Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Phylogenomic and syntenic data demonstrate complex evolutionary processes in early radiation of the rosids

Luxian Liu[1], Mengzhen Chen[1], Ryan A. Folk[2], Meizhen Wang[3], Tao Zhao[4], Fude Shang[1], Douglas Soltis[5], and Pan Li[3]

[1]Henan University
[2]Mississippi State University
[3]Zhejiang University
[4]Northwest A&F University
[5]University of Florida

March 11, 2023

Abstract

Some of the most vexing problems of deep level relationship that remain in angiosperms involve the superrosids. The superrosid clade contains a quarter of all angiosperm species, with 18 orders in three subclades (Vitales, Saxifragales and core rosids) exhibiting remarkable morphological and ecological diversity. To help resolve deep-level relationships, we constructed a high-quality chromosome-level genome assembly for Tiarella polyphylla (Saxifragaceae) thus providing broader genomic representation of Saxifragales. Whole genome microsynteny analysis of superrosids showed that Saxifragales shared more synteny clusters with core rosids than Vitales, further supporting Saxifragales as more closely related with core rosids. To resolve the ordinal phylogeny of superrosids, we screened 122 single copy nuclear genes from genomes of 36 species, representing all 18 superrosid orders. Vitales were recovered as sister to all other superrosids (Saxifragales + core rosids). Our data suggest dramatic differences in relationships compared to earlier studies within core rosids. Fabids should be restricted to the nitrogen-fixing clade, while Picramniales, the Celastrales-Malpighiales (CM) clade, Huerteales, Oxalidales, Sapindales, Malvales and Brassicales formed an "expanded" malvid clade. The Celastrales-Oxalidales-Malpighiales (COM) clade (sensu APG IV) was not monophyletic. Crossosomatales, Geraniales, Myrtales and Zygophyllales did not belong to either of our well-supported malvids or fabids. There is strong discordance between nuclear and plastid phylogenetic hypotheses for superrosid relationships; we show that this is best explained by a combination of incomplete lineage sorting and ancient reticulation.

Phylogenomic and syntenic data demonstrate complex evolutionary processes in early radiation of the rosids

Luxian Liu[1,2+], Mengzhen Chen[1+], Ryan A. Folk[3+], Meizhen Wang[2], Tao Zhao[4], Fude Shang[1,5] Douglas E. Soltis[6,7], and Pan Li[2*]

[1]Laboratory of Plant Germplasm and Genetic Engineering, School of Life Sciences, Henan University, Kaifeng, Henan, 475001, China

[2]Key Laboratory of Biosystems Homeostasis and Protection (Zhejiang University), Ministry of Education, Hangzhou, Zhejiang, 310058, China

[3]Department of Biological Sciences, Mississippi State University, Starkville, MS, United States

[4]State Key Laboratory of Crop Stress Biology for Arid Areas/Shaanxi Key Laboratory of Apple, College of Horticulture, Northwest A&F University, Yangling, Shaanxi, 712100, China

[5]Henan Engineering Research Center for Osmanthus Germplasm Innovation and Resource Utilization, Henan Agricultural University, Zhengzhou, Henan, 450002, China

[6]Florida Museum of Natural History, University of Florida, Gainesville, FL, 32611 United States

[7]Department of Biology, University of Florida, Gainesville, FL, 32611 United States

[+]These authors contributed equally to this work

[*]Corresponding author:

Pan Li (Email: panli_zju@126.com, Phone: +8613757152017)

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Abstract

Some of the most vexing problems of deep-level relationships in angiosperms involve superrosids. The superrosid clade contains a quarter of all angiosperm species, with 18 orders in three subclades (Vitales, Saxifragales, and core rosids) exhibiting remarkable morphological and ecological diversity. To help resolve deep-level relationships, we constructed a high-quality chromosome-level genome assembly for *Tiarella polyphylla* (Saxifragaceae), thereby providing a broader genomic representation of Saxifragales. Whole genome microarray analysis of superrosids showed that Saxifragales shared more synteny clusters with core rosids than Vitales, further supporting Saxifragales as being more closely related to core rosids. To resolve the ordinal phylogeny of superrosids, we screened 122 single-copy nuclear genes from the genomes of 36 species representing all 18 superrosid orders. Vitales were recovered as sisters to all other superrosids (Saxifragales + core rosids). Our data suggest dramatic differences in these relationships compared to earlier studies of core rosids. Fabids should be restricted to the nitrogen-fixing clade, while Picramniales, the Celastrales-Malpighiales (CM) clade, Huerteales, Oxalidales, Sapindales, Malvales, and Brassicales formed an "expanded" malvid clade. The Celastrales-Oxalidales-Malpighiales (COM) clade (sensu APG IV) was not monophyletic. Crossosomatales, Geraniales, Myrtales, and Zygophyllales did not belong to either our well-supported malvids or fabids.

There is a strong discordance between nuclear and plastid phylogenetic hypotheses for superrosid relationships, which can be best explained by a combination of incomplete lineage sorting and ancient reticulation.

.

Key words: genome assembly, *Tiarella polyphylla* , Angiosperm-mega 353, phylogeny, superrosids, ancient reticulation.

Introduction

The core eudicots consist of Gunnerales, Dilleniales, superrosids, and superasterids, with the latter two containing the vast majority of flowering plant diversity (**Drinnan et al., 1994; Soltis et al., 2018**). Superrosids, comprising core rosids (eurosids), Saxifragales, and Vitales, contain more than 90,000 species and thus represent more than a quarter of all angiosperms (**Wang et al., 2009; Sun et al., 2020**). Superrosid species exhibit remarkable morphological and ecological diversity and include herbs, shrubs, trees, vines, aquatics, succulents, and parasites (**Zhao et al., 2016**); Many important crops, as well as forest trees, are superrosids (**Wang et al., 2009**) including Rosales (e.g., apple, jujube, and mulberry), Vitales (grape), Cucurbitales (watermelon, cucumber), Fabales (peanut, soybean), Fagales (walnut, waxberry, oak), and Brassicales (radish, mustard, and cabbage). Several superrosid orders, such as Malvales, Myrtales, Cucurbitales, Fabales, Rosales, and Saxifragales, exhibit exceptionally high diversification rates among angiosperms (**Magallon & Sanderson, 2001; Folk et al., 2019; Sun et al., 2021**). The enormous diversity and ecological and economic importance of superrosid species highlights the importance of greater resolution in superrosid phylogeny.

The monophyly of superrosids has been recovered repeatedly in previous studies, with both organellar (**Moore et al., 2010; Sun et al., 2015; Li et al., 2019a**) and nuclear genes (**Zhang et al., 2012; One Thousand Plant Transcriptomes Initiative, 2019; Sun et al., 2021**), as well as combined datasets

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

(**Wang et al., 2009; Sun et al., 2020**). However, relationships within superrosids have proven more problematic. In APG IV (**2016)** , Saxifragales were sister to Vitales plus core rosids, a topology found in multiple phylogenetic studies of mostly plastid genes (e.g., **Wang et al., 2009; Soltis et al., 2011; Li et al., 2019a**). The core rosid clade, in turn, consisted of fabid and malvid subclades. The fabids contained the COM clade (Celastrales, Oxalidales, and Malpighiales), nitrogen-fixing clade (Fabales, Rosales, Cucurbitales, and Fagales), and Zygophyllales, which include Geraniales, Myrtales, Crossosomatales, Picramniales, Sapindales, Huerteales, Malvales, and Brassicales.

Although superrosids have long been the focus of phylogenetic research (**Wang et al., 2009; Soltis et al., 2011; Zhang et al., 2012; Li et al., 2019a; Sun et al., 2020**), relationships remain problematic, in part because of rapid radiation (**Wang et al., 2009)** combined with substantial recent evidence of incongruence between nuclear and plastid topologies (**Zhang et al., 2012; Li et al., 2019a; Sun et al., 2020**). Key problems in our understanding of relationships in superrosids remain:1) Are Saxifragales or Vitales the sister lineage of core rosids? 2) What are the major subclades within core rosids, and what orders should be included in fabids vs. malvids? 3) What are the relationships between COM clade members, and are they actually monophyletic? An improved nuclear-based phylogeny of superrosids and core rosids would help provide a better understanding of the evolutionary history of this enormous clade.

Previous phylogenetic studies of superrosids were primarily based on plastid and mitochondrial genes or relied on a small number of nuclear genes (**Wang et al., 2009; Moore et al., 2010; Zhang et al., 2012; Sun et al., 2016; Li et al., 2019a; Sun et al., 2020**), with a recent exception that includes numerous nuclear genes derived from transcriptomes (**One Thousand Plant Transcriptomes Initiative, 2019**). Organellar genomes (mitochondrial genomes and plastomes) are generally inherited uniparentally, and the mitochondrial genome is slowly evolving and sometimes affected by horizontal gene transfer, which introduces biases and errors in phylogenetic reconstruction (**Birky, 2001; Davis et al., 2014**); likewise The plastome is frequently transferred horizontally through introgression (**Okuyama et al., 2005; Stegemann et al., 2012**). In contrast, nuclear genes are inherited biparentally and show higher substitution rates than organellar genes, thereby overcoming many of these issues (**Springer et al., 2001; Davis et al., 2014**). In particular, low- or single-copy nuclear genes provide a crucial line of evidence for resolving angiosperm phylogeny (**Zeng et al., 2014; Zhang et al., 2020**), and the importance of using these genes for phylogenetic reconstruction has long been recognized (**Strand et al., 1997; Duarte et al., 2010; Zhang et al., 2012**). Therefore, the use of a sufficient number of single- or low-copy nuclear genes coupled with broad taxon sampling is a promising approach to elucidate angiosperm phylogeny (**Duarte et al., 2010; Soltis et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019**). In green plants, however, identifying orthologous loci has proven difficult because of frequent whole-genome duplication events, especially in angiosperms (**Blanc & Wolfe, 2004; Barker et al., 2009**). The increasing availability of genomic resources held in public repositories and the availability of many newly developed bioinformatic pipelines to identify low- or single-copy genes have enabled bait kit design for orthologous genes from a wide range of flowering plant groups (**Campana, 2018; Vatanparast et al., 2018; McLay et al., 2021**). Universal bait kits, such as Angiosperms353 loci used in this study, aim to capture the same set of loci from samples representing significant phylogenetic breadth and evolutionary timescales (**Bossert& Danforth, 2018; Johnson et al., 2019; Breinholt et al., 2021**). Currently, the Angiosperms353 probe set has been widely used to study the relationships between different groups (**Maurin et al., 2021; Thomas et al., 2021; Zuntini et al., 2021; Acha & Majure, 2022**).

Increasing amounts of genomic data have been sequentially applied to resolve rapid radiation in both green plant (**Carlsen et al., 2018; Rouard et al., 2018**) and animal (**Malinsky et al., 2018; Jensen et al., 2021**) lineages. Much of this work has used large numbers of coding regions extracted from genomes; however, chromosome-level genomes offer an additional path to assessing phylogenetic relationships via microsynteny, which is particularly valuable for resolving recalcitrant phylogenetic nodes (**Zhao et al., 2021**). A number of available genome assemblies have been published for Vitales (**Massonnet et al., 2020; Minio et al., 2022**), as well as for diverse families and orders of the core rosids (**Wang et al., 2021b; Wang et al., 2022a**), Rosales (**Jiao et al., 2020; Cao et al., 2022**), but few high-quality genomic

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

resources have been obtained for Saxifragales, preventing the use of this information to resolve phylogeny or understand genome evolution in the earliest radiation of the superrosids. Although small, Saxifragales are an ancient and morphologically diverse group (**Jian et al., 2008; Soltis et al., 2018**) with early and rapid radiation (~89.5 to 110 Ma) that has made resolving phylogenetic relationships challenging (**Fishbein et al., 2001; Wang et al., 2009; Jian et al., 2008; Dong et al., 2018; Folk et al., 2019**). For the 15 families of Saxifragales, seven whole-genome assemblies from four families are available: *Paeonia ostii* T. Hong and J. X. Zhang (**Yuan et al., 2022**), *Paeoniasuffruticosa* Andrews (Paeoniaceae, **Lv et al., 2020**),*Hamamelis virginiana* L. (Hamamelidaceae, **Korgaonkar et al., 2021**), *Cercidiphyllum japonicum* Siebold et Zucc. (Cercidiphyllaceae, **Zhu et al., 2020**), and three Crassulaceae species (*Kalanchoe fedtschenkoi* Raym.-Hamet et H. Perrier,**Yang et al., 2017** ; *Rhodiola crenulata* (Hook. f. et Thoms.) H. Ohba, **Fu et al., 2017** ; *Sedum album* L.,**Wai et al., 2019**). However, of these assembled genomes, only *C. japonicum* and *P. ostii*are assembled at the chromosomal level. To improve the genome resources for Saxifragales and provide genome-scale data needed for our analyses of relationships, we produced a chromosome-level genome assembly for*Tiarella polyphylla* D. Don (Saxifragaceae) (**Fig. 1-A**). This species has a wide distribution (**Wu & Raven, 2003**); it is an ideal model for use in future biogeographic studies as well as to investigate the features of Saxifragaceae (e.g., it is used in traditional medicine; **Lee et al., 2012; Kim et al., 2021**).

In this study we: (1) use gene sequence data for numerous nuclear loci representing all orders of superrosids to resolve relationships and evolutionary history; (2) constructed a high-quality chromosomal assembly reference genome for *T. polyphylla* to help elucidate evolutionary history; and (3) combined our newly generated complete genome and published complete nuclear genome sequences to conduct microsynteny analyses of superrosids to further resolve relationships.

Materials and methods

Genome sequencing, assembly and annotation

One living individual of *T. polyphylla* was collected from the Chongdugou scenic spot in Henan, China (111°39'41.64' 'E, 33°56'23.87 '' N) for whole genome sequencing. We sequenced and assembled the genome using a combination of Illumina short-read sequencing and Nanopore long-read sequencing. The completeness of the genome assembly was assessed with sets of both the Core Eukaryotic Genes Mapping Approach**(CEGMA; Parra et al., 2007)** and benchmarking universal single-copy orthologs **(BUSCO; Simao et al., 2015)** . For repetitive element annotation, simple sequence repeats (SSRs), tandem repeats and transposable elements (TEs) were identified in the *T. polyphylla* genome. We combined *de novo* , homology-based, and RNA sequencing-aided methods for gene prediction. For details, see Supporting Information Methods S1.

Hi-C library construction and chromosome assembly

To generate a chromosome-level assembly of the *T. polyphylla*genome, a Hi-C library was constructed following Rao's protocol (**Rao et al., 2014**). Fresh leaf cells were fixed in 1% formaldehyde for cross-linking. The cross-linked DNA was homogenized by tissue lysis, digested with *DpnII* restriction endonuclease, labelled with biotin-14-dCTP, and ligated using T4 DNA Ligase. After reversal of the cross-links, the ligated DNA was purified and sheared into 300–600 bp fragments. Biotinylated DNA fragments were extracted using streptavidin beads to construct the Hi-C fragment library. After PCR enrichment, high-quality libraries were sequenced on an Illumina NovaSeq 6000 platform to produce approximately 160.46 Gb data.

The cleaned Hi-C data were mapped to the initial genome assembly using BOWTIE2 v2.3.2 **(Langmead & Salzberg, 2012)** with the end-to-end model (-very-sensitive -L 30), and only unique mapped read pairs were retained in further analysis. Then, the valid mate pair reads were used for chromosome-level genome assembly, and the contigs of the draft genome were sorted, oriented, and divided into different chromosomal groups using the LACHESIS pipeline **(Burton et al., 2013)** with the following parameters: CLUSTER MIN RE SITES = 100, CLUSTER MAX LINK DENSITY = 2.5, CLUSTER NONINFORMATIVE RATIO = 1.4, ORDER MIN N RES IN TRUNK = 60, and ORDER MIN RES IN SHREDS = 60.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Whole genome duplication events of *T. polyphylla*

Three genomes were selected for comparison to investigate the whole genome duplication (WGD) history of *T. polyphylla* : *Vitis vinifera* (Vvi), *Cercidiphyllum japonicum* (Cja), and *Tiarella polyphylla* (Tpo). We identified paralogs (within Vvi, Cja, and Tpo) and orthologs (Tpo/Vvi and Tpo/Cja) using BLASTP (E-value [?] 1e-5). For each gene pair, the number of synonymous substitutions per synonymous site (*Ks*) was calculated using PAML v4.8 (**Yang, 2007**) using the YN00 NG model. MCScanX (**Wang et al., 2012**) was employed to identify syntenic blocks between genomes of Tpo, Vvi, and Cja based on the all-to-all BLASTP results, and a python version of MCScan was used to analyze the synteny (minspan = 100) to further detect whole genome duplication events.

Whole genome microsynteny of superrosids

Sixteen genomes, including *T. polyphylla* , representing 14 orders of superrosids, were used for microsynteny network construction (**Table S1**) using an approach described in detail previously (**Zhao et al., 2017** ; **Zhao et al., 2021**). Briefly, DIAMOND (v0.9.14.115) (**Buchfink et al., 2015**) was used for all pairwise intra- and inter-genome comparisons using all predicted protein sequences of each genome. Next, MCScanX (**Wang et al., 2012**) was used to detect all pairwise inter- and intra-synteny blocks under default settings. All synteny blocks were integrated into the total synteny network of syntenic genes. The Infomap algorithm (v0.20.0) (**Rosvall & Bergstrom, 2008**) was used for network clustering in the two-level partitioning mode with ten trials (-clu -N 10 -map -2). All synteny clusters identified were phylogenically profiled. A cluster profile recorded the number of nodes in a given cluster for each species. The collection of phylogenomic profiles (of all syntenic clusters) was summarized into a binary data matrix.

Identification of nuclear markers for phylogenetic analyses

To resolve the superrosid phylogeny, 34 transcriptomes and three nuclear genomic datasets were downloaded from GenBank and combined with the *T. polyphylla* sequences obtained in this study, representing all accepted orders of superrosids and outgroups (**Table S2**). Two species of Buxales (*Buxus semperivirens* and *Buxus sinica*var. *insularis*) were selected as outgroups (**Chanderbali et al., 2022**). HybPiper v1.2 (**Johnson et al., 2016**) was used to identify nuclear markers among the Angiosperm-mega 353 gene set (**McLay et al., 2021**) from all the datasets. The identified contigs matching probe can be extract using the following command line ". /reads_first.py -b mega353.fasta -r sample_R1.fastq sample_R2.fastq -prefix sample_-result -bwa", and we selected the genes commonly shared in 38 samples to construct the phylogenetic tree.

Phylogenetic inference

We used both coalescent and concatenation-based methods to reconstruct the phylogenetic trees. We first estimated individual gene trees using IQ-TREE v1.6.12 (**Minh et al., 2020**); ModelFinder (**Kalyaanamoorthy et al., 2017**) implemented in IQ-TREE enables a free-rate variation model for each partition alignment. The individual gene trees inferred by IQ-TREE were used as input in ASTRAL-III (**Zhang et al., 2018**) with default parameters to show the local posterior probabilities (LPPs).

A concatenated phylogeny was inferred using maximum likelihood (ML) and Bayesian inference (BI) analyses. ML analysis was performed using the CIPRES Science Gateway v3.3 (https://www.phylo.org/portal2) **(Miller et al., 2010**) and RAxML v8.1.11 (**Stamatakis et al., 2008**). One thousand rapid bootstrap iterations were used, and the default settings were used for the other parameters. BI analysis was carried out using MrBayes v3.2.3 (**Ronquist & Huelsenbeck, 2003**), and the posterior probability was estimated using four chains running 5,000,000 generations sampling every 1,000 generations. Convergence of the MCMC chains was assumed when the average standard deviation of split frequencies reached 0.01 or less, and the first 25% of the sampled trees were considered burn-in trees.

To investigate the gene tree expectations under coalescence, which can be used to gain insight into whether topological features are attributable to ILS (incomplete lineage sorting) or hybridization (reviewed in **Folk et al., 2018**), we used a previously described simulation pipeline (**Folk et al., 2017** ; **Garcia et al. 2017** ; https://github.com/ryanafolk/tree_utilities). Briefly, assuming a species tree with estimated branch length

measured in coalescent units, 1000 gene tree histories were simulated and clade probabilities were calculated for all observed relationships in the species tree, where $p \sim 0$ indicates a relationship not expected under ILS alone. Where the ILS is high, many relationships could be low in probability, so two further probabilistic tests were implemented. First, a significance test was conducted based on comparing the complete set of all pairwise Robinson-Foulds distances (1) among simulated gene trees and (2) between simulated and empirical gene trees, where a significantly higher empirical distance would suggest discord not predicted by the ILS, indicative of potential hybridization. Second, clade probabilities enumerated from the simulated gene tree set were compared to clade probabilities in the empirical gene set. Similarly, significantly lower clade probabilities in the empirical gene set are suggestive of potential hybridization. Both analyses were implemented as one-tailed t -tests.

Results

Genome sequencing and assembly

The diploid *Tiarella polyphylla* ($2n = 2x = 14$;**Fig. 1-B, C**) genome size was estimated to be 393.29 Mb based on the total number of 20,057,799,129 21-mer and a peak 21-mer depth of 51, and the estimated heterozygosity rate was approximately 0.273% (**Fig. S1; Table S3**). A total of 65.3 Gb of ONT (Oxford Nanopore Technology) reads was produced with an N50 of 30.0 kb from one PromethION R9.4.1 flow cell. The longest ONT read was 219.8 kb, and the genome coverage was c. x162 (**Table S4**). We trimmed the raw reads using the CANU software, and the corrected reads were assembled for resulting in an initial assembly with a genome size of ~ 404.3 Mb and a contig N50 of ~11.5 Mb (**Table 1**). After polishing using NextPolish, we retrieved a corrected genome with size 418.1 Mb and contig N50 12.0 Mb. Finally, after removing sequences originating from the plastome, the mitochondrial genome and bacteria, the de novo genome assembly was 412.2 Mb in size, with contig N50, longest contig, and contig number of 12.0 Mb, 26.3 Mb and 206, respectively.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

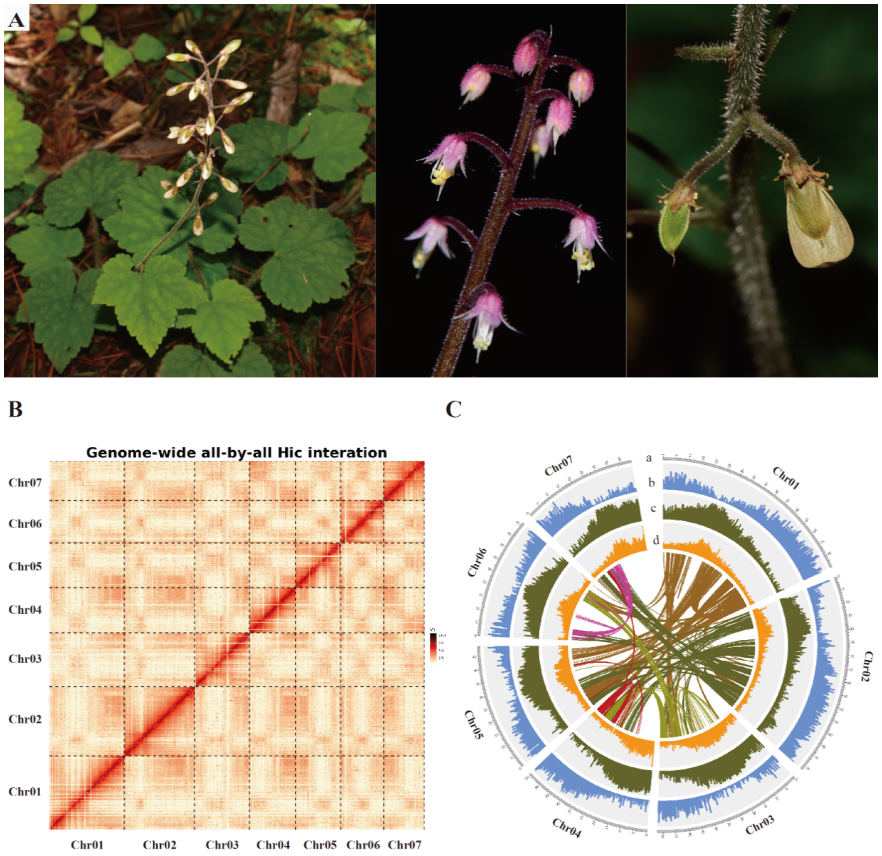

Fig. 1 Characterization of *Tiarella polyphylla* . (A) Whole plant, flowers, and fruits of *T. polyphylla* (photos by Pan Li). (B) Hi-C interaction heat map between 7 chromosomes for the *T. polyphylla*genome. (C) Characterization and synteny of the *T. polyphylla*genome. Circles from the outside inwards: (a) pseudochromosomes, (b) gene density, (c) repeat density and (d) GC content. The density was calculated with 500 kb sliding windows.

Table 1. Statistics of the *Tiarella polyphylla* genome assembly

Genome assembly of *T. polyphylla* (2*n* = 14)	Nanopore	Nanopore	Nanopore	Nanopore+Hi-C
	SMARTdenovo	NextPolish	blastn	LACHESIS
Total assembly size of contigs (bp)	404,318,547	418,052,640	412,241,708	-
Number of contigs	247	247	206	-
N50 contig length (bp)	11,502,989	12,010,362	12,010,362	-
N90 contig length (bp)	899,560	932,839	968,152	-

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Genome assembly of *T. polyphylla* (2*n* = 14)	Nanopore	Nanopore	Nanopore	Nanopore+Hi-C
Longest contig (bp)	25,752,925	26,317,235	26,317,235	-
Total assembly size of scaffolds (bp)	-	-	-	403,101,895
Number of scaffolds	-	-	-	160
N50 scaffold length (bp)	-	-	-	57,234,420
N90 scaffold length (bp)	-	-	-	44,079,820

Note: N50, shortest sequence length at 50% of the genome; N90, shortest sequence length at 90% of the genome. The dashed line indicates data not available.

The completeness of genome assembly has been validated using various approaches. More than 97.31% of the complete single-copy BUSCOs were found in the genome assembly, and only 2.18% of the BUSCOs were missing (**Fig. S2** ; **Table S5**). CEGMA assessment retrieved 241 (97.18%) (**Table S6**) of the 248 core eukaryotic genes (CEGs). Furthermore, Illumina short reads (65.3 Gb) were aligned to the assembled genome using BWA software, with a mapping efficiency of ~ 98.15% and coverage percentage of ~ 95.44%, suggesting a high consistency between Illumina reads and the assembled genome (**Table S7**). Together, these results show that the assembled *T. polyphylla* genome sequence was complete and had a low error ratio.

Chromosome level assembly of Hi-C data

We generated 171.54 Gb of raw Hi-C data, consisting of 1,148,659,116 paired-end reads (**Table S8**). After quality control, 169.83 Gb of clean data remained, containing 99.00% clean paired-end reads, which were used as input for the BOWTIE2 and LACHESIS Hi-C analysis pipelines. Finally, 160 scaffolds (representing 97.78% of the total genome length) were anchored to the seven chromosomes of *T. polyphylla* with a length of 403.10 Mb (**Fig. 1; Table 1**). The lengths of the chromosomes ranged from 44.08 Mb to 79.84 Mb with a scaffold N50 of 57.23 Mb (**Table S9**).

Genome annotation

Repeat sequences, accounting for 60.10% of the genome, were identified based on the assembled sequence of the *T. polyphylla* genome (**Table S10**). Of these, SSRs accounted for 0.18% of the repeat fraction, including 44,893 di-, 7,943 tri-, and 856 tetra-nucleotide repeats (**Table S10-S11**). We also identified 34,470 tandem repeats containing 2.39 Mb sequences, accounting for 0.58% of the *T. polyphylla* genome (**Table S11**). Overall, the combined results of the *de novo* and homology-based methods revealed that 57.29% of the *T. polyphylla*genome contained TEs, of which Class I (retrotransposons) and Class II (DNA transposons) comprised 49.57% and 7.71% of the genome, respectively (**Table S11**). Of these, long terminal repeat (LTR) retrotransposons constituted the predominant repeat element in the genome, accounting for 45.02%. Further examination showed that two types of LTRs, Gypsy and Copia, occupied 25.39% and 4.07% of the genome sequences, respectively.

We identified 25,319 protein-coding genes in the *T. polyphylla*genome (**Table S12, Table S15**), with average gene length, coding sequence length, and exon length estimated as 4192.8 bp, 1221.8 bp and 227.8

bp respectively, and the average exon number per gene was 5.36 (**Fig. S3**). In total, 23,041 genes were annotated in at least one of the five databases, accounting for 91% of the total genes (**Fig. S4; Table S13**). In addition to protein-coding genes, various non-coding RNA sequences were identified and annotated (**Table S14**), including 703 transfer RNAs, 607 ribosomal RNAs, 90 microRNAs, and 220 small nuclear RNAs.

Whole genome duplication events of *T. polyphylla*

Microsynteny analysis revealed that a typical ancestral region in the *T. polyphylla* genome could be linearly connected to one region in the *Vitis vinifera* and *Cercidiphyllum japonicum* genomes (**Fig. 2A**). Syntenic depth analyses showed that 45% of the *T. polyphylla* blocks were covered by one Tpo-Vvi block, 2% were covered by two Tpo-Vvi blocks, 51% of the *V. vinifera* blocks were covered by one Vvi-Tpo block, and 3% were covered by two Vvi-Tpo blocks. Similarly, we found that 62% of the *T. polyphylla* blocks were covered by one Tpo-Cja block, 2% were covered in two Tpo-Cja blocks, and 67% of the *C. japonicum* blocks were covered by one Cja-Tpo block and 1% were covered in two Cja-Tpo blocks (**Fig. 2B**). These results suggest a 1:1 syntenic depth pattern for *T. polyphylla* versus *V. vinifera* and *T. polyphylla* versus *C. japonicum* . Moreover, the K_s distribution of paralogs also showed only one dominant peak for *T. polyphylla* (1.39; **Fig. 2C**), which was the same as *V. vinifera* (1.12-1.40) and *C. japonicum* (0.79). The peak of *T. polyphylla* occurred before the divergence peak at 0.77 between Tpo and Vvi, and was earlier than the speciation peak at 0.46 between Tpo and Cja, confirming that the WGD event occurred in the ancestor of the three species. Hence, these results strongly suggest that, as expected, *T. polyphylla* experienced the same gamma WGD event as *V. vinifera* and *C. japonicum* .

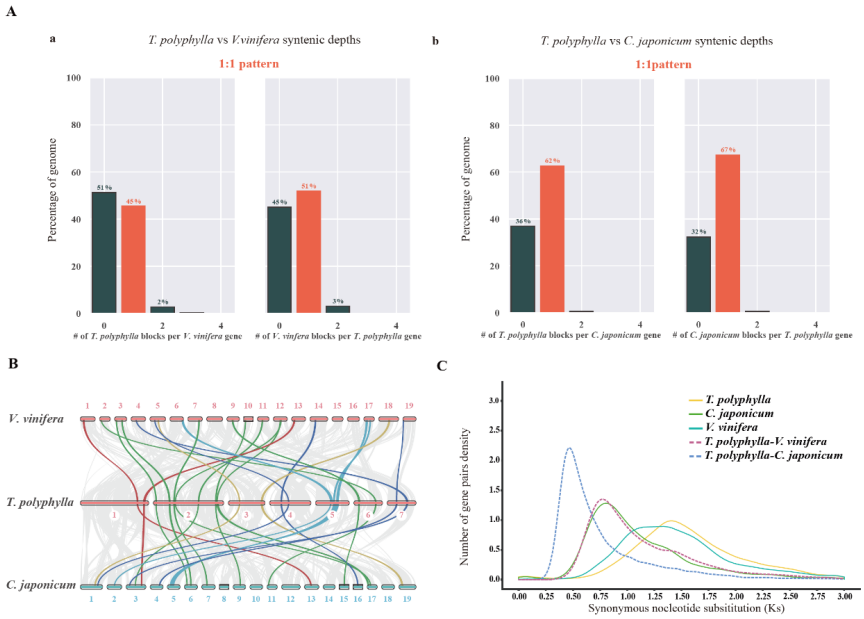

Fig. 2 Genome duplication in *Tiarella polyphylla* . (A) Syntenic depth pattern between (a) *T. polyphylla* vs *V. vinifera* and (b) *T. polyphylla* vs *C. japonicum* . (B) Macrosynteny patterns showing that a typical region in *T. polyphylla* can be traced to no more than one region in *V. vinifera* and *C. japonicum* respectively. (C) Distribution of synonymous nucleotide substitutions (*Ks*) among *T. polyphylla* , *V. vinifera* and *C. japonicum* .

Whole genome microsynteny of superrosids

The size of the matrix obtained from the microsynteny network construction was 16 × 21,326, which contained

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

a binary presence/absence coding for each cluster in the synteny network (**Table S16**). A total of 15,413 and 15,119 synteny clusters were detected in *Vitis riparia* and *V. vinifera* , while the numbers of synteny clusters were detected in *Cercidiphyllum japonicum* and *Tiarella polyphylla* were 13,537 and 12,728 respectively (Fig. 3A). For the remaining 12 species, the number of synteny clusters varied significantly, ranging from 6,131 in *Hibiscus cannabinus* to 13,497 in *Juglans regia* . Moreover, dividing the above 16 species into three major groups, namely Vitales, Saxifragales, and core rosids (Fig. 3B), 14,675, 15,833, and 21,326 synteny clusters were detected in Saxifragales, Vitales, and core rosids, respectively, and 12,870 synteny clusters were shared among the three major groups. Interestingly, we found that the number of synteny clusters shared between Saxifragales and core rosids (1,433) was greater than that shared by Vitales and core rosids (990) or Vitales and Saxifragales (170) apart from the 12,870 synteny clusters.

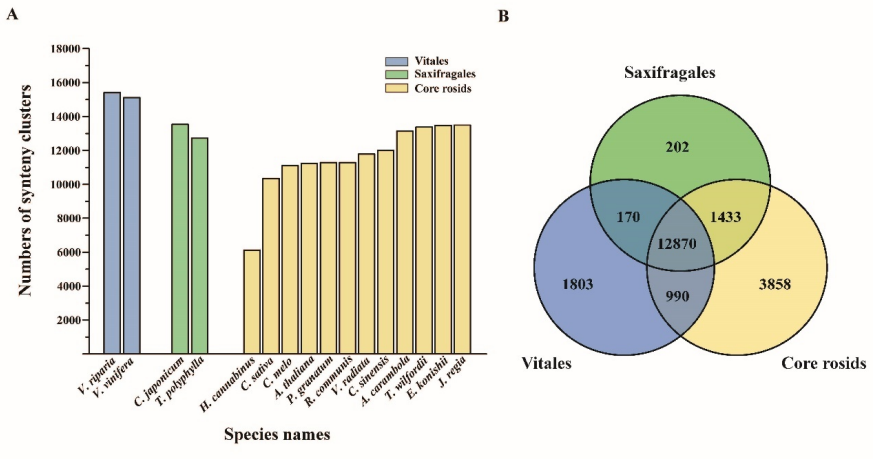

Fig. 3 Whole genome comparisons and microsynteny cluster detection. (A) The number of synteny clusters detected in the whole genome of the 16 studied species; (B) The number of synteny clusters recovered in Vitales, Saxifragales and core rosids, and the detail of the shared synteny clusters among the three groups.

Phylogenetic relationships within the superrosids

Gene recovery was successful for all species based on the Angiosperm-mega 353 gene set, with gene recovery rates of at least 91.22%, and the recovered gene number ranged from 322 genes recovered in *Carica papaya* to 350 genes in *Gerrardina foliosa*(**Fig. 4, Table S17**), resulting in a total of 166 putative single-copy nuclear genes shared among 38 species. Owing to the short length of some recovered genes, 44 genes with a gene length less than half of the target gene were removed, and the remaining 122 putative single-copy nuclear genes were used for phylogenetic inference.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

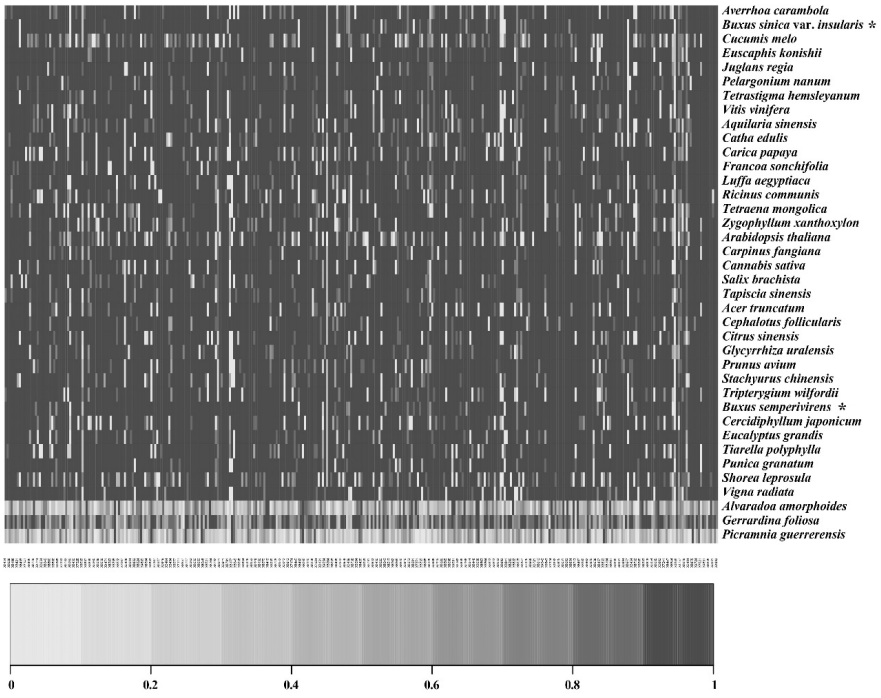

Fig. 4 Heatmap of gene recovery efficiency. Each row represents 36 superrosid species and two outgroups, and each column represents one target gene. Shading indicates the percentage of the target length recovered. Asterisk indicates the two outgroup species.

The phylogeny we obtained for superrosids was identical in both the ML and BI analyses, and we reconstructed a robust (ML bootstrap support (BS)/BI posterior probability (PP); BS/PP = 100/1) phylogeny for superrosids, with Vitales sister to Saxifragales plus the core rosids (**Fig. 5**). Saxifragales was strongly supported (BS/PP = 100/1) as a sister to the core rosids. Within the core rosids, the Geraniales + Crossosomatales clade (BS/PP = 60/1), followed by a clade of Zygophyllales + Myrtales (BS/PP = 100/1), were sisters to the remaining core rosids. The remaining core rosids comprised two major subclades, which we are referring to here as "fabids" and "malvids," although these differ from the circumscription given in APG IV; both clades as defined here received maximal BS support. Here, the fabids comprised only four orders known as the nitrogen-fixing clade, within which Fagales (BS/PP = 100/1) and Rosales (BS/PP = 100/1) were subsequent sisters to Fabales + Cucurbitales. Among the malvids defined here, Picramniales was sister to the remaining members. COM clade orders (Celastrales, Oxalidales, and Malpighiales) did not form a monophyletic group in our analyses. Celastrales and Malpighiales grouped together (BS/PP = 100/1) as sisters to the remaining orders (Huerteales, Oxalidales, Sapindales, Malvales, and Brassicales). For the remaining five orders, Huerteales was then resolved as sister to the other four orders with maximum support (BS/PP = 100/1); Oxalidales and Sapindales were subsequently recovered as successive sisters to Malvales + Brassicales with strong support (BS/PP = 98/1).

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

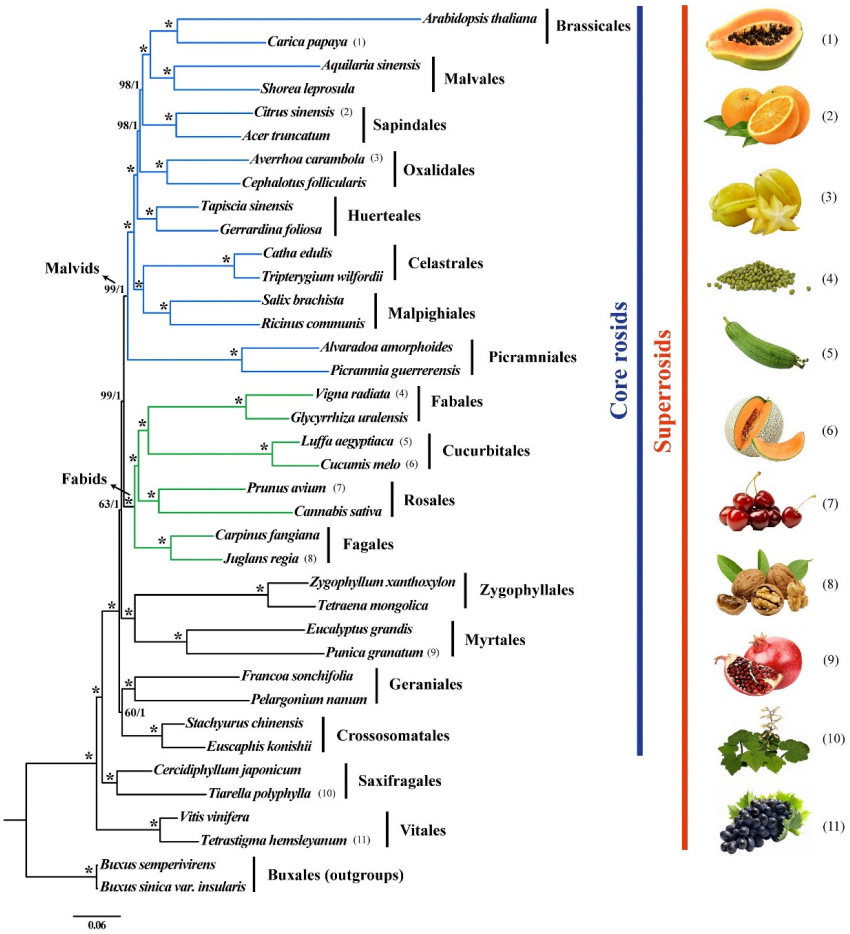

Fig. 5 Phylogenomic trees estimated using RAxML from 122 concatenated nuclear single-copy genes. Numbers above the branches represent ML bootstrap/Bayesian posterior probability (BS/PP), and "*" indicates that the clade is supported by BS/PP value 100/1. Numbers next to the species names correspond to the numbered plant photographs at the right. Fabids as defined here are shown with green branches; the expanded malvid clade is shown with blue branches.

Under the coalescence method, some backbone nodes had lower LPP compared to the two concatenation analyses, but the three analyses were topologically highly similar (**Fig. 6**). Vitales was sister to Saxifragales + core rosids (LPP = 0.77). The phylogenetic positions of COM clade members, Picramniales, Huerteales, Zygophyllales, and Myrtales were completely consistent in the coalescence and concatenation trees; incongruences between the two approaches were mainly observed for the positions of Crossosomatales and Geraniales. In the coalescence phylogeny, after the successive branching of Vitales and Saxifragales, Crossosomatales was sister to all other core rosids with maximum local branch support (LPP = 1.0). Subsequently, Geraniales was sister to a clade comprising Zygophyllales, Myrtales, fabids, and malvids, with moderate support (LPP = 0.55). Fabids comprised nitrogen-fixing clade orders (LPP = 1), and malvids consisted of Picramniales, Malpighiales, Celastrales, Huerteales, Brassicales, Malvales, Oxalidales, and Sapindales (LPP = 0.99).

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

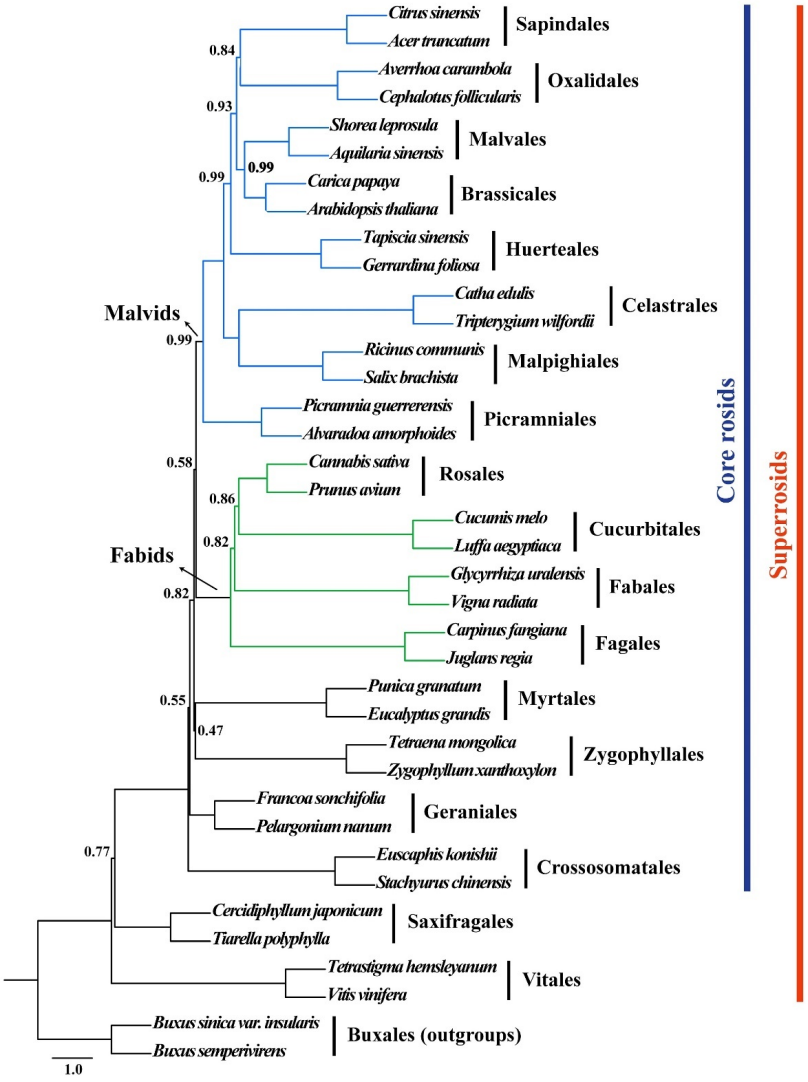

Fig. 6 Phylogenomic trees estimated using ASTRAL based on 122 nuclear single-copy genes. Branch support was calculated using local posterior probabilities (LPP), and the LPP value less than 1.0 are shown on branches. Fabids as defined here are shown with green branches; the expanded malvid clade is shown with blue branches.

Coalescent simulations demonstrated a significant role of the ILS in the backbone of the superrosids (**Fig. S5**). The branch subtending (core rosids + Saxifragales), the critical node for examining the relative placement of Vitales versus Saxifragales with respect to the core rosids, had a clade probability of 0.31. This value is close to the theoretical minimum probability (0.33 for the species tree clade under the ILS in the three-taxon case (**Pamilo & Nei, 1988**), suggesting that this branch is in the anomaly zone (**Degnan & Rosenberg, 2006**). Therefore, these results are consistent with the primary role of the ILS in conflicting interpretations of superrosid relationships. However, the expected clade probabilities were lower ([?] 0.1) in the backbone of the rosids, suggesting high discord compared to ILS expectations. Statistical tests suggest that ILS alone cannot explain gene tree incongruence; observed gene tree Robinson-Foulds distances were significantly higher than expected ($p <$ 1e-20), and the observed gene tree clade probabilities were lower than the ILS expectation ($p = 0.0043$). These results are therefore consistent with a potential role for early

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

reticulation in the radiation of rosids.

Discussion

Comparative genomics of Saxifragales

Although seven species from four families of Saxifragales have published whole-genome sequence assemblies (**Fu et al., 2017; Yang et al., 2017** ; **Wai et al., 2019** ; **Lv et al., 2020; Zhu et al., 2020; Korgaonkar et al., 2021**), only those of *Paeonia ostii* and *Cercidiphyllum japonicum* were assembled at the chromosomal level. Here, we successfully assembled the genome of*Tiarella polyphylla* , which is the first chromosome-level sequencing, assembly, and annotation of the genome of Saxifragaceae. The final size is 412.2 Mb, similar to the estimated genome size (393.29 Mb) based on k-mer analysis. It is much smaller than estimates reported in other genera of Saxifragaceae tribe Heuchereae (e.g., *Mitella diphylla* , 1C = 0.57 pg; **Bai et al., 2012** ; *Heuchera cylindrica,* 1C = 0.48–0.52 pg; **Godsoe et al., 2013**). Our results demonstrated that *T. polyphylla* , *C. japonicum* and*Hamamelis virginiana* had no additional genome duplication after the ancestral gamma hexaploidization event shared with all core eudicots. These three genera span Saxifragales, demonstrating that there is no ancestral genome-wide duplication characteristic of the entire order. More details on comparative genomics and whole-genome duplication are available in Discussion S1.

Vitales sister to core rosids + Saxifragales

Our whole-genome microarray analysis of superrosids showed that Saxifragales shared more synteny clusters with core rosids than Vitales (**Fig. 3**), suggesting that it has a closer relationship with core rosids. Similarly, Vitales and Saxifragales were supported as successive sisters to the core rosids based on the phylogenetic analysis of 122 nuclear putative single-copy genes with BS = 100% and PP = 1 in the concatenation tree (**Fig. 5**). The same position of Vitales was also reconstructed using the coalescent method with a 1.0 LPP support (**Fig. 6**). These results agree with the coalescent analysis in the 1KP study (**One Thousand Plant Transcriptomes Initiative, 2019**) and the coalescent and concatenation analyses in **Zeng et al. (2017)** , both of which relied on nuclear gene sequence data but do not agree with most other studies that largely relied on plastid gene sequences. While Vitales and Saxifragales are generally found to be closely associated with core rosids, relationships among them are uncertain, and all three possible relationships have been recovered (**APG III, 2009; Moore et al., 2010**). The consensus relationship in APG IV **(2016)** , which is different from that reported here, considers Saxifragales as sister to the remaining superrosids (Vitales + core rosids). For example,**Wang et al. (2009)** recovered the relationship of Saxifragales as sister to Vitales + core rosids (BS = 72%) using numerous genes (primarily plastid), a topology that has also been recovered in subsequent studies based on different molecular markers, for example, 17 loci including 11 plastid, two nuclear and four mitochondrial genes (BS = 85%, **Soltis et al., 2011**), four loci including one mitochondrial and three plastid genes (BS = 97%, **Sun et al., 2016**), four mitochondrial genes (BS = 13%, **Sun et al., 2015**), or even the entire protein-coding and rRNA genes of the plastid genome (BS = 60%, **Li et al., 2019a**). In other studies, Saxifragales plus Vitales occasionally formed a clade sister to core rosids based on plastid phylogenomic data (BS = 82%, **Moore et al., 2010;** BS = 73%, **Sun et al., 2015;** BS/PP = 91/0.99, **Zhang et al., 2016**), or sister to a clade comprising Caryophyllales and the remaining rosid species based on 5-gene nuclear sequences (PP=1, **Zhang et al., 2012** ; BS = 96%, **Sun et al., 2015**). The topology of Vitales sister to Saxifragales + core rosids was first reported by**Moore et al. (2011)** based on the plastid inverted repeat region. The same topology was also retrieved with strong support by**Shi et al. (2020)** using 44 plastid genes (PP = 1) as well as**Zeng et al. (2017)** and **Wang et al. (2022b)** based on numerous nuclear genes.

Major changes in circumscriptions of fabids and malvids

In most studies using organellar genes, core rosids consisted of two major groups: fabids and malvids. In our results, we recovered two major core rosid clades, but these differed in composition from those reported to date (**Fig. 5, Fig. 6**). The fabid clade consisted only of the nitrogen-fixing clade (Cucurbitales, Fagales, Fabales, and Rosales). Meanwhile, Picramniales, the CM clade, Huerteales, Oxalidales, Sapindales, Malvales, and Brassicales composed the "expanded" malvids. The remaining four core rosid orders (Geraniales,

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Crossosomatales, Zygophyllales, and Myrtales) were recovered in more early diverging positions and were not placed in either malvids or fabids, a relationship first reported by **Qiu et al. (2010)** . In our concatenation tree, Geraniales and Crossosomatales formed a strongly supported clade, Zygophyllales was sister to Myrtales with strong support, and these two clades were subsequently successive sisters to the fabids-malvids clade (**Fig. 5**). However, Geraniales and Crossosomatales did not form a clade in our coalescence analysis, although the support for relationships was weak in this part of the tree (**Fig. 6**). **Zhao et al. (2016)** recovered a similar rosid topology with fewer taxa based on 891 clusters of putative orthologous genes, except for the position of Zygophyllales; however, most studies did not recover early diverging positions for the four rosid orders. Missing data may be an important cause of inconsistent topologies because of their presence in previous nuclear phylogenomic studies (**Kvist & Siddall, 2013; Roure et al., 2013**). Further literature reviews concerning Picramniales and Huerteales are available in Discussion section S1.

The COM clade is non-monophyletic; constituent families should be placed in malvids

The COM clade, as circumscribed by two studies (**Matthews &Endress, 2006; Zhu et al., 2007**), contains approximately 19,000 species, or approximately one-fifth of all superrosids (**APG IV, 2016**). Despite rapid progress in elucidating the major branches of superrosid phylogeny, the position of the COM clade has been a subject of much debate. We found that the COM clade was non-monophyletic, and the three constituent families appeared with malvids based on coalescent and concatenation-based methods (**Fig. 5, 6**). The sister relationship between Celastrales and Malpighiales was strongly supported, and Oxalidales was sister to a clade comprising Sapindales, Malvales, and Brassicales with strong support in concatenation phylogeny, which was also reported by **Zhao et al. (2016)** . In our coalescent tree, Celastrales and Malpighiales were also sisters with strong support, and Oxalidales and Sapindales formed a clade with LPP = 0.89. Thus, the CM clade and Oxalidales should be members of malvids rather than fabids, which is also supported by floral features shared between the COM orders and malvids (e.g., the inner integument of the ovule, contorted petals, **Matthews & Endress, 2006**).

Cyto-nuclear discordance and ancient reticulation

The conflict between plastid trees and inferences from the nuclear genome goes back to the earliest studies on plastid phylogenetics (**Palmer et al., 1982**). Such conflict subsequently appeared in many early plastid restriction site analyses, in which one or more individuals of one species were nested within plastid-based clades of another species (**Rieseberg et al., 1991**). These results demonstrate the high frequency of cytoplasmic gene flow in angiosperms, as well as its extent within certain lineages (**Rieseberg & Soltis 1991**). Accordingly, the deep discordance between plastid and nuclear trees might be interpreted as evidence of ancient hybridization, given the propensity for interspecific hybridization among extant species (**Gitzendanner et al., 2018**), including more than hundred records of interspecific hybridization among rosid taxa alone (**Rieseberg & Soltis 1991; Rieseberg et al., 1996**).

Although conflict between plastid and nuclear trees is typically attributed to hybridization, other processes such as incomplete lineage sorting (ILS) may also cause phylogenetic incongruence between nuclear and plastid DNA (**Soltis & Kuzoff, 1995; Wendel & Doyle, 1998**). While hybridization and ILS were historically difficult to distinguish due to their similar phylogenetic signatures (**Wendel & Doyle, 1998**), the multispecies coalescence (**Mirarab et al., 2014; Mirarab & Warnow, 2015**) offers a clear path for testing the relative roles of ancient hybridization and ILS in explaining gene tree congruence (**Folk et al. 2018**). **Sun et al. (2015)**proposed that the incongruence in the positions of the COM orders between studies based on plastid, mitochondrial, and nuclear genes was possibly the result of ancient hybridization and introgression events. This hypothesis requires further study with probabilistic methods and larger samples of the three genomic compartments of plants.

Here, we retrieved different relationships of superrosids based on 122 single-copy nuclear genes compared to plastid genes, especially for the placement of Vitales and Saxifragales with respect to the rosids, non-monophyly of the COM clade, and the re-circumscription of rosids and fabids seen here. As reviewed above and in Discussion S1, while methodological decisions and data properties influence recovered topologies, the

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

primary cause for differing deep relationships among superrosids appears to rely on different studies of either cytoplasmic or nuclear markers. Although cytonuclear discordance is often attributed to hybridization, our simulation results suggest a role for both ILS and hybridization. The pattern of gene tree discord between the three major superrosid lineages (Vitales, Saxifragales, and rosids) was within ILS expectations and consistent with this branch of the tree being in the anomaly zone. However, the degree of gene tree heterogeneity related to backbone relationships within the core rosids was unexpected based on ILS alone, and is therefore likely due to ancient hybridization. These results contribute to a growing awareness that complex evolutionary processes should be considered, even for deep-level plant phylogenetics (**Folk et al. 2018** ; **Stull et al. 2022**).

Conclusion

We successfully assembled the genome of *Tiarella polyphylla* and reported the first chromosome-level assembly of Saxifragales. We leveraged this genome to generate a large nuclear gene dataset covering all superrosid orders, as well as microsynteny data from complete genome assemblies to resolve relationships. We provided strong support for Vitales as a sister to core rosids and Saxifragales. We also resolved the relationships within the core rosids, demonstrating new circumstances for fabids and malvids. There is a strong discordance between nuclear and plastid phylogenetic hypotheses for superrosid relationships, and our work demonstrates that this is best explained by a combination of incomplete lineage sorting and ancient reticulation.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (Grant Nos. 31900188 and 31970225) and Zhejiang Provincial Natural Science Foundation, China (Grant No. LY19C030022), and the Henan Province Major Research Fund of Public Welfare (Grant No. 201300110900).

Conflict of Interest

The authors declare that they have no competing interests.

Author Contributions

P.L., D.E.S. and F.D.S. designed and supervised the study. L.X.L. and M.Z.C collected the samples and extracted genomic DNA. L.X.L., M.Z.C., R.A.F., M.Z.W., and T.Z. performed genome assembly and data analyses. L.X.L., M.Z.C., and P.L. prepared the manuscript. R.A.F. and D.E.S. revised the manuscript. All the authors approved the final manuscript.

Data Availability Statement

The raw sequencing reads for *T. polyphylla* genome assembly have been deposited in the NCBI database with BioProject ID PRJNA870970 and the Nanopore, Illumina, and Hi-C raw data are available under accession nos. SRR21133002, SRR21133003, and SRR21132998, and RNA-seq data (roots, stems, leaves, and PacBio) are available under accession nos. SRR21132999, SRR21133001, SRR21133000, SRR21158993. The final assembly genome at contig level, chromosome level, and genome annotation information has been deposited in the National Genomics Data Center under accession no. PRJCA011134. The alignment files for the gene tree and species tree construction are available at https://doi.org/10.5061/dryad.fj6q573z3.

References

Acha S, Majure LC. 2022. A new approach using targeted sequence capture for phylogenomic studies across Cactaceae. *Genes* **13** : 350.

APG III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III.*Botanical Journal of the Linnean Society* **161** : 105–121.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV.*Botanical Journal of the Linnean Society* , **181** : 1–20.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Bai C, Alverson WS, Follansbee A, Waller DM. 2012. New reports of nuclear DNA content for 407 vascular plant taxa from the United States. *Annals of Botany* **110** : 1623-1629.

Barker MS, Vogel H, Schranz ME. 2009. Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* **1** : 391–399.

Birky Jr CW. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual review of genetics* **35** : 125–148.

Blanc G, Wolfe KH. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution.*The Plant Cell* **16** : 1679–1691.

Bossert S, Danforth BN. 2018. On the universality of target-enrichment baits for phylogenomic research. *Methods in Ecology and Evolution* **9** : 1453–1460.

Breinholt JW, Carey SB, Tiley GP, Davis EC, Endara L, McDaniel SF, Neves LG, Sessa EB, von Konrat M, Chantanaorrapint S *et al.* 2021. A target enrichment probe set for resolving the flagellate land plant tree of life. *Applications in Plant Sciences* **9** : e11406.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND.*Nature methods* **12** : 59–60.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature biotechnology* **31** : 1119–1125.

Campana MG. 2018. BaitsTools: Software for hybridization capture bait design. *Molecular ecology resources* **18** : 356–361.

Cao K, Peng Z, Zhao X, Li Y, Liu K, Arus P, Fang W, Chen C, Wang X, Wu J. 2022. Chromosome-level genome assemblies of four wild peach species provide insights into genome evolution and genetic basis of stress resistance. *BMC biology* **20** : 1-17.

Carlsen MM, Fér T, Schmickl R, Leong-Škorničková J, Newman M, Kress WJ. 2018. Resolving the rapid plant radiation of early diverging lineages in the tropical Zingiberales: pushing the limits of genomic data. *Molecular phylogenetics and evolution* **128** : 55–68.

Chanderbali AS, Jin L, Xu Q, Zhang Y, Zhang J, Jian S, Carroll E, Sankoff D, Albert VA, Howarth DG. 2022. *Buxus* and*Tetracentron* genomes help resolve eudicot genome history.*Nature Communications* **13** : 1-10.

Davis CC, Xi Z, Mathews S. 2014. Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. *BMC biology* **12** : 1–4.

Degnan JH, Rosenberg NA. 2006. Discordance of species trees with their most likely gene trees. *PLoS genetics* **2** : e68.

Dong W, Xu C, Wu P, Cheng T, Yu J, Zhou S, Hong D–Y. 2018. Resolving the systematic positions of enigmatic taxa: manipulating the chloroplast genome data of Saxifragales. *Molecular phylogenetics and evolution* **126** : 321–330.

Drinnan AN, Crane PR, Hoot SB. 1994. Patterns of floral evolution in the early diversification of non-magnoliid dicotyledons (eudicots). *Plant Systematics and Evolution* (Suppl.) **8** : 93–122.

Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires PK, Leebens-Mack J, Depamphilis CW. 2010. Identification of shared single copy nuclear genes in *Arabidopsis* , *Populus* , *Vitis*and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC evolutionary biology* **10** : 1–18.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Fishbein M, Hibsch-Jetter C, Soltis DE, Hufford L. 2001. Phylogeny of Saxifragales (angiosperms, eudicots): analysis of a rapid, ancient radiation. *Systematic Biology* **50** : 817–847.

Folk RA, Mandel JR, Freudenstein JV. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology***66** : 320–337.

Folk RA, Soltis PS, Soltis DE, Guralnick R. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. *American journal of botany* **105** : 364–375.

Folk RA, Stubbs RL, Mort ME, Cellinese N, Allen JM, Soltis PS, Soltis DE, Guralnick RP. 2019. Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proceedings of the National Academy of Sciences* **116** : 10874–10882.

Fu Y, Li L, Hao S, Guan R, Fan G, Shi C, Wan H, Chen W, Zhang H, Liu G *et al.* 2017. Draft genome sequence of the Tibetan medicinal herb *Rhodiola crenulata* . *Gigascience* **6** : gix033.

García N, Folk RA, Meerow AW, Chamala S, Gitzendanner MA, de Oliveira RS, Soltis DE, Soltis PS. 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). *Molecular phylogenetics and evolution* **111** : 231–247.

Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE. 2018. Plastome phylogenetics: 30 years of inferences into plant evolution.*In Advances in botanical research* **85** : 293–313.

Godsoe W, Larson MA, Glennon KL, Segraves KA. 2013. Polyploidization in *Heuchera cylindrica* (Saxifragaceae) did not result in a shift in climatic requirements. *American journal of botany* **100** : 496-508.

Jensen EL, Gaughran SJ, Garrick RC, Russello MA, Caccone A. 2021. Demographic history and patterns of molecular evolution from whole genome sequencing in the radiation of Galapagos giant tortoises.*Molecular Ecology* **30** : 6325–6339.

Jian S, Soltis PS, Gitzendanner MA, Moore MJ, Li R, Hendry TA, Qiu YL, Dhingra A, Bell CD, Soltis DE. 2008. Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology* **57** : 38–57.

Jiao F, Luo R, Dai X, Liu H, Yu G, Han S, Lu X, Su C, Chen Q, Song Q. 2020. Chromosome-level reference genome and population genomic analysis provide insights into the evolution and improvement of domesticated mulberry (*Morus alba*). *Molecular plant***13** : 1001-1012.

Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJ, Wickett NJ. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* **4** : 1600016.

Johnson MG, Pokorny L, Dodsworth S, Botigue LR, Cowan RS, Devault A, Eiserhardt WL, Epitawalage N, Forest F, Kim JT *et al.* 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering.*Systematic Biology* **68** : 594–606.

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14** : 587–589.

Kim HR, Kim S, Jie EY, Kim SJ, Ahn WS, Jeong SI, Yu KY, Kim SW, Kim SY. 2021. Effects of *Tiarella polyphylla* D. Don callus extract on photoaging in human foreskin fibroblasts Hs68 cells.*Natural Product Communications* **16** : 1–9.

Korgaonkar A, Han C, Lemire AL, Siwanowicz I, Bennouna D, Kopec RE, Andolfatto P, Shigenobu S, Stern DL. 2021. A novel family of secreted insect proteins linked to plant gall development. *Current Biology* **31** : 1836–1849.

Kvist S, Siddall ME. 2013. Phylogenomics of A nnelida revisited: a cladistic approach using genome-wide expressed sequence tag data mining and examining the effects of missing data. *Cladistics***29** : 435–448.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9** : 357–359.

Lee MY, Ahn KS, Lim HS, Yuk JE, Kwon OK, Lee KY, Lee HK, Oh SR. 2012. Tiarellic acid attenuates airway hyperresponsiveness and inflammation in a murine model of allergic asthma. *International Immunopharmacology* **12** : 117–124.

Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y *et al.* 2019a. Origin of angiosperms and the puzzle of the Jurassic gap. *Nature Plants* **5** : 461–470.

Lv S, Cheng S, Wang Z, Li S, Jin X, Lan L, Yang B, Yu K, Ni X, Li N *et al.* 2020. Draft genome of the famous ornamental plant*Paeonia suffruticosa* . *Ecology and Evolution* **10** : 4518–4530.

Magallon S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* **55** : 1762–1780.

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of *Malawi cichlids* reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution* **2** : 1940–1955.

Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, Garcia JF, Zhou Y, Delledonne M, Riaz S. 2020. The genetic basis of sex determination in grapes. *Nature Communications* **11** : 1-12.

Matthews ML, Endress PK. 2006. Floral structure and systematics in four orders of rosids, including a broad survey of floral mucilage cells. *Plant Systematics and Evolution* **260** : 199–221.

Maurin O, Anest A, Bellot S, Biffin E, Brewer G, Charles-Dominique T, Cowan RS, Dodsworth S, Epitawalage N, Gallego B*et al.* 2021. A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. *American Journal of Botany***108** : 1087–1111.

McLay TG, Birch JL, Gunn BF, Ning W, Tate JA, Nauheimer L, Joyce EM, Simpson L, Schmidt-Lebuhn AN, Baker WJ *et al.* 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences* **9** : e11420.

Miller M, Pfeiffer W, Schwartz T 2010 . Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)* **2010** : 1–8.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37** : 1530–1534.

Minio A, Cochetel N, Massonnet M, Figueroa-Balderas R, Cantu D. 2022. HiFi chromosome-scale diploid assemblies of the grape rootstocks 110R, Kober 5BB, and 101–14 Mgt. *Scientific Data* **9** : 1-8.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation.*Bioinformatics* **30** : i541–i548.

Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes.*Bioinformatics* **31** : i44–i52.

Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences* **107** : 4623–4628.

Moore MJ, Hassan N, Gitzendanner MA, Bruenn RA, Croley M, Vandeventer A, Horn JW, Dhingra A, Brockington SF, Latvis M *et al.* 2011. Phylogenetic analysis of the plastid inverted

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

repeat for 244 species: insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *International Journal of Plant Sciences* **172** : 541–558.

Okuyama Y, Fujii N, Wakabayashi M, Kawakita A, Ito M, Watanabe M, Murakami N, Kato M. 2005. Nonuniform concerted evolution and chloroplast capture: heterogeneity of observed introgression patterns in three molecular data partition phylogenies of Asian *Mitella*(Saxifragaceae). *Molecular Biology and Evolution* **22** : 285-296.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants.*Nature* **574** : 679–685.

Palmer JD, Zamir D. 1982. Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. *Proceedings of the National Academy of Sciences* **79** : 5006–5010.

Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5** : 568–583.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes.*Bioinformatics* **23** : 1061–1067.

Qiu YL, Li L, Wang B, Xue JY, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution***48** : 391–425.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al.* 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159** : 1665–1680.

Rieseberg LH, Beckstrom-Sternberg SM, Liston A, Arias DM. 1991. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asteraceae).*Systematic Botany* **16** : 50–76.

Rieseberg LH, Soltis D. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* ,**5** , 65–84.

Rieseberg LH, Whitton J, Linder CR. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica* **45** : 243–262.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics***19** : 1572–1574.

Rosvall M, Bergstrom CT. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* **105** : 1118–1123.

Rouard M, Droc G, Martin G, Sardos J, Hueber Y, Guignon V, Cenci A, Geigle B, Hibbins MS, Yahiaoui N *et al.* 2018. Three new genome assemblies support a rapid radiation in *Musa acuminata*(wild banana). *Genome Biology and Evolution* **10** : 3129–3140.

Roure B, Baurain D, Philippe H. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets.*Molecular Biology and Evolution* **30** : 197–214.

Shi C, Han K, Li L, Seim I, Lee SMY, Xu X, Yang H, Fan G, Liu X. 2020. Complete chloroplast genomes of 14 mangroves: phylogenetic and comparative genomic analyses. *BioMed Research International* . doi: 10.1155/2020/8731857.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31** : 3210–3212.

Soltis D, Soltis P, Endress P, Chase MW, Manchester S, Judd W, Majure L, Mavrodiev E. 2018.*Phylogeny and evolution of the angiosperms: revised and updated edition* . University of Chicago Press.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Soltis DE, Kuzoff RK. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae).*Evolution* **49** : 727–742.

Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* **98** : 704–730.

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, Stanhope MJ. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution* **18** : 132–143.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology***57** : 758–771.

Stegemann S, Keuthe M, Greiner S, Bock R. 2012. Horizontal transfer of chloroplast genomes between plant species. *Proceedings of the National Academy of Sciences* **109** : 2434-2438.

Strand A, Leebens-Mack J, Milligan B. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology***6** : 113–118.

Stull GW, Pham KK, Soltis PS, Soltis DE. 2022. Deep reticulation: the long legacy of hybridization in vascular plant evolution. *EcoevoRxiv* . doi: 10.32942/X24W2K

Sun M, Soltis DE, Soltis PS, Zhu X, Burleigh JG, Chen Z. 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae.*Molecular Phylogenetics and Evolution* **83** : 156–166.

Sun M, Naeem R, Su JX, Cao ZY, Burleigh JG, Soltis PS, Soltis DE, Chen ZD. 2016. Phylogeny of the Rosidae: A dense taxon sampling analysis. *Journal of Systematics and Evolution* **54** : 363–391.

Sun M, Folk RA, Gitzendanner MA, Soltis PS, Chen Z, Soltis DE, Guralnick RP. 2020. Recent accelerated diversification in rosids occurred outside the tropics. *Nature Communications* **11** : 1–12.

Sun WH, Li Z, Xiang S, Ni L, Zhang D, Chen DQ, Qiu MY, Zhang QG, Xiao L, Din L *et al.* 2021. The *Euscaphis japonica* genome and the evolution of malvids. *The Plant Journal* **108** : 1382–1399.

Thomas SK, Liu X, Du ZY, Dong Y, Cummings A, Pokorny L, Xiang QY, Leebens-Mack JH. 2021. Comprehending Cornales: phylogenetic reconstruction of the order using the Angiosperms353 probe set.*American Journal of Botany* **108** : 1112–1121.

Vatanparast M, Powell A, Doyle JJ, Egan AN. 2018. Targeting legume loci: A comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Applications in Plant Sciences* **6** : e1036.

Wai CM, Weise SE, Ozersky P, Mockler TC, Michael TP, VanBuren R. 2019. Time of day and network reprogramming during drought induced CAM photosynthesis in *Sedum album* . *PLoS Genetics* **15** : e1008209.

Wang H, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences* **106** : 3853–3858.

Wang M, Li J, Wang P, Liu F, Liu Z, Zhao G, Xu Z, Pei L, Grover CE, Wendel JF. 2021b. Comparative genome analyses highlight transposon-mediated genome expansion and the evolutionary architecture of 3D genomic folding in cotton. *Molecular biology and evolution***38** : 3621-3636.

Wang S, Liang H, Wang H, Li L, Xu Y, Liu Y, Liu M, Wei J, Ma T, Le C. 2022a. The chromosome-scale genomes of *Dipterocarpus turbinatus* and *Hopea hainanensis* (Dipterocarpaceae) provide insights into fragrant oleoresin biosynthesis and hardwood formation.*Plant biotechnology journal* **20** : 538-553.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H *et al.* 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40** : e49.

Wang Z, Li Y, Sun P, Zhu M, Wang D, Lu Z, Hu H, Xu R, Zhang J, Ma J *et al.* 2022b. A high-quality *Buxus austro -yunnanensis* (Buxales) genome provides new insights into karyotype evolution in early eudicots. *BMC Biology* **20** : 1–17.

Wendel JF, Doyle JJ. 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. *Molecular Systematics of Plants II* . doi: 10.1007/978-1-4615-5419-6_10.

Wu Z, Raven P. 2003 . Flora of China Illustrations. 8. Brassicaceae through Saxifragaceae. Beijing: Science Press/St Louis: Missouri Botanical Garden Press.

Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Cheol Yim W, Ha J *et al.* 2017. The *Kalanchoe* genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications* **8** : 1–15.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24** : 1586–1591.

Yuan J, Jiang S, Jian J, Liu M, Yue Z, Xu J, Li J, Xu C, Lin L, Jing Y *et al.* 2022. Genomic basis of the giga-chromosomes and giga-genome of tree peony *Paeonia ostii* . *Nature Communications* **13** : 1–16.

Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. 2014. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications* **5** : 1–12.

Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. 2017. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytologist* **214** : 1338–1354.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19** : 15–30.

Zhang C, Zhang T, Luebert F, Xiang Y, Huang CH, Hu Y, Rees M, Frohlich MW, Qi J, Weigend M *et al.* 2020. Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Molecular Biology and Evolution* **37** : 3188–3210.

Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist* **195** : 923–937.

Zhang N, Wen J, Zimmer EA. 2016. Another look at the phylogenetic position of the grape order Vitales: Chloroplast phylogenomics with an expanded sampling of key lineages. *Molecular Phylogenetics and Evolution* **101** : 216–223.

Zhao L, Li X, Zhang N, Zhang SD, Yi TS, Ma H, Guo ZH, Li DZ. 2016. Phylogenomic analyses of large-scale nuclear genes provide new insights into the evolutionary relationships within the rosids. *Molecular Phylogenetics and Evolution* **105** : 166–176.

Zhao T, Holmer R, de Bruijn S, Angenent GC, van den Burg HA, Schranz ME. 2017. Phylogenomic synteny network analysis of MADS-box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. *The Plant Cell* **29** : 1278–1292.

Zhao T, Zwaenepoel A, Xue JY, Kao SM, Li Z, Schranz ME, Van de Peer Y. 2021. Whole-genome microsynteny-based phylogeny of angiosperms. *Nature Communications* **12** : 1–14.

Zhu S, Chen J, Zhao J, Comes HP, Li P, Fu C, Xie X, Lu R, Xu W, Feng Y. 2020. Genomic insights on the contribution of balancing selection and local adaptation to the long-term survival of a widespread living fossil tree, *Cercidiphyllum japonicum* . *New Phytologist* **228** : 1674–1689.

Zhu XY, Chase MW, Qiu YL, Kong HZ, Dilcher DL, Li JH, Chen ZD. 2007. Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evolutionary Biology***7** : 1–15.

Zuntini AR, Frankel LP, Pokorny L, Forest F, Baker WJ. 2021. A comprehensive phylogenomic study of the monocot order Commelinales, with a new classification of Commelinaceae. *American Journal of Botany* , **108** : 1066–1086.

Posted on 11 Mar 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.167853251.10525445/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.