



## OPEN ACCESS

## EDITED BY

Huihui Li,  
Chinese Academy of Agricultural Sciences,  
China

## REVIEWED BY

Dongyan Zhao,  
Cornell University, United States  
Diaga Diouf,  
Cheikh Anta Diop University, Senegal  
Xuming Li,  
Hugo Biotechnologies Co., Ltd., China

## \*CORRESPONDENCE

Yang Dong  
✉ loyang@163.com  
Wen-Bin Yu  
✉ yuwenbin@xtbg.ac.cn

<sup>†</sup>These authors have contributed equally to this work and share first authorship

RECEIVED 05 June 2024

ACCEPTED 16 July 2024

PUBLISHED 02 August 2024

## CITATION

Chen B-Z, Li D-W, Luo K-Y, Jiu S-T, Dong X, Wang W-B, Li X-Z, Hao T-T, Lei Y-H, Guo D-Z, Liu X-T, Duan S-C, Zhu Y-F, Chen W, Dong Y and Yu W-B (2024) Chromosome-level assembly of *Lindenbergia philippensis* and comparative genomic analyses shed light on genome evolution in Lamiales. *Front. Plant Sci.* 15:1444234. doi: 10.3389/fpls.2024.1444234

## COPYRIGHT

© 2024 Chen, Li, Luo, Jiu, Dong, Wang, Li, Hao, Lei, Guo, Liu, Duan, Zhu, Chen, Dong and Yu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Chromosome-level assembly of *Lindenbergia philippensis* and comparative genomic analyses shed light on genome evolution in Lamiales

Bao-Zheng Chen<sup>1,2†</sup>, Da-Wei Li<sup>2†</sup>, Kai-Yong Luo<sup>1,2</sup>, Song-Tao Jiu<sup>3</sup>, Xiao Dong<sup>2</sup>, Wei-Bin Wang<sup>2</sup>, Xu-Zhen Li<sup>2</sup>, Ting-Ting Hao<sup>2</sup>, Ya-Hui Lei<sup>2</sup>, Da-Zhong Guo<sup>1,2</sup>, Xu-Tao Liu<sup>1,2</sup>, Sheng-Chang Duan<sup>2</sup>, Yi-Fan Zhu<sup>1,2</sup>, Wei Chen<sup>2</sup>, Yang Dong<sup>2\*</sup> and Wen-Bin Yu<sup>4,5\*</sup>

<sup>1</sup>College of Food Science and Technology, Yunnan Agricultural University, Kunming, Yunnan, China, <sup>2</sup>Yunnan Provincial Key Laboratory of Biological Big Data, Yunnan Agricultural University, Kunming, Yunnan, China, <sup>3</sup>Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China, <sup>4</sup>Center for Integrative Conservation and Yunnan Key Laboratory for the Conservation of Tropical Rainforests and Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan, China, <sup>5</sup>Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Mengla, Yunnan, China

Lamiales, comprising over 23,755 species across 24 families, stands as a highly diverse and prolific plant group, playing a significant role in the cultivation of horticultural, ornamental, and medicinal plant varieties. Whole-genome duplication (WGD) and its subsequent post-polyploid diploidization (PPD) process represent the most drastic type of karyotype evolution, injecting significant potential for promoting the diversity of this lineage. However, polyploidization histories, as well as genome and subgenome fractionation following WGD events in Lamiales species, are still not well investigated. In this study, we constructed a chromosome-level genome assembly of *Lindenbergia philippensis* (Orobanchaceae) and conducted comparative genomic analyses with 14 other Lamiales species. *L. philippensis* is positioned closest to the parasitic lineage within Orobanchaceae and has a conserved karyotype. Through a combination of Ks analysis and syntenic depth analysis, we reconstructed and validated polyploidization histories of Lamiales species. Our results indicated that *Primulina huaijiensis* underwent three rounds of diploidization events following the  $\gamma$ -WGT event, rather than two rounds as reported. Besides, we reconfirmed that most Lamiales species shared a common diploidization event (L-WGD). Subsequently, we constructed the Lamiales Ancestral Karyotype (LAK), comprising 11 proto-chromosomes, and elucidated its evolutionary trajectory, highlighting the highly flexible reshuffling of the Lamiales paleogenome. We identified biased fractionation of subgenomes following the L-WGD event across eight species, and highlighted the positive

impacts of non-WGD genes on gene family expansion. This study provides novel genomic resources and insights into polyploidy and karyotype remodeling of Lamiales species, essential for advancing our understanding of species diversification and genome evolution.

#### KEYWORDS

*Lindenbergia philippensis*, polyploidization history, karyotype evolutionary trajectories, Lamiales, genome assembly

## Introduction

Whole-genome duplication (WGD) or polyploidization is a prevalent process in terrestrial plants, contributing to genetic diversity, particularly in ferns and angiosperms (Julca et al., 2018; Kubis et al., 1998; Jurka, 2000; Korf, 2004; Majoros et al., 2004; Li and Durbin, 2009; Katoh and Standley, 2013; Kellogg, 2016; Li et al., 2016; Kalyaanamoorthy et al., 2017; Landis et al., 2018; Luo et al., 2018; Mandáková and Lysak, 2018; Li et al., 2019; Lovell et al., 2021; Kong et al., 2023; Liao et al., 2023; Liu et al., 2023; Letunic and Bork, 2024). An increasing number of WGD events have been identified across various lineages from whole-genomic sequencing and comparative genomic analyses (Cui et al., 2006; Soltis et al., 2009; Jiao et al., 2011; Vanneste et al., 2014; Van de Peer et al., 2017). WGD events generally arise through two primary mechanisms: autopolyploidization, involving whole-genome duplication within a single species, and allopolyploidization, resulting from the hybridization of two distinct species (Stebbins, 1947; Cheng et al., 2018). WGD events can provide their ancestors with a 'genomic playground', enabling new mutations to arise and tend to be fixed (through gene sub-functionalization and/or neofunctionalization). Consequently, these may contribute to physiological and morphological innovations, making WGD events as a significant driving force for species diversification and environmental adaptation (Cheng et al., 2018; Ren et al., 2018).

WGD events play important roles in promoting angiosperm diversification. However, whether these events are correlated with higher diversification rates remains a subject of debate (Tank et al., 2015; Kellogg, 2016; Landis et al., 2018). The 'lag phase' model, positing a delay between polyploidization events and subsequent lineage diversification, offers critical insights into influences of WGD events on species diversification (Dodsworth et al., 2015; Tank et al., 2015; Clark and Donoghue, 2017; Mandáková and Lysak, 2018). In other words, the WGD event likely initiated many speciation events across angiosperm lineages and also provided the genetic basis for the post-polyploid diploidization (PPD) process (Mandáková and Lysak, 2018). PPD process is different from WGD events by involving a process of karyotype evolutionary trajectories, which primarily includes changes in genome size, chromosomal rearrangements (alterations in chromosomal number and structure), subgenome-specific fractionation (including biased

gene retention/loss and gene sub-/neofunctionalization), differential expression of homologous genes, activation of transposable elements (TE), and epigenetic reprogramming (Paterson et al., 2004; Wang et al., 2005; Mandáková and Lysak, 2018; Zhuang et al., 2019). Therefore, the PPD process may also play a significant role in promoting the diversification rate of angiosperms.

The evolutionary mechanism and significance of promoting species diversity through the PPD process have been elucidated and reviewed by several studies (Mandáková et al., 2017; Mandáková and Lysak, 2018; Mayrose and Lysak, 2021). Generally, dysploid or non-dysploid changes in chromosome number and the fractionation of duplicated genes represent the primary aspects of the PPD process. Among them, chromosomal changes arising from dysploid alterations can radically increase or decrease the base number of chromosomes. Both descending and ascending dysploidy are significant in karyotype evolution, with the latter primarily observed in a few plant groups possessing monocentric chromosomes, such as the cycad genus *Zamia* (Rastogi and Ohri, 2019; Mayrose and Lysak, 2021). The evolution of land plant chromosomes is predominantly characterized by descending dysploidy (Carta et al., 2020; Mayrose and Lysak, 2021; Wang et al., 2022c; Kong et al., 2023). Centric fission is traditionally considered the most common form of ascending dysploidy (Birchler and Han, 2018). Unlike ascending dysploidy, descending dysploidy can be initiated by two mechanisms, including end-to-end joining (EEJ) and nested chromosome fusion (NCF) (Morin et al., 2017; Ren et al., 2019; Sun et al., 2022). In general, chromosomal diploidization can also be accompanied by various non-dysploid chromosomal rearrangements (CRs), such as inversions, reciprocal translocations, deletions, and duplications (Schubert and Lysak, 2011; Sun et al., 2022). With the alterations in dysploidy and non-dysploidy, the karyotype of specific lineages will undergo significant reshuffling, leading to karyotype modifications and potentially initiating interspecific reproductive barriers. Consequently, these processes may enable some species to acquire evolutionarily advantageous genetic diversity, thus adapting to a changing environment (Soltis et al., 2009; Clark and Donoghue, 2017).

In addition to dysploid or non-dysploid changes, the prevalence of dominant subgenomes, resulting from the preferential retention

of genes, is notable in many lineages that have undergone WGD events (Edger et al., 2017; Lovell et al., 2021; Wang et al., 2022b). Consequently, compared to a submissive subgenome, a dominant subgenome often retains more ancestral genes, exhibits higher levels of homologous gene expression, and undergoes stronger purifying selection (Sun et al., 2023). The biased retention (fractionation) of redundant genes resulting from WGD events may facilitate the adaptation of lineage-specific species to diverse ecological environments during speciation. Wu et al. (2020), for example, investigated gene duplicates across 25 genomes, revealing that duplicates retained following WGD events often correlate with environmental adaptability. Specifically, gene families associated with cold and dark conditions were frequently preserved in several lineages following WGD events around the Cretaceous-Paleogene boundary, a period marked by significant global cooling and darkness. Benefiting from karyotype changes, lineage-specific species evolve towards advantageous genetic diversity through the PPD process. This evolutionary advantage provides them with greater buffering capacity against mutations than their ancestors, thereby aiding speciation and enhancing adaptability in harsh environments (Comai, 2005; Ren et al., 2018; Clo, 2022). However, elucidating the complex process of PPD is challenging because, in most species, the ancestral chromosome tend to scatter and fragment within the new karyotype due to changes following the long evolutionary history (Damas et al., 2018; Ren et al., 2019; Zhao et al., 2021). Consequently, the intricate process of PPD, which involves a range of evolutionary modifications, remains a largely overlooked and understudied topic, particularly in certain specific lineages.

Representing one of the most abundant and diverse plant groups, the order Lamiales comprises over 23,755 species and 24 families (<https://www.britannica.com/plant/Lamiales>). These plants play a crucial role in providing a wide variety of horticultural, ornamental, and medicinal species. Besides, a variety of ecotype plants can be found in this lineage, including autotrophic and heterotrophic (parasitic and carnivorous) plants, aquatic and terrestrial plants. The high species diversity in Lamiales can be directly reflected in the abundant genetic materials. More importantly, almost all Lamiales species shared a common WGD event (the *L event*), and most retain a relatively complete ancestral karyotype, making them as ideal resources for investigating the PPD process (Feng et al., 2020).

The history of polyploidization and the PPD process have long been subjects of extensive study due to their significant roles in species adaptation and evolution. However, research across many lineages has been limited by a lack of comprehensive genomic resources. Encouragingly, the increasing availability of chromosomal-level genome assemblies is now enabling more detailed investigations into the history of polyploidization and the evolutionary trajectories of karyotypes within specific lineages. Significant advances have been made in some specific lineages such as Asteraceae (Kong et al., 2023), Cucurbitaceae (Wang et al., 2022a), and Nyssaceae (Feng et al., 2024). *L. philippensis* is part of Orobanchaceae in Lamiales with a unique taxonomic status, being closest to the parasitic lineage within Orobanchaceae (Li et al.,

2019; Mutuku et al., 2021). Besides, *L. philippensis* exhibited a conserved karyotype according to our previously exploration. In this study, to provide more insightful information about the polyploidization history, karyotype evolutionary trajectories, and the subgenomes evolutionary traits in the Lamiales, we assembled a chromosome-level genome of *L. philippensis* using Oxford Nanopore Technology (ONT) sequencing, Illumina sequencing, and high-throughput chromosome conformation capture (Hi-C) technology. Furthermore, we conducted a comparative genomic analysis on *L. philippensis* and other 14 genomes from 12 families within the order Lamiales, with *Vitis vinifera* and *Ophiorrhiza pumila* as outgroup references. The polyploidization histories of most Lamiales genomes were validated and corrected through combined Ks and syntenic depth analyses. Additionally, an ancestral karyotype of Lamiales species was constructed, and its evolutionary trajectories were deciphered in eight Lamiales species. Our study provides valuable genomic resources and will facilitate further research into genome evolution and the PPD process in Lamiales.

## Materials and methods

### Plant materials and DNA extraction

The plant samples of *L. philippensis* were collected from the same adult plant cultivated at Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, and identified by Professor Wen-Bin Yu. Fresh leaves were stored in liquid nitrogen and sent to Novogene Co., Ltd. for sequencing (Beijing, China). The high-quality genomic DNA of *L. philippensis* was prepared by a modified CTAB method (Karoonthaisiri et al., 2020) and purified with QIAGEN<sup>®</sup> Genomic kit (QIAGEN, USA) at Novogene Co., Ltd. (Beijing, China). The quality and quantity of the extracted genomic DNA were assessed using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), Qubit dsDNA HS Assay Kit on a Qubit 3.0 fluorometer (Life Technologies, Carlsbad, CA, USA) and electrophoresis on a 0.8% agarose gel, respectively.

### Long read sequencing

For long-read sequencing, a total of 2  $\mu$ g DNA was used for the ONT library construction. After the sample was qualified, long DNA fragments are selected using the BluePippin system (Sage Science, Beverly, MA, USA). Further, the ends of DNA fragments were repaired and a ligation reaction was conducted using the NEBNext<sup>®</sup> Ultra<sup>™</sup> II End Repair/dA-Tailing Module Kit. The ONT library with an insert size of 30 kb was prepared using the ligation sequencing kit 1D (SQKLSK109; Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's instructions. The ONT sequencing was then performed on an Oxford Nanopore PromethION 48 platform at Novogene Co., Ltd. (Beijing, China).

## Illumina short read sequencing

In total, 1 µg DNA was used as the input material and sequencing library was generated using the VAHTS Universal DNA Library Prep Kit for MGI (Vazyme, Nanjing, China). Following the manufacturer's recommendations, and index codes were added to attribute sequences to sample. The Library quantification and size were measured using Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A paired-end library was created with a 350 bp insert size using the GenElute Plant Genomic DNA Miniprep kits following the manufacturer's instructions (Sigma-Aldrich, Corp., St. Louis, MO, USA). Subsequently, the short-read library was performed on the Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA).

## Hi-C library construction and sequencing

The Hi-C libraries were constructed following established protocols (Padmarasu et al., 2019). Initially, samples were cross-linked under vacuum infiltration using formaldehyde. Subsequently, the cross-linked samples were subsequently digested using *DpnII*. After reversing cross-links, the ligated DNA was extracted using the QIAamp DNA Mini Kit (Qiagen) according to the manufacture's instruction. Purified DNA was then sheared to 300 bp to 500 bp fragments, which underwent blunt-end repair, A-tailing, and adaptor addition. The resulting fragments were purified through biotin-streptavidin-mediated pull-down and subjected to PCR amplification. Finally, the Hi-C libraries were quantified and sequenced on the Illumina NovaSeq 6000 platform (Illumina Inc., San Diego, CA, USA).

## Genome assembly and quality evaluation

Prior to conducting the assembly, it is imperative to conduct a comprehensive survey of the genomic features. To accomplish this, we utilized clean paired-end short reads and employed GenomeScope (v2.0) and Jellyfish (v2.2.10) with default parameters to assess the genome size, heterozygosity, and repeat content of the *L. philippensis* genome (Marçais and Kingsford, 2011; Ranallo-Benavidez et al., 2020). Furthermore, flow cytometry (BD FACSCalibur) was also used to investigate the genome size. For the genome assembly, we initially assembled the clean long reads to generate the draft assembly using NextDenovo (v2.4.0) with the following parameters "task = all; rerun = 3; read\_cuoff = 1k; seed\_cutoff = 8k; seed\_cuoff = 8k; genome\_size = 400 m; seed\_cutfiles = 80; blocksize = 10g; pa\_correction = 80; minimap2\_options\_raw = -x ava-ont -t 16; sort\_options = -m 10g -t 16 -k 50; correction\_options = -p 32 random\_round = 100 minimap2\_options\_cns = -x ava-ont -t 20 -k17 -w17; nextgraph\_options = -a 1". Subsequently, the draft assembly underwent three rounds of polishing using NextPolish (v1.3.1) with the following parameters "rerun = 3; parallel\_jobs = 8;

multithread\_jobs = 8; sgs\_options = -max\_depth 100 -bwa". To obtain a preliminary genome assembly, haplotyped duplication sequences were filtered using Redundans (v1.01) with parameters "ident=0.95, ovl=0.95" (Pryszcz and Gabaldón, 2016). For scaffolding contigs, Hi-C data were mapped to the *L. philippensis* preliminary assembly using Juicer (v1.6.2) with parameters of "-s *DpnII* -t 40" (Durand et al., 2016). Subsequently, the valid reads were utilized to order and orient the contigs by employing 3D-DNA (Dudchenko et al., 2017). Any missing joins were rectified based on the Hi-C contact signals using Juicebox (v1.11.08) (<https://github.com/aidenlab/Juicebox>). The completeness of the genome assembly was evaluated using BUSCO (v5.1.2) with "eukaryota\_odb10" dataset downloaded from the BUSCO website (<https://busco-archive.ezlab.org/v3/>) (Seppey et al., 2019). We utilized BWA-MEM (v0.7.12) (Li and Durbin, 2009) for mapping Illumina reads to the assembly and computed mapping statistics with SAMtools (v1.9) using the "flagstat" module (Danecek et al., 2021).

For transcriptome assembly, we downloaded the raw reads of RNA sequencing data from NCBI (ERR2040586, ERR2040587) and used Fastp (v0.20.1) to filter the low quality reads with the following parameters "-q 30 -u 40 -l 50 -w 16". Trinity (v2.11.0) with the following parameters of "-seqType fq -JM 300G -CPU 20" was used to perform the transcriptome *de novo* assembly (Grabherr et al., 2011).

## Genome annotation

Repetitive elements (REs) across all 17 species were predicted through a combination of evidence-based and *ab initio* methods. For the evidence-based method, we predicted repeats within the target genome using RepeatMasker with the following parameters "-a -nolow -no\_is -norna" and RepeatProteinMask with parameters of "-engine ncbi -noLowSimple -pvalue 0.0001" (vopen-4.0.9) (Chen, 2004) based on the Repbase (v24.06) (Jurka, 2000). For the *ab initio* method, we first constructed a *de novo* repeat library of the target genome using RepeatModeler (v2.0) with the parameter "-engine rmblast". Long terminal retrotransposons (LTRs) were identified using both LTR\_FINDER\_parallel (v1.1) (Ou and Jiang, 2019) with the following parameters "-harvest\_out -size 1000000 -time 300 -finder" and LTRharvest v1.0 (Ellinghaus et al., 2008) with the following parameters "-minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -motif TGCA -motifmis 1 -similar 85 -vic 10 -seed 20 -seqids yes". Then, the LTRs candidates were further passed to LTR\_retriever (v2.8) (Ou and Jiang, 2018) with default parameters to filter out false LTRs, and calculate the LTR Assembly Index (LAI). Finally, the repeat libraries from LTR\_retriever and RepeatModeler were merged to complete *de novo* prediction of REs using RepeatMasker with the following parameters "-nolow -no\_is -norna". In addition, tandem repeats were predicted by using the Tandem Repeat Finder (TRF v4.09) package (Benson, 1999) with the following parameters "2 7 7 80 10 50 2000 -d -h".

The prediction of protein-coding genes in the *L. philippensis* genome involved the integration of three distinct methods, including *ab initio* gene prediction, homology-based gene

prediction, and RNA-Seq-assisted gene prediction. Before proceeding with protein-coding gene prediction, we soft-masked the assembled *L. philippensis* genome using Bedtools (Quinlan and Hall, 2010) according to the annotated file of TEs. For *ab initio* gene prediction, we employed GenScan (v1.0) (Aggarwal and Ramaswamy, 2002), GlimmerHMM (v3.0.3) (Majoros et al., 2004), Augustus (v3.2.2) (Stanke et al., 2008), and SNAP (v1.0) (Korf, 2004) to predict protein-coding genes. Next, homology-based gene prediction was performed using TBLASTN (Altschul et al., 1990) with a cutoff threshold of  $1e^{-5}$ , searching against protein sequences from five reference species, including *A. thaliana*, *V. vinifera*, *Solanum lycopersicum*, *S. indicum*. To execute RNA-Seq-assisted gene prediction, the transcriptome assembly was used for gene prediction by comparing it with genomes using the Program to Assemble Spliced Alignments (PASA) (Haas et al., 2003). Finally, a non-redundant gene set was integrated using EvidenceModeler (v1.1.1) (Haas et al., 2008) and updated with PASA. Based on sequence similarity and domain conservation, functional annotations of gene models were predicted by the online EggNOG (v5.0.0) database (Cantalapiedra et al., 2021).

## Phylogenetic reconstruction and comparative genomics analysis

The longest protein-coding sequences of *L. philippensis* and the other 16 species were extracted and clustered using OrthoFinder (v2.5.2) (Emms and Kelly, 2019). Subsequently, the protein-coding sequences of single-copy gene were subjected to multiple sequence homology alignment using Mafft (v7.471) (Katoh and Standley, 2013) with the following parameters “-localpair -maxiterate 1000”. Each coding sequence (CDS) was aligned separately according to the corresponding amino acid alignments using PAL2NAL (v14) (Suyama et al., 2006), and then all CDS matrixes were concatenated into a supermatrix. After filtering the poorly aligned regions of integrated CDS alignments using Gblocks (v0.91b) (Castresana, 2000), a maximum likelihood (ML) tree was constructed using IQ-TREE v2.2.0.3 (Kalyaanamoorthy et al., 2017) with the following parameters “-m MFP; -bb 1000; -nt 10” and with the best-fit model (GTR+F+I+G4). Divergence times for single-copy gene supermatrix dataset were estimated based on the ML tree using MCMCTree module from the PAML package with the following parameters “burnin = 50000; nsample = 100000” (Yang, 2007). Two fossil calibration points for divergence time estimation were searched from the TimeTree database (<http://www.timetree.org/>). One is *L. philippensis* versus *V. vinifera* (range: 111.4~123.9 Mya) and another is *L. philippensis* versus *B. alternifolia* (range: 31.5~56.1 Mya). The resulting phylogenetic tree was visualized using FigTree (v1.4.3) (<https://github.com/rambaut/figtree>). The expansion and contraction of gene family in *L. philippensis* were determined using Computational Analysis of Gene Family Evolution (CAFE v5.0) (Mendes et al., 2021) with the following parameters “-k 3 -cores 30”. This process through comparing orthologs groups of itself with other 16 species based on the cluster results of OrthoFinder (v2.5.2) (Emms and Kelly, 2019) and the ultrametric phylogeny generated

from r8s (Mulcahy et al., 2012). Finally, ortholog groups with  $P < 0.05$  were considered as gene families undergoing significant expansion or contraction. The correlation between genome size and repeat content was calculated using the “cor.test” function in R 4.2.1 with the Pearson method.

## Analyses of whole-genome duplication

The WGD events experienced by *L. philippensis* and the other 16 species were determined by combining the analysis of synonymous substitutions per synonymous site (Ks) and the syntenic analysis that reflects the syntenic depth of intergenomic collinear blocks.

Firstly, syntenic blocks (paralogous genes) within each species were identified using WGDI (v0.6.2) (Sun et al., 2022) with the parameters “-d, -icl, -ks, -bi, -c, -bk”, and then the Ks between collinear genes were calculated by using the Nei-Gojobori approach as implemented in the PAML (v4.9h) package (Yang, 2007). Median Ks values were used to represent each syntenic block, and Ks peak fitting was performed using WGDI with the “-pf” option (Sun et al., 2022). Secondly, the syntenic depths of collinear genes within other species were employed to determine syntenic ratios between different species, confirming their polyploidy levels. To exactly detect the polyploidization levels, we detected the syntenic depth via two methods. One method involved using WGDI (Sun et al., 2022) with the “-bk” option, while the other one utilized JCVI (v1.3.8) with two sets of parameters: “jcv.compara.catalog ortholog; -no\_strip\_names -cscore=0.99” and “jcv.compara.synteny depth -histogram” (Tang et al., 2008).

## Inference of Lamiales ancestral karyotype and analyses of karyotype evolutionary trajectory

We used the ‘Telomere-centric genome repatterning model’ proposed in previous study (Wang et al., 2015; Sun et al., 2022) to construct the LAK and infer its evolutionary trajectory in Lamiales plants. Given the conserved karyotype of *L. philippensis*, its genome was chosen to complete the construction of LAK. The construction process was delineated into three key steps: Step 1 entailed the detection of Whole Genome Duplication (WGD); Step 2 involved the reconstruction of the ancestral karyotype; and Step 3 focused on validating the accuracy of the reconstructed ancestral karyotype. A more detailed description was provided in Supplementary Materials (Note 1).

To analyze the evolutionary history of karyotype among Lamiales species, 13 species from eight families were chosen. Similar to the process of LAK inference, we utilized WGDI (Sun et al., 2022) with the parameters “-d, -icl, -bi, -c, -km, -d” to complete karyotype mapping between different species with LAK. Additionally, the dynamic evolutionary trajectory of LAK and post-LAK following the  $\gamma$ -WGT event was illustrated using Adobe Animate software.

## Construction of subgenome and comparative analyses of eight Lamiales species

Eight species, each representing a distinct family and possessing a relatively complete ancestral karyotype, were selected to investigate the traits of karyotype evolutionary trajectory. To precisely build the sub-genomes, two LAK copies in *L. philippensis* (post-LAK1-22) were created to aid in constructing the sub-genome of other species. Similar to the previous reconstruction of LAK, we utilized WGDI with the parameters “-d, -icl, -bi, -c, -km, -ak, -d” to construct the sub-genomes of the eight species. The syntenic relationship between the 16 subgenomes was then visualized using JCVI (Tang et al., 2008). For further characterization of the subgenome, each subgenome was tackled as species, and their corresponding protein-coding sequences were clustered into orthogroups using OrthoFinder (v2.5.2) (Emms and Kelly, 2019). The intersection of different groups was visualized using a website tool at <https://bioinformatics.psb.ugent.be/webtools/Venn/>.

## The identification of different modes of gene duplication and the analysis of CYP superfamily

Various gene duplication modes were identified utilizing the “DupGen\_finder-unique.pl” module of DupGen\_finder (Qiao et al., 2019) with default parameters, and *O. pumila* was set as the reference. We identified CYP genes using HMMER v3.3.2 (Potter et al., 2018) with parameter ‘-cut\_tc’. The Pfam HMM models, namely PF00067 was set as queries for the identification of CYPs. The previously characterized *A. thaliana* CYPs genes was downloaded from <http://p450.kvl.dk/index.shtml> and used as outgroups. To construct the phylogenies for CYPs, the protein sequences were aligned using MAFFT (Katoh and Standley, 2013). The poor alignments were trimmed using trimAl (Capella-Gutiérrez et al., 2009). ML phylogenetic trees were constructed with IQ-TREE (Kalyaanamoorthy et al., 2017) and visualized using iTOL (Letunic and Bork, 2024).

## Results

### Genome assembly and annotation of *Lindenbergia philippensis*

Through the analysis of 17-kmer frequencies from Illumina short-reads and flow cytometry, the genome size of *L. philippensis* was estimated at approximately 416.78 Mb and 396.66 Mb, respectively, with a heterozygosity rate of 0.706% (Supplementary Tables S1, S2; Supplementary Figures S1, S2). The consistency of genome size estimation was observed between these two methods. A total of 40.13 Gb (101×) of raw ONT long-reads were utilized for the initial assembly of contigs using NextDenovo (v2.5) (<https://github.com/Nextomics/NextDenovo>) (Supplementary Table S3).

After two rounds of polish of the 80.49 Gb (202×) Illumina short-reads using Nextpolish v1.2.1 (<https://github.com/Nextomics/NextPolish>), we obtained 949 final contigs with a total size of 406.79 Mb and a N50 of 1.79 Mb (Supplementary Tables S3, S4). Subsequently, the polished contigs were clustered and ordered using 131.51 Gb (331×) Hi-C data through Juicer (Durand et al., 2016) and 3D-DNA (Dudchenko et al., 2017), resulting in the successful construction of 16 pseudo-chromosomes with a scaffold N50 of 23.51 Mb, covering approximately 96.55% of the final assembled sequences (393.39 Mb/407.46 Mb) (Figure 1A; Supplementary Figure S3).

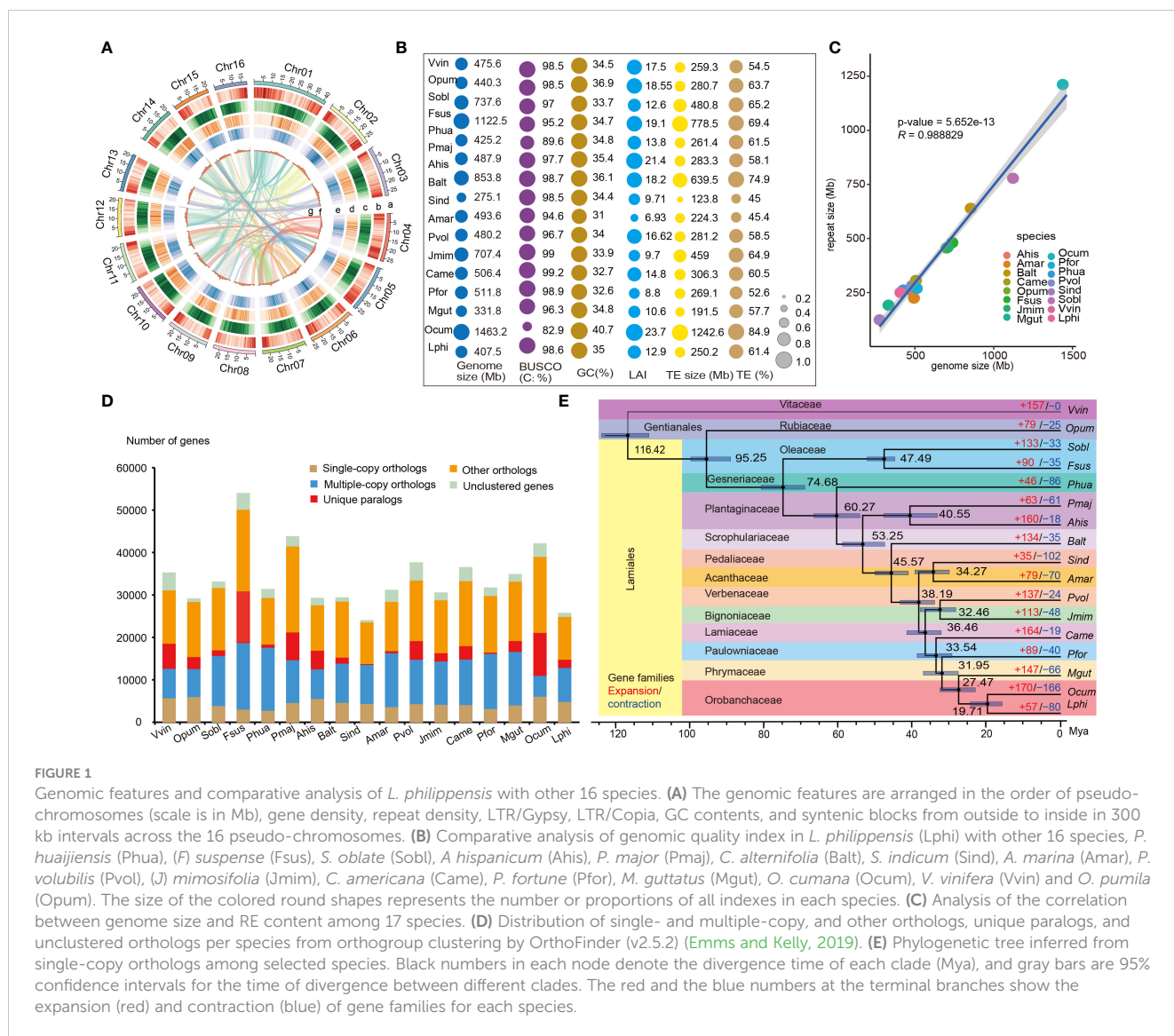
The final assembled genome size was nearly close to the size estimated by the flow cytometry and the 17 kmer frequency distribution (Supplementary Figure S1; Supplementary Table S4). Furthermore, mapping 536,633,198 Illumina reads to the final assembly resulted in a mapping rate of 99.15% and a coverage rate of 95.05% (Supplementary Table S5). The completeness of genome assembly is 98.6% of BUSCO genes based on the embryophyta\_10 dataset (Seppey et al., 2019), which was comparable with 14 genomes of Lamiales species (Figure 1B; Supplementary Table S6). *Lindenbergia philippensis* genome had a high Long-terminal repeat (LTR) Assembly Index (LAI) score of 12.9 (Figure 1B), meeting the “reference standard” (LAI value > 10) of genome assembly proposed by Finn et al. (2011).

Based on homologous and *de novo* prediction, 250.16 Mb of repetitive elements (REs) were identified in the *L. philippensis* genome, constituting 61.40% of the assembly genome. These elements included LTRs (39.68%), DNA transposons (6.63%), LINEs (0.86%), SINEs (0.02%), and unclassified sequences (15.78%) (Supplementary Table S8). After masking the REs, 25,693 protein-coding genes were identified by combining *de novo*, homology-based, and RNA-Seq-based predictions. On average, each predicted gene had an average length of 3,800 bp and contained five exons with an average length of 232 bp (Supplementary Table S9). Approximately 94.89% of protein-coding genes were functionally annotated by existing databases (Supplementary Table S10).

### Comparative and evolutionary genomics of *Lindenbergia philippensis* and its relatives

To investigate genomic characteristics of *L. philippensis* and its relatives, comparative genomic analyses were performed on 15 representative genomes from 12 families of Lamiales and two outgroups, *V. vinifera* and *O. pumila* (Figure 1E, Supplementary Table S7). The annotation and comparison of their REs revealed that the repeat size was widely distributed in these 17 genomes, varying from 88.8 Mb to 1,242.6 Mb, with *O. cumana* exhibiting the highest repeat content (Figure 1B; Supplementary Table S8). Meanwhile, correlation analysis showed that the genome size was positively correlated with the repeat contents ( $R = 0.97$ ,  $P < 0.05$ ), which was consistent with previous studies (Figure 1C) (Shao et al., 2019; de Lima and Ruiz-Ruano, 2022).

By employing OrthoFinder (v2.5.2) (Emms and Kelly, 2019) to cluster orthologs, a total of 576,537 genes from 17 genomes were

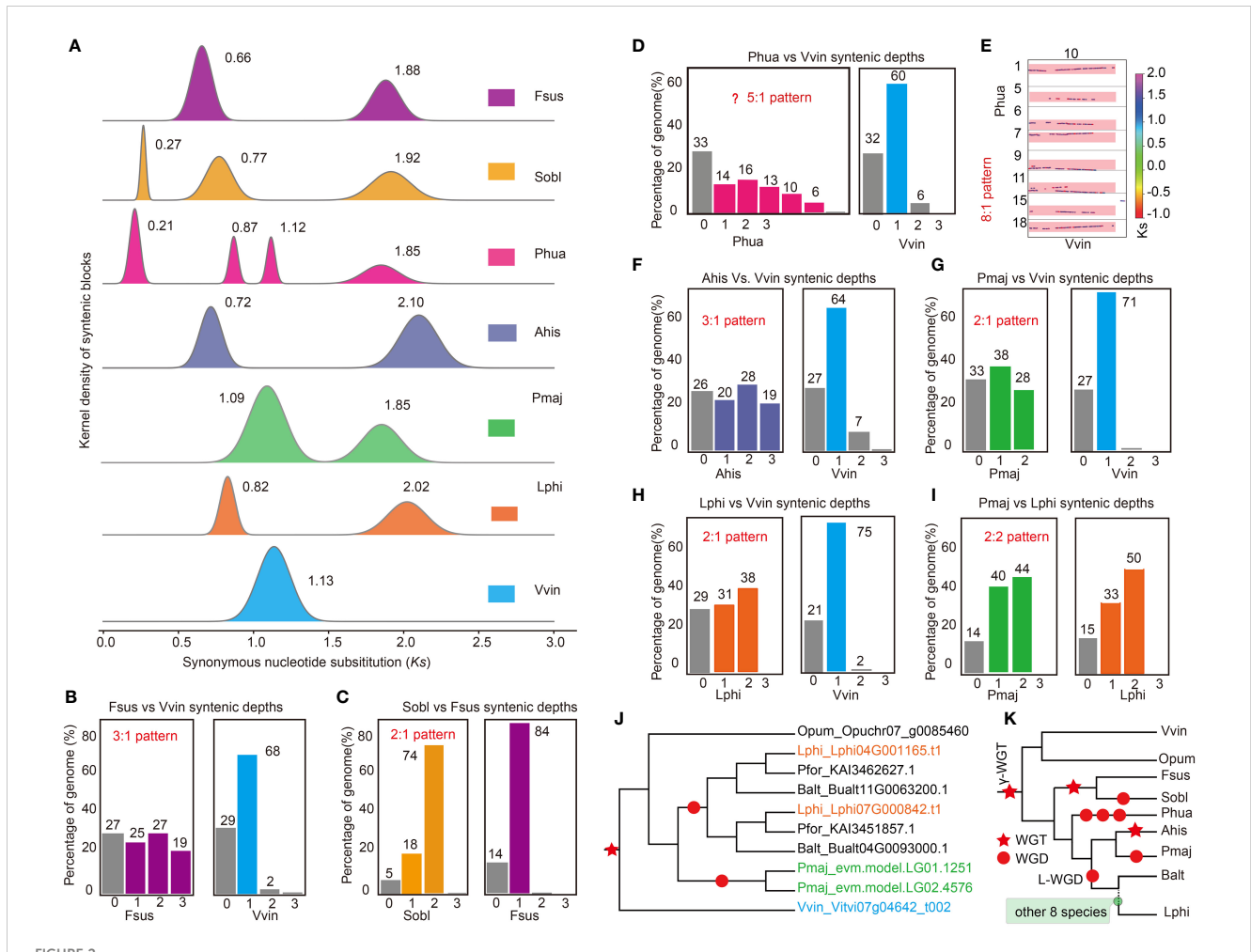


classified into 540,506 orthologs groups and 36,031 unclustered genes. Among them, 7,775 groups were shared by all 17 species, including 326 single-copy orthologs groups (Figure 1D; Supplementary Table S11). *Lindenbergia philippensis* possessed 1,934 species-specific genes, including 289 orthologs genes and 628 unclustered genes (Figure 1D; Supplementary Table S12). The biological processes of species-specific genes were mainly distributed in ‘host cellular response’, ‘metabolic process’ and ‘biosynthetic process’ (Supplementary Figure S4), suggesting the evolution of key enzyme genes associated with metabolite synthesis and pathways for environmental adaptation in *L. philippensis*. The phylogenetic tree constructed using 326 conserved single-copy genes from 17 genomes using the maximum-likelihood method showed that *L. philippensis* was sister to parasitic species in Orobanchaceae, aligning with prior research (Mutuku et al., 2021). Divergence time estimation showed that the divergence between *L. philippensis* and *O. cumana* occurred at ~19.71 million years ago (Mya), and the Lamiales diverged from the Gentianales at ~95.25 Mya (Figure 1E). Expansion or contraction of gene families

is often associated with adaptive divergence in closely related species (Cheng et al., 2017). Therefore, we investigated changes in gene families using the estimated phylogeny to capture key genomic information associated with *L. philippensis* adaptability. Compared to related species, a total of 57 gene families (including 496 genes) and 80 gene families (including 90 genes) exhibited significant expansion and contraction in the *L. philippensis* genome, respectively ( $P < 0.05$ ) (Figure 1E). Interestingly, the expanded genes were primarily enriched in many secondary metabolite biosynthetic pathways (e.g., flavonoid biosynthesis and metabolic process, glucan metabolic process and cellulose biosynthetic process), suggesting that *L. philippensis* produces some active substances such as phenols (Supplementary Figure S5).

### Polyploidization history of Lamiales species

To unveil the ancient polyploidization history of Lamiales species, we examined the distribution of substitutions per



**FIGURE 2**  
 Inference of polyploidization histories in the genomes of the studied Lamiales species. (A) The synonymous substitution (Ks) distributions of gene pairs in syntenic blocks among compared genomes. (B) The ratio of orthologous genes between *F. suspense* (Fsus) and *V. vinifera* (Vvin). (C) The ratio of orthologous genes between *S. oblate* (Sobl) and *F. suspense* (Fsus). (D) The ratio of orthologous genes between *P. huaijiensis* (Phua) and Vvin. (E) The syntenic depth of homologues blocks between Phua and Vvin. (F) Ratio of orthologous genes between *A. hispanicum* (Ahis) and Vvin. (G) The ratio of orthologous genes between *P. major* (Pmaj) and Vvin. (H) The ratio of orthologous genes between *L. philippensis* (Lphi) and Vvin. (I) The ratio of orthologous genes between *P. major* (Pmaj) and Lphi. (J) The phylogenetic tree of ten orthologous genes, derived from four Lamiales species, Vvin and *O. pumila* (Opum). (K) Overview of WGD events in those 15 Lamiales species. Polyploidization events are indicated by red pentagram (triploidization, WGT) and red round shape (diploidization, WGD).

synonymous site (Ks) of intra-genomic collinear blocks in the 15 Lamiales species. Two to four separate peaks were detected in the Ks distribution for species-specific paralogous pairings in those species (Figure 2A, Supplementary Figure S6), indicating that at least one round of WGD events occurred in this lineage following the  $\gamma$ -WGT event. For example, four obvious Ks peaks were observed in *P. huaijiensis*, reflecting a younger WGD event at Ks 0.21, two distinct WGD events at Ks 0.87 and 1.12 respectively, and  $\gamma$ -WGT event at Ks 1.85. In *S. oblate*, three Ks peaks indicated a younger WGD event at Ks 0.27, a WGD event at Ks 0.77, and the  $\gamma$ -WGT event at Ks 1.92 (Figure 2A). Other species such as *F. suspensa*, *A. hispanicum*, *P. major*, *B. alternifolia*, *S. indicum*, *P. volubilis*, *P. fortunei*, *J. mimosifolia*, *C. americana*, *M. guttatus*, *O. cumana* and *L. philippensis* exhibited two peaks (Figure 2A; Supplementary Figure S6). The first peak indicated a recent WGD event, while the second peak corresponded to the  $\gamma$ -WGT event. *Vitis vinifera* and *O. pumila* displayed only a single peak representing  $\gamma$ -WGT

event (Figure 2A; Supplementary Figure S6). The distribution of Ks peaks showed differences among species (Figure 2A; Supplementary Figure S6), which was usually caused by evolutionary rate variations in habitat divergence (Sensalari et al., 2022). For example, besides *V. vinifera*, *P. fortunei* exhibited the lowest Ks value in  $\gamma$ -WGT event (Supplementary Figure S6), indicating it may have a lower evolutionary rate than other species.

To determine the polyploidization level of Lamiales species after the  $\gamma$ -WGT event, their ratio of orthologous genes with *V. vinifera* was examined with precision. Generally, a species experienced the WGD event will have a corresponding orthologous gene ratio with another species, which only retained their common ancestral karyotype. For instance, Hoang et al. (2023) demonstrated that *Cleome violacea* did not undergo Gg- $\alpha$  (diploidization) event after divergence from a shared ancestor with *Gynandropsis gynandra*, which had experienced a diploidization event. Consequently, *C. violacea* exhibited a 1:2 orthologous gene ratio with *G. gynandra*.



Through comparative analysis of genome syntenic blocks, we have successfully determined the level of polyploidization in those 17 species following  $\gamma$ -WGT event. *Ophiorrhiza pumila* exhibited a 1:1 orthologous gene ratio with *V. vinifera* (Supplementary Figure S7), suggesting that it underwent only the shared  $\gamma$ -WGT event and did not experience additional WGD events after diverging from their common ancestor. *Forsythia suspensa* had a 3:1 orthologous gene ratio with *V. vinifera* (Figure 2B; Supplementary Figure S8), indicating it experienced a triploidization at Ks peak 0.66. *Syringa oblata* had a 6:1 orthologous gene ratio with *V. vinifera* and a 2:1 orthologous gene ratio with *F. suspensa*, respectively (Figure 2C; Supplementary Figures S9, S10), indicating it experienced a common triploidization with *F. suspensa* at Ks peak 0.77 and an independent diploidization event at Ks peak 0.27 (Figure 2A). The two rounds of WGD events were also proved by previous studies (Julca et al., 2018; Feng et al., 2020). *Primulina huaijiensis* showed an 8:1 orthologous gene ratio with *V. vinifera* (Figures 2D, E; Supplementary Figure S11). Therefore, the orthologous gene ratio between *P. huaijiensis* and *V. vinifera* could be explained as  $2 \times 2 \times 2:1$  according to the Ks distribution, corresponding to three rounds of diploidization events rather than two rounds of diploidization events reported in a previous study (Feng et al., 2020). *Antirrhinum hispanicum* had a 3:1 orthologous gene ratio with *V. vinifera* (Figure 2F; Supplementary Figure S12), indicating that it underwent a triploidization event at Ks peak 0.72 (Figure 2A), which was consistent with previous results (Zhu et al., 2023b). *Avicennia marina* had a 4:1 orthologous ratio with *V. vinifera* and a 2:1 orthologous gene ratio with *L. philippensis* (Supplementary Figures S13; S33), respectively, indicating it experienced a common diploidization with *L. philippensis* at Ks peak 0.77 and an independent diploidization event at Ks peak 0.27 (Figure 2A).

In addition, other ten species, including *P. major*, *B. alternifolia*, *S. indicum*, *P. volubilis*, *P. fortunei*, *J. mimosifolia*, *C. americana*, *M. guttatus*, *O. cumana* and *L. philippensis*, had a 2:1 orthologous gene ratio with *V. vinifera* (Figures 2G, H; Supplementary Figures S14, S23). This suggests that they could have experienced a common diploidization event after the  $\gamma$ -WGT event, corresponding to *L\_event* revealed by previous results (Julca et al., 2018; Feng et al., 2020). To better determine whether the ten species underwent a common diploidization event, we examined the inter-genomic collinearity relationships among orthologous genes, using *L. philippensis* as the reference. Except for *P. major*, which exhibited a 2:2 orthologous gene ratio with *L. philippensis* (Figure 2I; Supplementary Figure S24), the remaining eight species displayed a 2:1 orthologous gene ratio with *L. philippensis* (Supplementary Figures S25-S32). This suggests that *P. major* may have undergone a diploidization event independently, while the remaining nine species shared a common diploidization event after the  $\gamma$ -WGT event. Furthermore, phylogenetic analyses of orthologous genes derived from four paired subgenomes and using *O. pumila* and *V. vinifera* as outgroups, further corroborating this hypothesis (Figure 2J).

In summary, after the  $\gamma$ -WGT event, Lamiales species underwent multiple WGD events based on Ks and syntenic analyses. *Syringa oblata* and *F. suspensa* underwent a shared triploidization event, and *S. oblata* subsequently underwent an independent diploidization event, in line with the previous

finding (Julca et al., 2018; Feng et al., 2020). *Primulina huaijiensis* experienced three rounds of diploidization events. *Antirrhinum hispanicum* experienced a triploidization event, while *P. major* underwent a diploidization event. Integrating phylogenetic and syntenic analyses, we found that the remaining ten species from nine families, including *B. alternifolia* (Scrophulariaceae), *S. indicum* (Pedaliaceae), *A. marina* (Acanthaceae), *P. volubilis* (Verbenaceae), *J. mimosifolia* (Bignoniaceae), *C. americana* (Lamiaceae), *P. fortunei* (Paulowniaceae), *M. guttatus* (Phrymaceae), *O. cumana*, and *L. philippensis* (Orobanchaceae), underwent a shared diploidization event, known as L-WGD event (Figure 2K).

## Construction of Lamiales ancestral karyotype and analyses of karyotype evolutionary trajectories

After polyploidization, substantial karyotype changes frequently occur in many plant genomes. These changes can alter the basic chromosome number and trigger species diversification. In Lamiales, the chromosome numbers of 15 selected species range from  $2n=12$  to  $2n=64$  (Supplementary Table S7). These variations are primarily caused by karyotype changes. To uncover the karyotype evolutionary trajectories, the *L. philippensis* genome was used to reconstruct the Lamiales ancestral karyotype (LAK).

To comprehensively delineate the karyotype evolutionary trajectory of the LAK in Lamiales species, we defined the 21 proto-chromosomes of the Ancestral Core Eudicot Karyotypes (ACEK), derived from the triplication of seven ancestral eudicot karyotypes (AEK), as A1-7, B1-7, and C1-7. As a result, a putative LAK was constructed consisting of 11 proto-chromosomes (LAK1-LAK11), which shared the same base chromosomal number with the sister clade species such as *O. pumila* ( $2n=22$ ), *Morinda officinalis* ( $2n=22$ ), and *Leptodermis oblonga* ( $2n=22$ ) from the Gentianales order. This suggests a possible common ancestral karyotype between Lamiales and Gentianales. To validate this hypothesis, we generated a dot plot by comparing the LAK and *O. pumila* (Rubiaceae family) with ACEK. Rubiaceae, positioned at the root of Gentianales, exhibits a higher likelihood of sharing the same karyotype with LAK among its species. The dot plot analysis indicated that nine chromosomes of the LAK and *O. pumila* exhibited a one-to-one correspondence in their collinearity relationship (Supplementary Figures S34-S36). The primary distinction between them lies in the rearrangement of proto-chromosome B6 (Figure 3). Following the methodology used in constructing the LAK, we constructed a hypothetical common ancestral karyotype for Lamiales and Gentianales orders, labeled as LG1-LG12 (Figure 3; Supplementary Figure S37). In summary, LAK evolved into 11 proto-chromosomes through a series of chromosomal rearrangements, including nine end-to-end joining (EEJ), two nested chromosome fusions (NCF), and ten reciprocal translocations of chromosome arms (RTA). For example, the formation of proto-chromosomes LAK1 was mainly explained by the fusion of A6 and C6 initially with the EEJ pattern and then further fused with B6 through the EEJ pattern (Figure 3). Similarly,

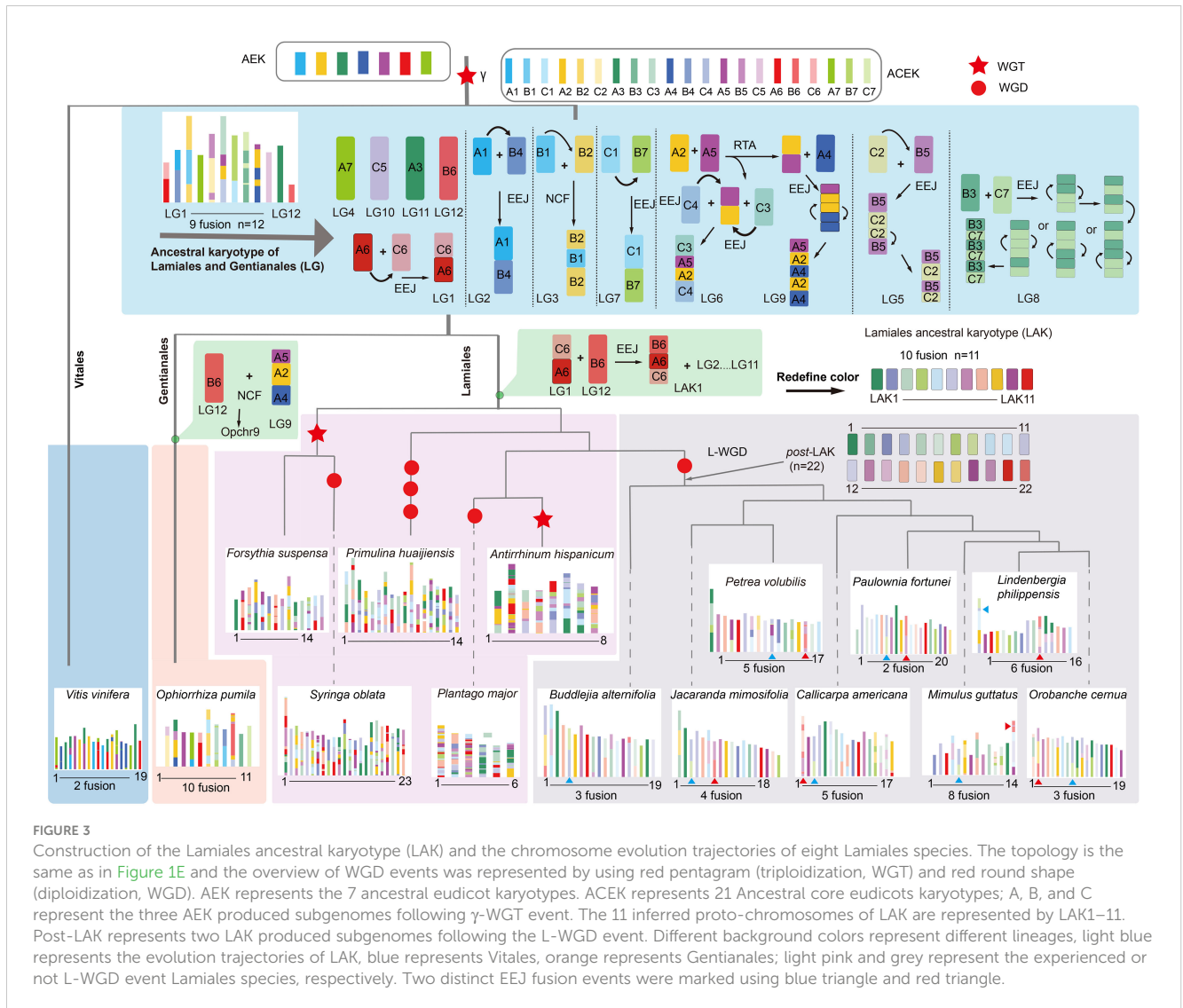


FIGURE 3

Construction of the Lamiales ancestral karyotype (LAK) and the chromosome evolution trajectories of eight Lamiales species. The topology is the same as in Figure 1E and the overview of WGD events was represented by using red pentagram (triploidization, WGT) and red round shape (diploidization, WGD). AEK represents the 7 ancestral eudicot karyotypes. ACEK represents 21 Ancestral core eudicots karyotypes; A, B, and C represent the three AEK produced subgenomes following  $\gamma$ -WGT event. The 11 inferred proto-chromosomes of LAK are represented by LAK1–11. Post-LAK represents two LAK produced subgenomes following the L-WGD event. Different background colors represent different lineages, light blue represents the evolution trajectories of LAK, blue represents Vitales, orange represents Gentianales; light pink and grey represent the experienced or not L-WGD event Lamiales species, respectively. Two distinct EEJ fusion events were marked using blue triangle and red triangle.

the evolutionary trajectory of the other ten proto-chromosomes of LAK was inferred in Figure 3.

Based on previous results, eight Lamiales species, including *B. alternifolia*, *P. volubilis*, *P. fortunei*, *J. mimosifolia*, *C. americana*, *M. guttatus*, *O. cumana* and *L. philippensis*, have been identified as sharing the L-WGD event. This makes them ideal candidates for exploring the evolutionary characteristics of the LAK. Following the L-WGD event, the LAK underwent duplication, resulting in the formation of 22 proto-chromosomes (post-LAK). To investigate the evolutionary characteristics of the post-LAK in these species, two sets of LAK generated from *L. philippensis* were used to represent post-LAK karyotype and labeled as post-LAK1 to post-LAK22. Two distinct EEJ fusion events were identified by analyzing the dot plot comparing these eight species with the post-LAK (Figure 3). The first EEJ fusion event, which involved post-LAK4 and post-LAK8, occurred in all eight species. In contrast, the second EEJ fusion event, involving post-LAK20 and post-LAK22, was only present in seven of these species, with *B. alternifolia* being the sole exception (Figure 3). Subsequently, the eight species separated and evolved with different chromosome evolutionary trajectories.

Given the non-dysploid chromosomal changes were prevalent, we focused primarily on depicting the dysploid chromosomal rearrangements in these species. In summary, *B. alternifolia* genome experienced three chromosomal fusions, consisting of two EEJ fusions and one NCF fusion, leading to the current chromosome number  $n=19$  (Figure 3; Supplementary Figure S38); *J. mimosifolia* genome experienced two EEJ fusions and two NCF fusions to form the current chromosome number  $n=18$  (Figure 3; Supplementary Figure S39); the *P. volubilis* genome experienced five chromosomal fusions, composing of four EEJ and one NCF, resulting in the current chromosome number  $n=17$  (Figure 3; Supplementary Figure S40); *C. americana* genome experienced three EEJ fusions, one NCF fusion and one EEJ or NCF fusion, leading to the current chromosome number  $n=17$  (Figure 3; Supplementary Figure S41); *P. fortunei* genome experienced the fewest karyotype change events to form the current chromosome number  $n=20$ , with just two EEJ fusions and no further karyotype evolutionary events (Figure 3; Supplementary Figure S42); *M. guttatus* genome experienced three EEJ fusions and five NCF fusions to form the current chromosome number  $n=14$  (Figure 3;

Supplementary Figure S43); *O. cumana* genome experienced two EEJ fusions and one NCF fusion to form the current chromosome number  $n=19$  (Figure 3; Supplementary Figure S44); *L. philippensis* genome experienced two EEJ fusions and four NCF fusions to form the current chromosome number  $n=16$  (Figure 3; Supplementary Figure S45).

## Comparative analyses of subgenomes in eight Lamiales species

Following polyploidization, most duplicated genes would experience drastic changes due to the sensitivity of dosage balance (Li et al., 2016). To elucidate the fractionation characteristics of duplicated genes in Lamiales, two sets of post-LAK subgenomes were initially classified as least fractionated (LF, 22A) and most fractionated (MF, 22B) based on their gene counts. Subsequently, 16 subgenomes were constructed using the WGDI, and their grouping was determined based on the collinear relationship with post-LAK. The one-to-one collinear correspondence of these subgenomes with post-LAK confirmed the reliability of these subgenomes (Figure 4A; Supplementary Figures S46–S52), making them suitable for further

research. Like the post-LAK, all subgenomes exhibited subgenome dominance. For example, 22A subgenomes (with 14,306 – 24,133 genes) had more gene counts than 22B subgenomes (with 9,364 – 16,727 genes) among these 16 subgenomes (Supplementary Table S13). The BUSCO analyses also showed that eight 22A subgenomes had over 50% complete BUSCO genes from the embryophyta\_10 dataset, whereas the completeness level in the eight 22B subgenomes was below the threshold of 50% (Figure 4B). This phenomenon suggested that these eight species exhibit consistently biased preservation and display a dominance within their respective subgenomes.

To uncover the fractionation pattern of the subgenomes, we utilized OrthoFinder (v2.5.2) (Emms and Kelly, 2019) to group their protein-coding genes into orthogroups, with *V. vinifera* and *O. pumila* as the reference. Stringent criteria were applied to choose representative orthogroups, necessitating orthogroups with a minimum of eight distinct subgenomes, encompassing *V. vinifera* and *O. pumila*. In total, 10,083 orthogroups were selected to comprise the core set of orthogroups (CSOs) for our further analyses. Based on the observed number of gene copies in each CSO, those CSOs were categorized into four distinct types: ‘Absent’ (no gene copies present), ‘Single Copy’ (one gene copy), ‘Two’

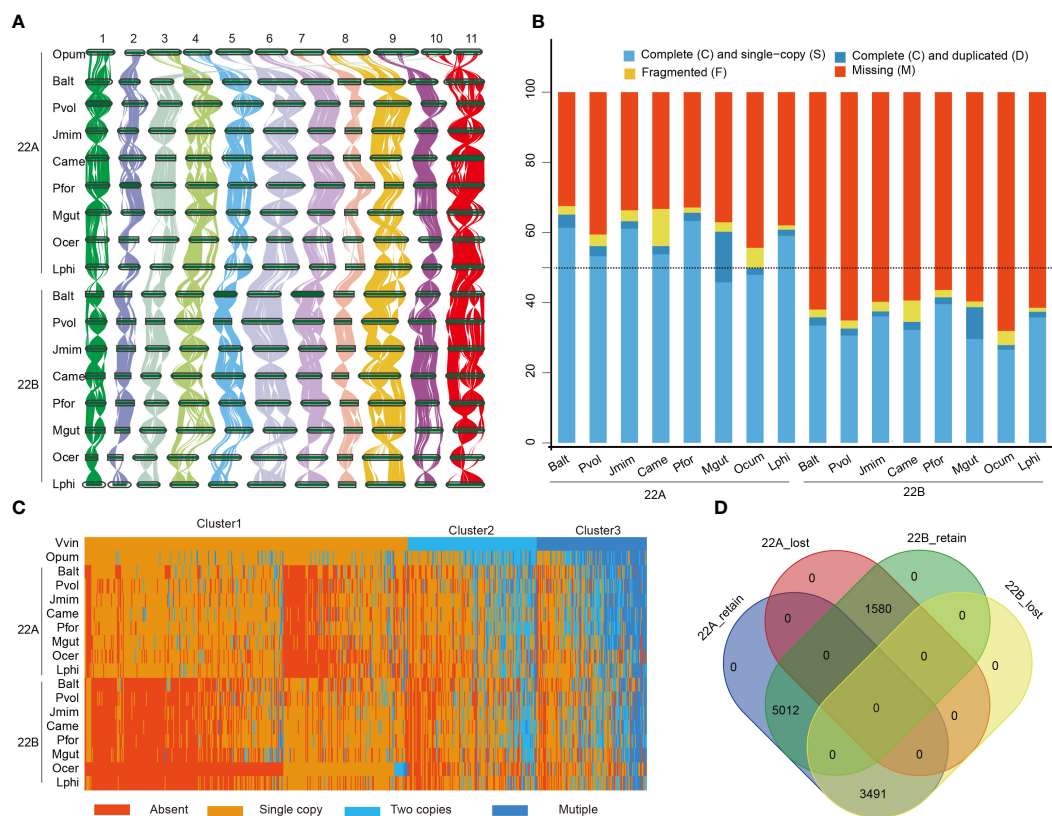


FIGURE 4

Comparative analysis of subgenomes of 8 species in the Lamiales. (A) The synteny plot across *O. pumila* (Opum) genome and sixteen subgenomes, *B. alternifolia* (Balt), *P. volubilis* (Pvol), *J. mimosifolia* (Jmim), *C. americana* (Came), *P. fortune* (Pfor), *M. guttatus* (Mgut), *O. cumana* (Ocum) and *L. philippensis* (Lphi); 22A represents least fractionated subgenome and 22B represents most fractionated subgenome. (B) Assessment of Benchmarking Universal Single-Copy Orthologs (BUSCOs) of those sixteen subgenomes with embryophyta\_10 (1614) databases. (C) Heat map of the clustered copy-number profile matrix in Opum, *V. vinifera* (Vvin), and sixteen subgenomes. Core gene families could be partitioned into four based on the clustering of the copy-number profile data. Rows represent species and columns represent the 10,083 CSOs. Gene families are sorted according to the three different clusters of Vvin. (D) Venn diagram showing the distribution of the retained and lost CSO sets.

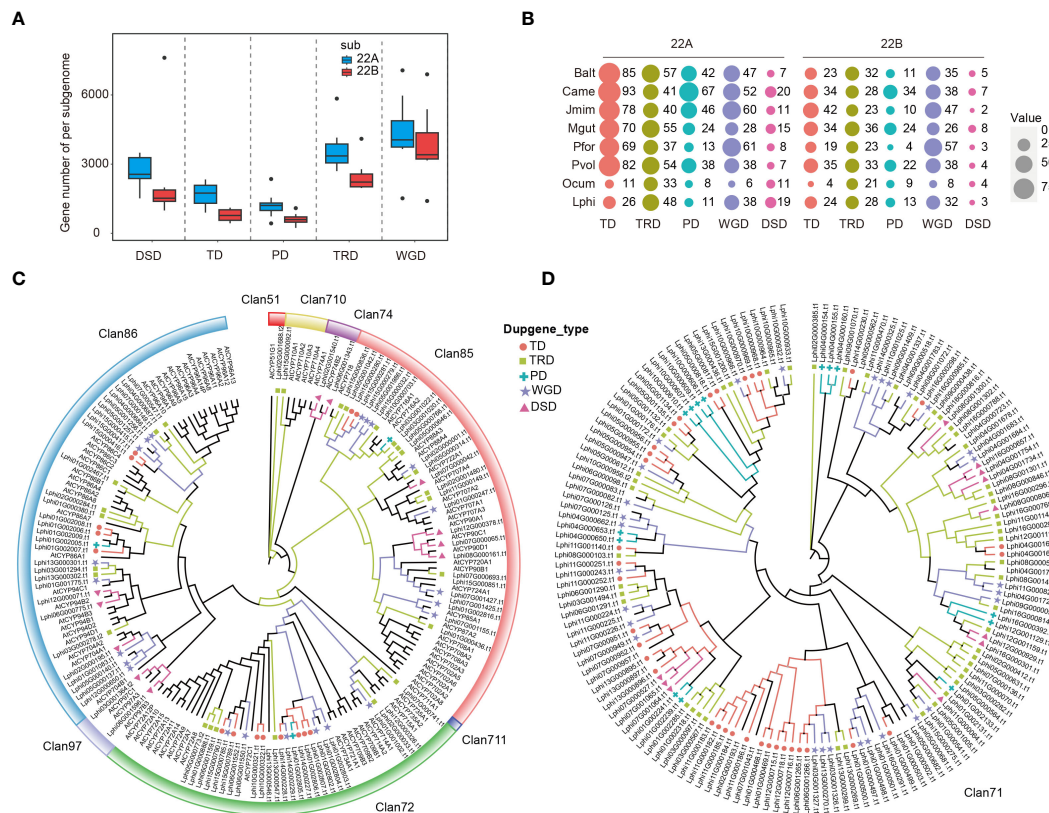
Copies' (exactly two gene copies), and 'Multiple' (more than two gene copies). Furthermore, according to the gene number in *V. vinifera*, those CSOs were organized into three clusters (cluster 1, cluster 2, and cluster 3) (Figure 4C). In cluster 1, CSOs mainly consisted of absent or singleton genes in 16 subgenomes (Supplementary Figure S53). In cluster 2, CSOs are composed of either absent or present genes in single or two-copy forms. In cluster 3, CSOs mainly consisted of orthogroups that are single, two, or multiple copies. In all three clusters, the number of absent CSOs in the 22B subgenome was significantly greater than that in the 22A subgenome ( $P < 0.05$ ). Conversely, the remaining three types showed the opposite trend, except for the single copy gene in cluster 3 (Supplementary Figure S53). Therefore, we speculated that the dominance of the subgenome could primarily originate from a higher frequency of loss and a lower rate of retention. Besides, our results also indicated that the distribution of CSOs across the 22A and 22B subgenome had a nested complementary profile, particularly evident in cluster 1.

We further defined the CSOs present in over 50% of the subgenome as retained CSO sets, while those not maintained are referred to as lost CSO sets. Based on these criteria, slightly over half of the CSO sets (5,071/10,083) displayed a complementary distribution across the two subgenomes, corroborating earlier

findings presented in the heat map analysis. Specifically, 3,491 CSOs were conserved exclusively in the 22A subgenome sets, 1,580 CSOs were solely retained in the 22B subgenome sets, and 5,012 CSOs were present in both 22A and 22B subgenomes (Figure 4D).

### Different modes of gene duplications driving the dominance of subgenome

In addition to WGD events, gene duplication is also a crucial process in expanding the gene family (Fajardo et al., 2023). To determine if gene duplication caused the dominance of subgenome, we conducted statistical analyses on non-WGD (Dispersed duplication, DSD, Tandem duplicate, TD; Proximal duplication, PD; and Transposed duplication, TRD) and WGD genes, as well as on unduplicated genes (UD) within those subgenomes. Our results indicated that all the modes of gene duplications were higher in 22A subgenome sets than in 22B subgenome sets (Figure 5A). Cytochrome P450s (CYPs) form the largest enzyme family in plants, representing around 1% of protein-coding genes in various flowering plants (Liu et al., 2023). They can be ideal candidates to study different modes of gene duplications. The distribution of CYP genes in the various



**FIGURE 5** The numbers and the distributions of CYP genes from different gene duplications. (A) Distribution of non-WGD (TRD, DSD, TD, PD, and TRD) and WGD genes across two sets of subgenomes in 16 subgenomes. (B) The proportions and numbers of CYP genes from different gene duplications are estimated in 16 subgenomes. The size of the colored round shapes represents the number, or proportions of all genes in each gene duplication mode. (C) The phylogenetic tree of 8 Clan CYP genes from *L. philippensis* and *A. thaliana*. (D) The phylogenetic tree of Clan71 CYP genes from *L. philippensis*.

modes of gene duplications showed more copies in the 22A subgenome sets than in the 22B subgenome sets across the 16 subgenomes (Figure 5B). Interestingly, there are more CYP genes in TD genes than in WGD genes, and the number of CYP genes in TRD and PD was also similar to that in WGD genes, despite their lower total gene count compared to WGD genes (Figure 5B). This observation suggests that, besides WGD events, the non-WGD genes also play a crucial role in the expansion of gene families. To detail the influence of the gene duplication on gene family expansion, the phylogenetic tree of CYP genes in *L. philippensis* genome were constructed and with *Arabidopsis thaliana* as references. In total, 242 CYP genes in *L. philippensis* genome were cluster into 9 subfamilies according to the result of Williams et al. (2000) (<http://p450.kvl.dk/p450.shtml>) (Figures 5C, D). Within the phylogenetic tree, the gene duplication modes are distinct among major subfamilies like Clan71, Clan72, Clan85, and Clan86. The remaining subfamilies only exhibit one type of gene duplication mode. Clan71, as the largest subfamily in the CYP superfamily, contains more gene duplication copies of various modes than other subfamilies (Figure 5D).

## Discussion

### Genome assembly of *Lindenbergia philippensis* provides an important genomic resource

*Lindenbergia philippensis* belongs to the tribe Lindenbergieae, besides the tribe Rehmannieae, and it is the closest autotrophic sister clade to all parasitic plant lineages in the family Orobanchaceae (Mutuku et al., 2021; Jiang et al., 2022; Xu et al., 2022) (Figure 1E). Here, the *L. philippensis* genome was achieved by combining Illumina paired-end sequencing data, Oxford Nanopore data and Hi-C data. The new genome assembly size was 407.46 Mb, close to the estimated size of 396.66 Mb via flow cytometry and 17-kmer frequency estimation (Supplementary Tables S2, S4). The completeness of the genome assembly was comparable with 15 species in Lamiales (Supplementary Table S6). Therefore, the assembly of *L. philippensis* genome had good quality, making it suitable for further analyses. Additionally, the anchored 16 pseudochromosomes had good intra-genomic collinear blocks (Note 1), which makes it the high-quality reference genome to deduce the karyotype evolutionary trajectory among relative species. These results provide important genomic resources for further genome study on *L. philippensis* as well as Orobanchaceae in the future.

### Combining Ks and syntenic depth analyses reconstruct the accurate evolutionary history of polyploidization and WGD events

Polyploidization, or WGD events, have been identified as a critical mechanism in facilitating species evolution and diversification across a vast majority of plant lineages (Zhang et al., 2019; Clo, 2022). Additionally, the profound influence of

WGD events goes beyond its initial occurrence, and could primarily serve as a catalyst to drive a subsequent PPD process (Soltis and Soltis, 2016; Zhang et al., 2019). However, the PPD process has negative effect on the identification of WGD events and the determination of polyploidization levels.

Currently, although an increasing number of WGD events are being reported through Ks or 4Dtv analyses, syntenic depth analyses, or a combination of these methods, some WGD events are inaccurately determined due to low-quality and limited genomic data and analytical method constraints. For instance, Feng et al. (2020) used Ks analysis to reveal a WGD (the *L event*) present in almost all Lamiales except the lineage of Oleaceae, which conflicted with the results of Zhu et al. (2023b). Zhu and his colleagues substantiated that the Plantaginaceae underwent a distinct WGD event, diverging from the shared *L event* (Feng et al., 2020; Zhu et al., 2023b). This independent WGD event was confirmed in this study, as well as a recent research by Huang et al. (2023). By combining Ks and inter-species syntenic depth analyses, we validated that *P. huaijiensis* experienced three diploidization events following the  $\gamma$ -WGT event, rather than two WGD events in the previous report (Feng et al., 2020). This discrepancy primarily derived from that Feng et al. (2020) relied on the solely Ks analysis to survey the WGD event, without integrating syntenic depth comparisons across different species. Additionally, two separate Ks values (0.87 and 1.12) (Figure 2A) suggested that *P. huaijiensis* underwent two WGD events within a relatively close timeframe. Consequently, these two WGD events could easily be overlooked and misinterpreted as a single event.

WGDI (Sun et al., 2022) and JCVI (Tang et al., 2008) are both popular software options for analyzing WGD events through syntenic depth analysis, but WGDI has advantages over JCVI in distinguishing the level of polyploidization. For example, *O. pumila* and *V. vinifera* had been shown to share the  $\gamma$ -WGT event, the syntenic depths or orthologous gene ratio between them should theoretically be 1:1, ignoring the non-WGD effects, whereas their syntenic depths were determined at 2:2 in the research of Rai et al. (2021) by using JCVI, which cannot identify whether they shared this WGD or not. In our study, the 1:1 orthologous gene ratio of *O. pumila* and *V. vinifera* was validated using WGDI and confirmed that they shared the  $\gamma$ -WGT event, which was aligned with the previous results (Wang et al., 2022c). Besides, the orthologous gene ratio of *P. huaijiensis* compared to *V. vinifera* was showed to be 6:1, corresponding to its three diploidization events. However, their orthologous gene ratio was 5:1 using JCVI, conflicting with its polyploidization history.

Overall, it is imprudent to crudely estimate polyploidization events based solely on the Ks distribution or syntenic depth analysis. While the analysis of Ks can indicate the occurrence of WGD events, it is challenging to clearly distinguish the polyploidization histories. Essentially, Ks analysis only reveals whether the species underwent WGD events, making it hard to ascertain whether the WGD event led to diploidization, triploidization, or other forms of polyploidization. This and previous studies have revealed some misunderstandings regarding the evolutionary history of WGD events, such as the genomic researches of *C. americana* (Hamilton et al., 2020), watermelon (Guo et al., 2013), black pepper (Hu et al., 2019), Olive (Ren et al., 2018), and *Prunus mongolica* (Zhu et al., 2023a). These mistakes

significantly increase the chance of misinterpreting the evolutionary history of these events, hindering our comprehensive understandings of the functional evolution of subgenomes, gene families, pathways, and genomic structures. Integrating genomic collinearity analysis with Ks information provides a more accurate and effective method for inferring polyploidization events, as supported by our findings in this study and previous studies (Kong et al., 2023; Sun et al., 2024). Based on this theoretical framework, Sun et al. (2022) have developed an integrated tool WGD I that combines functions for detecting WGD events, analyzing karyotype evolution, and constructing ancestral karyotypes, among other functions, providing an effective and more accurate method for the WGD events analyses. Using this tool, WGD events of 15 species in Lamiales were corrected and validated, providing significant insights for the analysis of WGD events. Moreover, the L-WGD shared by most Lamiales species was validated by combing the Ks and syntenic depth analyses.

## Construction and evolutionary trajectory of ancestral karyotypes in Lamiales

The identification and construction of ancestral karyotypes play a crucial role in confirming the phylogenetic positions of species and elucidating the impact of various polyploidy events on species diversity and evolution (Murat et al., 2017; Kong et al., 2023). The recursive dysploid or non-dysploid changes have reshuffled the ancestral karyotypes of Lamiales, complicating the clear interpretation of polyploidization events (Ren et al., 2018; Feng et al., 2020). In this study, *L. philippensis* was used to construct the LAK, following the theoretical framework that suggested by Sun et al. (2022), consisting of 11 proto-chromosomes. The two complete copies of the paleogenome within the *P. fortunei* genome validated its reliability (Supplementary Figure S42).

The evolutionary path analyses of LAK and post-LAK showed that the base number deduction of chromosomes was caused by fusions (Wang et al., 2022c; Feng et al., 2024). This suggested that descending dysploidy may play a major role in karyotype evolution after WGD events, consistent with previous studies indicating that the chromosomal evolution in land plants is mostly characterized by descending dysploidy (Carta et al., 2020; Wang et al., 2022c; Kong et al., 2023). Two distinct EEJ fusion events were detected in those eight species, the first fusion shared by all studied species, while the second fusion event was observed in seven of the eight species, with *B. alternifolia* as the notable exception. This divergence may be a significant factor for its speciation from the other species. This finding also indicates that the PPD process plays a significant role in promoting species diversification. Usually, the reduction of chromosome number critically resulted in the abnormal pairing of gametes, ultimately leading to reproductive isolation (Paliulis and Nicklas, 2000; Luo et al., 2018). Additionally, the eight species showed a lower frequency of non-dysploidy alterations, with dysploidy changes being easily identifiable (Figure 3). Interestingly, a higher frequency of EEJ fusion compared to NCF fusion was observed in most species, suggesting that EEJ fusion may have a competitive advantage over

NCF fusion in the process of karyotype evolution. While a similar phenomenon was also reported in previous studies (Wang et al., 2022a), the reliability of this advantage is still an understudied topic. The construction of the LAK and the elucidation of its evolutionary trajectory address a significant gap in our understanding of chromosome karyotype evolution within Lamiales. Furthermore, the discovery revealed that the genomes of the eight karyotype-conserved species possess more complete ancestral chromosomal structures, which suggests their potential as model organisms for future genomic research in Lamiales.

## Genomic fractionation and the role of different modes of gene duplications in driving genome evolution

Following polyploidization events, extensive chromosome rearrangements and large-scale gene loss are prevalent due to the dosage balance, particularly in allopolyploids. In this study, following the construction of the post-LAK, we constructed two sets of subgenomes for the eight representative species, respectively. The subgenomes display biased preservation and subgenome dominance, aligning with the lineage-specific hexaploidization seen in *Lupinus* (Xu et al., 2020). This indicated that the L-WGD event may be an allopolyploid event. After observing the fractionation pattern of duplicated genes in these species, we hypothesized that plant species had undergone WGD events that tend to selectively retain these genes within subgenomes in a complementary manner. This suggests that species that underwent WGD events may optimize their genetic repertoire to achieve a more adaptable genetic system in response to changing environments. REs play important roles in driving genome evolution and regulating gene expression (Kubis et al., 1998; Novák et al., 2020; Liao et al., 2023). In this study, we confirmed that the expansion of REs is a key factor influencing genome size variations, which is consistent with some previous studies, and besides polyploidization and gene duplications, repeat expansion was the main factor in amplifying the genome size (Nishihara, 2019; Shao et al., 2019; Novák et al., 2020; de Lima and Ruiz-Ruano, 2022).

Besides, the investigation of different modes of gene duplications across 16 subgenomes revealed that the subgenome 22A exhibited a higher number of duplicate genes than subgenome 22B. This phenomenon shows that gene duplication may play important roles in driving subgenome dominance. Distribution of gene duplication modes across several larger subfamilies in the phylogenetic tree of the *L. philippensis* CYPs superfamily. This diverse distribution also indicates the duplicated gene as a significant force in expanding the gene family (Liu et al., 2023).

## Data availability statement

The raw genome sequencing data of *L. philippensis* are available at the National Genomics Data Center (<https://ngdc.cnpc.ac.cn/>) under BioProject number PRJCA010538 (CRA013614). All data are available from the corresponding author upon request.

## Author contributions

B-ZC: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation. D-WL: Writing – review & editing, Formal analysis. K-YL: Writing – review & editing. S-TJ: Writing – review & editing. XD: Writing – review & editing. W-BW: Writing – review & editing. X-ZL: Writing – review & editing. T-TH: Writing – review & editing. Y-HL: Writing – review & editing. D-ZG: Writing – review & editing. X-TL: Writing – review & editing. S-CD: Writing – review & editing. Y-FZ: Writing – review & editing. WC: Writing – review & editing. YD: Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing – review & editing. W-BY: Conceptualization, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (31870196, 32371700), the Key Programs of Yunnan Province, China (202101BC070003), the Hainan Province Science and Technology Special Fund (ZDYF2023RDYL01), the Hainan Institute of National Park Fund (KY-24ZK02), and Yunnan Revitalization Talent Support Program “Young Talent” and “Innovation Team” Projects.

## References

- Aggarwal, G., and Ramaswamy, R. (2002). Ab initio gene identification: prokaryote genome annotation with GeneScan and GLIMMER. *J. Biosci.* 27, 7–14. doi: 10.1007/bf02703679
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/s0022-2836(05)80360-2
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573
- Birchler, J. A., and Han, F. (2018). Barbara McClintock’s unsolved chromosomal mysteries: parallels to common rearrangements and karyotype evolution. *Plant Cell* 30, 771–779. doi: 10.1105/tpc.17.00989
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinf. (Oxford England)* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Carta, A., Bedini, G., and Peruzzi, L. (2020). A deep dive into the ancestral chromosome number and genome size of flowering plants. *New Phytol.* 228, 1097–1106. doi: 10.1111/nph.16668
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* doi: 10.1002/0471250953.bi0410s05
- Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* 4, 258–268. doi: 10.1038/s41477-018-0136-7
- Cheng, T. C., Wu, J., Wu, Y., Chilukuri, R. V., Huang, L., Yamamoto, K., et al. (2017). Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat. Ecol. Evol.* 1, 1747–1756. doi: 10.1038/s41559-017-0314-4
- Clark, J. W., and Donoghue, P. C. J. (2017). Constraining the timing of whole genome duplication in plant evolutionary history. *Proc. Biol. Sci.* 284, 20170912. doi: 10.1098/rspb.2017.0912
- Clo, J. (2022). Polyploidization: Consequences of genome doubling on the evolutionary potential of populations. *Am. J. Bot.* 109, 1213–1220. doi: 10.1002/ajb2.16029
- Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 6, 836–846. doi: 10.1038/nrg1711
- Cui, L. Y., Wall, P., Leebens-Mack, J., Lindsay, B., Soltis, D., Doyle, J., et al. (2006). Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16, 738–749. doi: 10.1101/gr.4825606
- Damas, J., Kim, J., Farre, M., Griffin, D. K., and Larkin, D. M. (2018). Reconstruction of avian ancestral karyotypes reveals differences in the evolutionary history of macro- and microchromosomes. *Genome Biol.* 19, 155. doi: 10.1186/s13059-018-1544-8
- Danecek, P., Bonfield, J., Liddle, J., Marshall, J., Ohan, V., Pollard, M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. doi: 10.1093/gigascience/giab008
- de Lima, L. G., and Ruiz-Ruano, F. J. (2022). In-depth satelitome analyses of 37 drosophila species illuminate repetitive DNA evolution in the drosophila genus. *Genome Biol. Evol.* 14, evac064. doi: 10.1093/gbe/evac064
- Dodsworth, S., Chase, M. W., and Leitch, A. R. (2015). Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Botanical J. Linn. Soc.* 180, 1–5. doi: 10.1111/boj.12357%[BotanicalJournaloftheLinneanSociety
- Dudchenko, O., Batra, S., Omer, A., Nyquist, S., Hoeger, M., Durand, N., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Sci. (New York N.Y.)* 356, 92–95. doi: 10.1126/science.aal3327

## Acknowledgments

We are grateful to Mr. Bing Hao and Ms. Shuang Ye for their help in collecting samples. We are grateful to Ms. Yun-bin Pan for his help in data analysis.

## Conflict of interest

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

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1444234/full#supplementary-material>

- Durand, N. C., Shamim, S., Machol, I., Rao, S., Huntley Miriam, H., Lander, E., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Syst.* 3, 95–98. doi: 10.1016/j.cels.2016.07.002
- Edger, P. P., Smith, R., McKain, R., Cooley, A., Vallejo-Marin, M., Yuan, Y.-W., et al. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* 29, 2150–2167. doi: 10.1105/tpc.17.00010
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinf.* 9, 18. doi: 10.1186/1471-2105-9-18
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y
- Fajardo, D., Saint Jean, R., and Lyons, P. J. (2023). Acquisition of new function through gene duplication in the metallocarboxypeptidase family. *Sci. Rep.* 13, 2512. doi: 10.1038/s41598-023-29800-9
- Feng, C., Wang, J., Wu, L., Kong, H., Yang, L., Feng, C., et al. (2020). The genome of a cave plant, *Primulina huaijensis*, provides insights into adaptation to limestone karst habitats. *New Phytol.* 227, 1249–1263. doi: 10.1111/nph.16588
- Feng, Y., Wang, Z., Xiao, Q., Teng, J., Wang, J., Yu, Z., et al. (2024). A likely paleo-autotetraploidization event shaped the high conservation of Nysaceae genome. *Hortic. Plant J.* doi: 10.1016/j.hpj.2022.09.010
- Finn, R. D., Clements, J., and Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 39, W29–W37. doi: 10.1093/nar/gkr367
- Grabherr, M. G., Haas, B., Yassour, M., Levin, J., Thompson, D., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W. J., Zhang, H., et al. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* 45, 51–58. doi: 10.1038/ng.2470
- Haas, B. J., Delcher, A., Mount, S., Wortman, J., Smith, R., Hannick, L., et al. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666. doi: 10.1093/nar/gkg770
- Haas, B. J., Salzberg, S., Zhu, W., Pertea, M., Allen, J., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7. doi: 10.1186/gb-2008-9-1-r7
- Hamilton, J. P., Godden, G., Lanier, T. E., Bhat, W. W., Kinser, T. J., Vaillancourt, B., et al. (2020). Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing Lamiaceae species, *Callicarpa americana*. *GigaScience* 9, g1aa093. doi: 10.1093/gigascience/g1aa093
- Hoang, N. V., Sogbohossou, O., Xiong, W., Simpson, C., Singh, P., Walden, N., et al. (2023). The Gynandropsis gynandra genome provides insights into whole-genome duplications and the evolution of C4 photosynthesis in Cleomaceae. *Plant Cell* 35, 1334–1359. doi: 10.1093/plcell/koad018
- Hu, L., Xu, Z., Wang, M., Fan, R., Yuan, D., Wu, B., et al. (2019). The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat. Commun.* 10, 4702. doi: 10.1038/s41467-019-12607-6
- Huang, H., Wang, C., Pei, S., and Wang, Y. (2023). A chromosome-level reference genome of an aromatic medicinal plant *Adenosma buchneroides*. *Sci. Data* 10, 660. doi: 10.1038/s41597-023-02571-8
- Jiang, N., Dong, L. N., Yang, J. B., Tan, Y., Wang, H., Randle, C., et al. (2022). Herbarium phylogenomics: Resolving the generic status of the enigmatic *Pseudobartsia* (Orobanchaceae). *J. Systematics Evol.* 60, 1218–1228. doi: 10.1111/jse.12829
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature* 473, 97–100. doi: 10.1038/nature09916
- Julca, I., Marcet-Houben, M., Vargas, P., and Gabaldon, T. (2018). Phylogenomics of the olive tree (*Olea europaea*) reveals the relative contribution of ancient allo- and autopolyploidization events. *BMC Biol.* 16, 15. doi: 10.1186/s12915-018-0482-y
- Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends genetics: TIG* 16, 418–420. doi: 10.1016/s0168-9525(00)02093-x
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Karoonuthaisiri, N., Anghong, P., Uengwetwanit, T., Pootakham, W., Sittikankaw, K., Sonthirod, C., et al. (2020). Optimization of high molecular weight DNA extraction methods in shrimp for a long-read sequencing platform. *PeerJ* 8, e10340. doi: 10.7717/peerj.10340
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kellogg, E. A. (2016). Has the connection between polyploidy and diversification actually been tested? *Curr. Opin. Plant Biol.* 30, 25–32. doi: 10.1016/j.pbi.2016.01.002
- Kong, X., Zhang, Y., Wang, Z., Bao, S., Feng, Y., Wang, J., et al. (2023). Two-step model of paleohexaploidy, ancestral genome reshuffling and plasticity of heat shock response in Asteraceae. *Horticulture Res.* 10, uhad073. doi: 10.1093/hr/uhad073
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinf.* 5, 59. doi: 10.1186/1471-2105-5-59
- Kubis, S., Schmidt, T., and Heslop-Harrison, J. S. (1998). Repetitive DNA elements as a major component of plant genomes. *Ann. Bot.* 82, 45–55. doi: 10.1006/anbo.1998.0779
- Landis, J. B., Soltis, D., Li, Z., Marx, H., Barker, M., Tank, D., et al. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *Am. J. Bot.* 105, 348–363. doi: 10.1002/ajb2.1060
- Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52, W78–W82. doi: 10.1093/nar/gkac268
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinf. (Oxford England)* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., De Smet, R., et al. (2016). Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* 28, 326–344. doi: 10.1105/tpc.15.00877
- Li, X., Feng, T., Randle, C., and Schneeweiss, G. (2019). Phylogenetic relationships in orobanchaceae inferred from low-copy nuclear genes: consolidation of major clades and identification of a novel position of the non-photosynthetic orobanche clade sister to all other parasitic orobanchaceae. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00902
- Liao, X., Zhu, W., Zhou, J., Li, H., Xu, X., Zhang, B., et al. (2023). Repetitive DNA sequence detection and its role in the human genome. *Commun. Biol.* 6, 954. doi: 10.1038/s42003-023-05322-y
- Liu, X., Gong, Q., Zhao, C., Wang, D., Ye, X., Zheng, G., et al. (2023). Genome-wide analysis of cytochrome P450 genes in *Citrus clementina* and characterization of a CYP gene encoding flavonoid 3'-hydroxylase. *Horticulture Res.* 10, uhac283. doi: 10.1093/hr/uhac283
- Lovell, J. T., MacQueen, A. H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J. D., et al. (2021). Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* 590, 438–43+. doi: 10.1038/s41586-020-03127-1
- Luo, J., Sun, X., Cormack, B. P., and Boeke, J. D. (2018). Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. *Nature* 560, 392–396. doi: 10.1038/s41586-018-0374-x
- Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinf. (Oxford England)* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315
- Mandáková, T., MacQueen, A. H., Mamidi, S., Bonnette, J., Jenkins, J., and Napier, J. D. (2017). Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Mol. Ecol.* 26, 6445–6462. doi: 10.1111/mec.14379
- Mandáková, T., and Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysplid changes. *Curr. Opin. Plant Biol.* 42, 55–65. doi: 10.1016/j.pbi.2018.03.001
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinf. (Oxford England)* 27, 764–770. doi: 10.1093/bioinformatics/btr011
- Mayrose, I., and Lysak, M. A. (2021). The evolution of chromosome numbers: mechanistic models and experimental approaches. *Genome Biol. Evol.* 13, evaa220.
- Mendes, F. K., Vanderpool, D., Fulton, B., and Hahn, M. W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinf. (Oxford England)* 36, 5516–5518. doi: 10.1093/bioinformatics/btaa1022
- Morin, S. J., Eccles, J., Iturriaga, A., and Zimmerman, R. S. (2017). Translocations, inversions and other chromosome rearrangements. *Fertility sterility* 107, 19–26.
- Mulcahy, D. G., Noonan, B., Moss, T., Townsend, T., Reeder, T., Sites, J. J., et al. (2012). Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Mol. Phylogenet. Evol.* 65, 974–991. doi: 10.1016/j.ympev.2012.08.018
- Murat, F., Armero, A., Pont, C., Klopp, C., and Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat. Genet.* 49, 490–496. doi: 10.1038/ng.3813
- Mutuku, J. M., Cui, S., Yoshida, S., and Shirasu, K. (2021). Orobanchaceae parasitism interactions. *New Phytol.* 230, 46–59. doi: 10.1111/nph.17083
- Nishihara, H. (2019). Transposable elements as genetic accelerators of evolution: contribution to genome size, gene regulatory network rewiring and morphological innovation. *Genes Genet. Syst.* 94, 269–281. doi: 10.1266/ggs.19-00029
- Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížková, A., et al. (2020). Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nat. Plants* 6, 1325–1329. doi: 10.1038/s41477-020-00785-x
- Ou, S., and Jiang, N. (2018). LTR\_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310
- Ou, S., and Jiang, N. (2019). LTR\_FINDER\_parallel: parallelization of LTR\_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mobile DNA* 10, 48. doi: 10.1186/s13100-019-0193-0
- Padmarasu, S., Himmelbach, A., Mascher, M., and Stein, N. (2019). *In situ* hi-C for plants: an improved method to detect long-range chromatin interactions. *Methods Mol. Biol. (Clifton N.J.)* 1933, 441–472. doi: 10.1007/978-1-4939-9045-0\_28



- Paliulis, L. V., and Nicklas, R. B. (2000). The reduction of chromosome number in meiosis is determined by properties built into the chromosomes. *J. Cell Biol.* 150, 1223–1232. doi: 10.1083/jcb.150.6.1223
- Paterson, A. H., Bowers, J. E., and Chapman, B. A. (2004). Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *PNAS* 101, 9903–9908. doi: 10.1073/pnas.0307901101
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448
- Pryszcz, L. P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 44, e113. doi: 10.1093/nar/gkw294
- Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20, 38. doi: 10.1186/s13059-019-1650-2
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinf. (Oxford England)* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Rai, A., Hirakawa, H., Nakabayashi, R., Kikuchi, S., Hayashi, K., Rai, M., et al. (2021). Chromosome-level genome assembly of *Ophiorrhiza pumila* reveals the evolution of camptothecin biosynthesis. *Nat. Commun.* 12, 405. doi: 10.1038/s41467-020-20508-2
- Ranallo-Benavidez, T. R., Jaron, K. S., and Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* 11, 1432. doi: 10.1038/s41467-020-14998-3
- Rastogi, S., and Ohri, D. (2019). Karyotype evolution in cycads. *Nucleus* 63, 131–141. doi: 10.1007/s13237-019-00302-2
- Ren, R., Wang, H., Guo, C., Zhang, N., Zeng, L., Chen, Y., et al. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* 11, 414–428. doi: 10.1016/j.molp.2018.01.002
- Ren, L. H., Huang, W., and Cannon, T. E. (2019). Reconstruction of ancestral genome reveals chromosome evolution history for selected legume species. *New Phytol.* 223, 2090–2103. doi: 10.1111/nph.15770
- Schubert, I., and Lysak, M. A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* 27, 207–216. doi: 10.1016/j.tig.2011.03.004
- Sensalari, C., Maere, S., and Lohaus, R. (2022). ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. *Bioinf. (Oxford England)* 38, 530–532. doi: 10.1093/bioinformatics/btab602
- Seppy, M., Manni, M., and Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol. (Clifton N.J.)* 1962, 227–245. doi: 10.1007/978-1-4939-9173-0\_14
- Shao, F., Han, M., and Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Sci. Rep.* 9, 15399. doi: 10.1038/s41598-019-51888-1
- Soltis, D. E., Albert, V., Leebens-Mack, J., Bell, C., Paterson, A., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079
- Soltis, P. S., and Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.cpb.2016.03.015
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntactically mapped cDNA alignments to improve *de novo* gene finding. *Bioinf. (Oxford England)* 24, 637–644. doi: 10.1093/bioinformatics/btn013
- Stebbins, G. L. Jr. (1947). Types of polyploids; their classification and significance. *Adv. Genet.* 1, 403–429. doi: 10.1016/s0065-2660(08)60490-3
- Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., et al. (2022). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* 15, 1841–1851. doi: 10.1016/j.molp.2022.10.018
- Sun, Y., Liu, Y., Shi, J., Wang, L., Liang, C., Yang, J., et al. (2023). Biased mutations and gene losses underlying diploidization of the tetraploid broomcorn millet genome. *Plant J.* 113, 787–801. doi: 10.1111/tpj.16085
- Sun, P., Lu, Z., Wang, Z., Wang, S., Zhao, K., Mei, D., et al. (2024). Subgenome-aware analyses reveal the genomic consequences of ancient allopolyploid hybridizations throughout the cotton family. *Proc. Natl. Acad. Sci.* 121, e2313921121. doi: 10.1073/pnas.2313921121
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315
- Tang, H., Bowers, J., Wang, X., Ming, R., Alam, M., and Paterson, A. (2008). Synteny and collinearity in plant genomes. *Sci. (New York N.Y.)* 320, 486–488. doi: 10.1126/science.1153917
- Tank, D. C., Eastman, J. M., Pennell, M. W., Soltis, P. S., Soltis, D. E., Hinchliff, C. E., et al. (2015). Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytol.* 207, 454–467. doi: 10.1111/nph.13491
- Van de Peer, Y., Mizrahi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* 18, 411–424. doi: 10.1038/nrg.2017.26
- Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* 24, 1334–1347. doi: 10.1101/gr.168997.113
- Wang, X., Jin, D., Wang, Z., Guo, H., Zhang, L., Wang, L., et al. (2015). Telomeric genome repatterning determines recurring chromosome number reductions during the evolution of eukaryotes. *New Phytol.* 205, 378–389. doi: 10.1111/nph.12985
- Wang, J. Q., Yuan, M., Feng, Y., Zhang, Y., Bao, S., Hao, Y., et al. (2022a). A common whole-genome paleotetraploidization in Cucurbitales. *Plant Physiol.* 190, 2430–2448. doi: 10.1093/plphys/kiac410
- Wang, L. F., Sun, X., Peng, Y., Chen, K., Wu, S., Guo, Y., et al. (2022b). Genomic insights into the origin, adaptive evolution, and herbicide resistance of *Leptochloa chinensis*, a devastating tetraploid weedy grass in rice fields. *Mol. Plant* 15, 1045–1058. doi: 10.1016/j.molp.2022.05.001
- Wang, Z. Y., Li, Y., Sun, P., Zhu, M., Wang, D., Lu, Z., et al. (2022c). A high-quality *Buxus austro-yunnanensis* (Buxales) genome provides new insights into karyotype evolution in early eudicots. *BMC Biol.* 20, 216. doi: 10.1186/s12915-022-01420-1
- Wang, X. Y., Shi, X. L., Hao, B. L., Ge, S., and Luo, J. C. (2005). Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol.* 165, 937–946. doi: 10.1111/j.1469-8137.2004.01293.x
- Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F., and McRee, D. E. (2000). Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell* 5, 121–131. doi: 10.1016/S1097-2765(00)80408-6
- Wu, S. D., Han, B. C., and Jiao, Y. N. (2020). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* 13, 59–71. doi: 10.1016/j.molp.2019.10.012
- Xu, W., Zhang, Q., Yuan, W., Xu, F., Muhammad Aslam, M., Miao, R., et al. (2020). The genome evolution and low-phosphorus adaptation in white lupin. *Nat. Commun.* 11, 1069. doi: 10.1038/s41467-020-14891-z
- Xu, Y., Zhang, J., Ma, C., Lei, Y., Shen, G., Jin, J., et al. (2022). Comparative genomics of orobanchaceous species with different parasitic lifestyles reveals the origin and stepwise evolution of plant parasitism. *Mol. Plant* 15, 1384–1399. doi: 10.1016/j.molp.2022.07.007
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Zhang, K., Wang, X., and Cheng, F. (2019). Plant polyploidy: origin, evolution, and its influence on crop domestication. *Hortic. Plant J.* 5, 231–239. doi: 10.1016/j.hpj.2019.11.003
- Zhao, Q. Z., Meng, Y., Wang, P., Qin, X., Cheng, C., Zhou, J., et al. (2021). Reconstruction of ancestral karyotype illuminates chromosome evolution in the genus *Cucumis*. *Plant J.* 107, 1243–1259. doi: 10.1111/tpj.15381
- Zhu, Q., Wang, Y., Yao, N., Ni, X., Wang, C., Wang, M., et al. (2023a). Chromosome-level genome assembly of an endangered plant *Prunus mongolica* using PacBio and Hi-C technologies. *DNA Res.* 30, dsad012. doi: 10.1093/dnares/dsad012
- Zhu, S., Zhang, Y., Copsy, L., Han, Q., Zheng, D., Coen, E., et al. (2023b). The snapdragon genomes reveal the evolutionary dynamics of the S-locus supergene. *Mol. Biol. Evol.* 40, msad080. doi: 10.1093/molbev/msad080
- Zhuang, W. J., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–864. doi: 10.1038/s41588-019-0402-2