Theor. Appl. Genet. 97 (1998) 937–945.
[72] N. Servant, N. Varoquaux, B.R. Lajoie, E. Viara, C.J. Chen, J.P. Vert, E. Heard, J. Dekker, E. Barillot, HiC-pro: an optimized and flexible pipeline for Hi-C data processing, Genome Biol. 1 (2015) 259.
[73] Y. Shi, Y. Ding, S. Yang, Molecular regulation of CBF signaling in cold acclimation, Trends Plant Sci. 23 (2018) 623–637.
[74] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210–3212.
[75] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30 (2014) 1312–1313.
[76] J.D. Storey, A direct approach to false discovery rates, J. R. Stat. Soc. Ser. B Stat Methodol. 64 (2002) 479–498.
[77] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, Syst. Biol. 56 (2007) 564–577.
[78] M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences, Curr. Protoc. Bioinformatics 25 (4) (2009), 10.1–4.10.14.
[79] TransposonPSI, An Application of PSI-Blast to Mine (Retro-) Transposon ORF Homologies. http://transposonpsi.sourceforge.net/. Accessed 18 March 2018.
[80] D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies, Genom. Proteom. Bioinform. 8 (2010) 77–80.
[81] L. Wang, F. He, Y. Huang, J. He, S. Yang, J. Zeng, C. Deng, X. Jiang, Y. Fang, S. Wen, R. Xu, H. Yu, X. Yang, G. Zhong, C. Chen, X. Yan, C. Zhou, H. Zhang, Z. Xie, R.M. Larkin, X. Deng, Q. Xu, Genome of wild mandarin and domestication history of mandarin, Mol. Plant 11 (2018) 1024–1037.
[82] M. Wang, W. Dai, J. Du, R. Ming, B. Dahro, J.H. Liu, ERF109 of trifoliate orange (*Poncirus trifoliata* (L.) Raf.) contributes to cold tolerance by directly regulating expression of Prx1 involved in antioxidative process, Plant Biotechnol. J. 17 (2019) 1316–1332.
[83] M. Wang, X. Zhang, J. Liu, Deep sequencing-based characterization of transcriptome of trifoliate orange (*Poncirus trifoliata* (L.) Raf.) in response to cold stress, BMC Genomics 16 (2015) 555.
[84] Y. Wang, H. Tang, J.D. Debarry, X. Tan, J. Li, X. Wang, T.H. Lee, H. Jin, B. Marler, H. Guo, J.C. Kissinger, A.H. Paterson, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res. 40 (2012), e49.
[85] T.L. Wei, Y. Wang, J.H. Liu, Comparative transcriptome analysis reveals synergistic and disparate defense pathways in the leaves and roots of trifoliate orange (*Poncirus trifoliata*) autotetraploids with enhanced salt tolerance, Hortic. Res. 7 (2020) 1689–1702.
[86] R.R. Wick, L.M. Judd, K.E. Holt, Performance of neural network basecalling tools for Oxford Nanopore sequencing, Genome Biol. 20 (2019) 129.
[87] X.E. Wu, M.T. Fan, J. Gao, W.X. Li, W.H. Liu, H. Xu, Studies on the RAPD analysis and taxonomic relationship of trifoliate orange, fuming trifoliate orange and Citrange, J. Yunnan Agric. Univ. 18 (2003) 277–280.
[88] X.X. Wu, Y. Tang, C.L. Deng, B.H. Liu, C.W. Chen, Observation of citrus pollen morphology by scanning electron microscopy, J. Fruit Sci. 35 (2018) 794–801.
[89] D.S. Xing, Carry out investigation on wild plants and alien invasive organisms, Yuxi Daily (2008) 005.
[90] Q. Xu, L.L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.B. Jiao, B. H. Hao, M.P. Lyon, J. Chen, S. Gao, F. Xing, H. Lan, J.W. Chang, X. Ge, Y. Lei, Q. Hu, Y. Miao, L. Wang, S. Xiao, K.B. Manosh, W. Zeng, F.H. Guo, X. Cao, X. Yang, Y.J. Xu, J. Cheng, J.H. Xu, J.L. Liu, Z. Oscar, W. Tang, H. Guo, H. Y. Kuang, L.R. Zhang, N. Mikeal, X.X. Niranjan, Y. Ruan Deng, The draft genome of sweet orange (*Citrus sinensis*), Nat. Genet. 45 (2013) 59–66.
[91] Y. Yang, H. Wu, L. Zhao, S. He, Analysis of *Poncirus polyandra* (Rutaceae) chloroplast genome and its phylogenetic implications, Mitochondrial. DNA B 4 (2019) 2287–2288.
[92] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (2007) 1586–1591.
[93] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS: Integr. Biol. 16 (2012) 284–287.
[94] X. Yu, S. Zhong, H. Yang, C. Chen, W. Chen, H. Yang, J. Guan, P. Fu, F. Tan, T. Ren, J. Shen, M. Zhang, P. Luo, Identification and characterization of *NBS* resistance genes in *Akebia trifoliata*, Front. Plant Sci. 29 (2021), 758559.
[95] S.S. Zhang, J. Chen, W.Z. Yang, Genetic diversity analysis of *Poncirus polyandra* by reduced-representation sequencing technique, J. Northeast For. Univ. 48 (2020) 36–41.
[96] S.S. Zhang, Y.H. Gan, W.H. Yang, Z.L. Duan, H.Q. Ding, H.M. Kang, N.S. Nama, Genetic diversity of *Poncirus polyandra* resources, J. West China Sci. 47 (2018), 14–20, 25.
[97] S.S. Zhang, W.Z. Yang, Y.P. Zhang, X.S. Gao, H.M. Kang, Effects of *Poncirus polyandra* sample size on genetic diversity index by SSR markers, J.Northeast Univ. 45 (2017), 35–39, 44.
[98] Y. Zhang, X. Gong, X.Y. Feng, Phylogenetic position of *Poncirus poluandra* by DNA sequencing, Guihaia 41 (2021) 114–122.
[99] Y.Y. Zhang, R.H. Ming, M. Khan, Y. Wang, B. Dahro, W. Xiao, C.L. Li, J.H. Liu, RF9 of Poncirus trifoliata (L.) Raf. Undergoes feedback regulation by ethylene and modulates cold tolerance via regulating a glutathione S-transferase U17 gene, Plant Biotechnol. J. 20 (2022) 183–200.
[100] R.F. Zhong, S.Y. Wang, M.Q. Liang, Y.M. Du, Z.Y. Duan, Studies on biological characteristics and application of *Poncirus polyandra*, J. Yunnan Agric. Univ. 16 (2001) 244–246.
[101] C. Zhu, X. Zheng, Y. Huang, J. Ye, P. Chen, C. Zhang, F. Zhao, Z. Xie, S. Zhang, N. Wang, H. Li, L. Wang, X. Tang, L. Chai, Q. Xu, X. Deng, Genome sequencing and CRISPR /Cas9 gene editing of an early flowering mini-Citrus (*Fortunella hindsii*), Plant Biotechnol. J. 17 (2019) 2199–2210.
[102] S. Zhu, F. Wang, W. Shen, D. Jiang, Q. Hong, X. Zhao, Genetic diversity of *Poncirus* and phylogenetic relationships with its relatives revealed by SSR and SNP/InDel markers, Acta Physiol. Plant. 37 (2015) 141.