Genomics Communications

ARTICLE

Open Access

https://doi.org/10.48130/gcomm-0026-0014

Genomics Communications **2026**, 3: e015

Gapless genome assembly and evolutionary analysis of *Cnidium monnieri* (Apiaceae)

Yang Liu[1#], Ling Zhao[2,3#], Jiaxin Zhang[1], Qian Ju[4], Xurong Fan[1,5], Zhuo Li[2], Xingrui Zhang[2], Xiaoyi Liang[2], Feng Ge[2*] and Jinfeng Chen[1*]

[1] *State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China*
[2] *Shandong Key Laboratory for Green Prevention and Control of Agricultural Pests, Institute of Plant Protection, Shandong Academy of Agricultural Sciences, Jinan 250100, China*
[3] *College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China*
[4] *Shandong Peanut Research Institute, Qingdao 266100, China*
[5] *College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*
[#] Authors contributed equally: Yang Liu, Ling Zhao
* Correspondence: gef@ioz.ac.cn (Ge F); chenjinfeng@ioz.ac.cn (Chen J)

Abstract

As a member of the Apiaceae family, *Cnidium monnieri* holds significant medicinal and ecological value. It is highly regarded not only for its rich content of pharmacologically active coumarins but also for its role in sustainable pest control by preserving natural predator populations. Despite its importance, comprehensive genomic data for this species have been limited. In this study, we present a gapless, telomere-to-telomere genome assembly of *C. monnieri*. The final assembly has a total size of 1.13 Gb, organized into 10 distinct chromosomes. Evaluation metrics confirm the assembly's exceptional precision and continuity, with telomeric repeats successfully anchored at all 20 chromosomal ends. We annotated 39,218 protein-coding genes and characterized the repeat landscape, which is dominated by long terminal repeat retrotransposons (51.45%). We then reconstructed the phylogeny of Apiaceae and found that *C. monnieri* is sister to *Peucedanum praeruptorum* and *Saposhnikovia divaricata*. Comparative synteny analyses revealed that species with 11 chromosomes retain an ancestral conserved Apiaceae karyotype, whereas the 10-chromosome *C. monnieri* exhibits substantial lineage-specific interchromosomal rearrangements. Our study provides a high-quality genomic resource to support future studies in *C. monnieri* research and sheds light on chromosome evolution within the Apiaceae.

Keywords: *Cnidium monnieri*, Gapless genome, Apiaceae phylogeny, Chromosome evolution

Citation: Liu Y, Zhao L, Zhang J, Ju Q, Fan X, et al. 2026. Gapless genome assembly and evolutionary analysis of *Cnidium monnieri* (Apiaceae). *Genomics Communications* 3: e015 https://doi.org/10.48130/gcomm-0026-0014

Introduction

Cnidium monnieri (L.) is an annual herb widely distributed across China and serves as a highly significant traditional Chinese medicine[1]. Its fruit, known as "She Chuang Zi", has been used for centuries to treat dermatological and gynecological conditions[1]. Modern phytochemical and pharmacological studies have revealed that the plant is rich in diverse active constituents, such as coumarins, chromones, volatile oils, glycosides, and terpenoids. A large proportion of these metabolites display notable biological activities, including antiosteoporotic, antiviral, and anti-inflammatory effects[2–7].

Beyond its medicinal value, *C. monnieri* is ecologically crucial. As a nectar- and pollen-rich umbellifer, *C. monnieri* supports beneficial arthropods, including Coccinellidae (ladybird beetles), Chrysopidae (lacewings), and Syrphidae (hoverflies), by providing supplemental food resources and shelter[8]. Field studies have shown that planting *C. monnieri* flower strips along the edges of wheat (*Triticum aestivum*)–maize (*Zea mays*) rotation fields creates "bridge habitats" that sustain ladybird beetle populations during the wheat harvest period and facilitate their migration into the adjacent maize fields, thereby enhancing aphid suppression[9]. When infested by aphids, *C. monnieri* releases β-myrcene, a key herbivore-induced plant volatile (HIPV). This terpene specifically attracts the predatory ladybird beetle *Harmonia axyridis* by binding to the beetle's odorant-binding proteins (HaxyOBPs), thereby mediating tri-trophic plant–aphid–predator interactions and promoting natural pest regulation[10]. These characteristics make *C. monnieri* an attractive candidate for ecological pest management strategies aimed at enhancing natural enemy conservation and reducing dependence on chemical pesticides.

Despite this dual importance, genomic research on *C. monnieri* has been limited, hindering efforts to fully elucidate the regulatory and evolutionary mechanisms underlying coumarin biosynthesis and the genetic basis of its ecological interactions. A recently published chromosome-level genome of *C. monnieri*[11] provides an initial foundation; however, a gapless assembly and high-confidence genome evolution analyses are still lacking.

Here, we present a telomere-to-telomere (T2T) genome assembly for *C. monnieri*. By integrating complementary long-read sequencing technologies with high-throughput chromosome conformation capture (Hi-C) scaffolding, we achieved a substantial improvement in the genome's continuity and completeness. Through phylogenetic reconstruction and large-scale synteny analyses, we clarified the chromosomal evolution across the Apiaceae family. This establishes a high-quality genomic foundation for understanding trait diversification in a species of both medicinal and ecological significance.

https://doi.org/10.48130/gcomm-0026-0014

Methods

Genomic DNA isolation and sequencing

High-quality genomic DNA was extracted and purified from the young leaves of *C. monnieri* via an optimized cetyltrimethylammonium bromide (CTAB) protocol. The paired-end sequencing libraries (150 bp) were prepared and subjected to sequencing on the Illumina NovaSeq X Plus platform (Illumina Inc., San Diego, CA, USA). For long-read sequencing, PacBio High-Fidelity (HiFi) and Oxford Nanopore Technologies (ONT) libraries were constructed from the young leaf tissues following the standard protocols of Pacific Biosciences Co. (Menlo Park, CA, USA) and Oxford Nanopore Technologies (Oxford, UK), respectively. The PacBio Revio platform was utilized to produce the HiFi long reads, whereas the ONT ultra-long reads were generated from the PromethION platform. To generate the Hi-C library, fresh seedling leaves of *C. monnieri* were harvested and treated with 2% formaldehyde under a vacuum to crosslink the chromatin. Glycine was then applied to halt the crosslinking reaction. After purification of the nuclei, chromatin was digested utilizing 100 units of the DpnII restriction enzyme, followed by biotin-14-deoxyadenosine triphosphate (dATP) labeling. The religated DNA was subsequently fragmented to a size range of 300–700 bp. These fragments underwent end-repair, A-tailing, and purification steps. Ultimately, the resulting Hi-C library was measured for concentration and sequenced on the Illumina NovaSeq X Plus platform.

Genome survey

We evaluated the genome size of *C. monnieri* by conducting a k-mer analysis based on the Illumina paired-end sequencing data. Specifically, KMC v3.2.1[12] was applied to determine the frequency of the k-mers. The frequency profile (k = 17) was subsequently fed into GenomeScope v1.0[13] to predict the overall genome size.

RNA-seq sequencing

Samples from five tissues (root, stem, leaf, flower, and fruit) were harvested from *C. monnieri*. Total RNA was isolated from these tissues utilizing TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA). RNA sequencing libraries were then constructed according to the standard Illumina protocol and paired-end reads sequenced on the Illumina NovaSeq X Plus platform.

Genome assembly, gap-filling, and evaluation

To assemble PacBio HiFi reads, we used Hifiasm v0.19.3[14] with the parameters '-l3 -u1'. We additionally used ONT ultra-long reads that were >100 kb long as the option '-ul'. To assemble the ONT ultra-long reads, NextDenovo v2.5.0[15] was applied with the parameters 'read_cutoff = 10k; seed_depth = 45; nextgraph_options = -a 1 -q 16'. Chromosome-level scaffolding for both the HiFi and ONT assemblies was achieved by integrating the Hi-C data using ALLHiC v0.9.13[16]. Finally, the interaction matrix of each chromosome was checked using HiC-Pro v3.1.0[17]. To further fill the gap in the HiFi assembly, we mapped the ONT contigs to the pseudochromosomes of HiFi using Winnowmap v2.03[18], and then filled the gap using the ONT contigs spanning the gap.

The final assembly quality was comprehensively evaluated. First, we ran Benchmarking Universal Single Copy Orthologs (BUSCO) v5.0.0[19] against the embryophyta_odb10 and the viridiplantae_odb10 databases to evaluate the gene space's completeness. Then the long terminal repeat (LTR) assembly index (LAI)[20] was computed to assess the repeat completeness using LTR_retriever v2.9.0[21]. Finally, the assembly accuracy metrics, including the base pairs' quality value (QV) and k-mer completeness, were determined utilizing Merqury v1.3[22].

Annotation of the transposable elements and protein-coding genes

To annotate the repeat sequences, a *C. monnieri*-specific repeat database was established using RepeatModeler v2.0.1[23]. To locate repetitive elements across the genome, we ran RepeatMasker v4.0.9 (www.repeatmasker.org) against the database with the following parameters: '-e rmblast -div 40 -norna'. We used barrnap v0.9 (https://github.com/tseemann/barrnap) to identify the ribosomal DNA (rDNA) tandem repeats.

For the annotation of protein-coding genes, we used Braker v3.0.8[24], integrating *ab initio* predictions with supporting evidence from both the transcriptome data and homologous proteins. The transcriptome evidence was derived from the RNA-seq data of five tissues. The raw RNA-seq reads underwent quality trimming and adapter removal via Trimmomatic v0.39[25]. The resulting high-quality reads were mapped back to the genome with HISAT2 v2.1.0[26], utilizing the settings '--min-intronlen 20 --max-intronlen 15,000'. The protein evidence included the protein sequences of *Arabidopsis thaliana*[27], *Daucus carota*[28], and the UniProt proteins of Embryophyta (www.uniprot.org). Finally, functional assignments for the predicted coding genes were performed with InterProScan v5.52-86.0[29].

Phylogenetic analysis and estimation of divergence times

To resolve the evolutionary relationships within the Apiaceae family, we used protein sequences from *C. monnieri* and 10 previously reported Apiaceae species, including *Centella asiatica*[30], *D. carota*[28], *Cryptotaenia japonica*[31], *Oenanthe sinensis*[32], *Angelica sinensis*[33], *Apium graveolens*[34], *Heracleum sosnowskyi*[35], *Coriandrum sativum*[36], *Peucedanum praeruptorum*[37], and *Saposhnikovia divaricata*[38] (Supplementary Table S1). The outgroups were *A. thaliana*[27] and *Vitis vinifera*[39]. OrthoFinder v2.5.4[40] was applied to cluster the orthologous gene families. Following this, the protein sequences of identified single-copy genes from each species were extracted. These sequences were aligned using MAFFT v7.487[41] and subsequently trimmed by removing poorly aligned regions with trimAl v1.5.0[42]. The aligned sequences were then used to reconstruct a maximum likelihood phylogenetic tree via IQ-TREE v3.0.1[43], accompanied by an ultrafast bootstrap analysis consisting of 1,000 replicates (-bb 1,000). To infer the temporal divergence of these lineages, r8s v1.81[44] was utilized, incorporating fossil calibration points retrieved from the TimeTree database[45].

Comparative synteny analysis

For evaluating the chromosomal synteny, collinearity among different species was examined via mcscan (JCVI v1.1.19[46]), utilizing orthologous single-copy genes derived from OrthoFinder. The syntenic blocks between species were identified by applying MCScanX[47], setting a threshold of at least 10 matching genes per syntenic block (-s 10 -m 50 -b 2). The resulting synteny relationships were subsequently plotted with the SynVisio (https://synvisio.github.io)[48].

Results

Assembly of the *C. monnieri* genome

The analysis of the karyotype indicated that *C. monnieri* was a diploid species containing 20 chromosomes (2*n* = 2*x* = 20) (Supplementary Fig. S1). According to the k-mer analysis, the approximate size of the *C. monnieri* genome was 1.17 Gb (Supplementary Fig. S2). To construct a high-quality genome assembly for *C. monnieri*, we performed PacBio HiFi long reads and ONT ultra-long reads sequencing. We generated ~99.7 Gb PacBio HiFi long reads (read N50 = 18.3 kb) and ~111.2 Gb ONT ultra-long reads (read N50 = 100.9 kb), which covered ~85× and ~95× of the *C. monnieri* genome, respectively (Supplementary Tables S2 and S3). Initial assembly of the PacBio HiFi data was conducted using Hifiasm[14], yielding 512 contigs with an N50 of 100.9 Mb (Supplementary Table S2). We also assembled the ONT ultra-long reads using NextDenovo[15], resulting in 670 contigs with an N50 of 5.8 Mb (Supplementary Table S2). Since the contigs assembled from the HiFi reads exhibited a notably higher N50 than those assembled from the ONT reads, we adopted the HiFi contigs for chromosome scaffolding (Supplementary Fig. S3). These contigs were anchored, oriented, and manually curated into 10 pseudochromosomes using Hi-C data with ALL-HiC[16]. The resulting assembly contained only one gap, and the anchoring rate of the HiFi contigs was 95.9% (Supplementary Table S2), indicating the high continuity of the HiFi assembly. We then filled the gap by using the ONT contigs spanning the gap (Supplementary Fig. S4). Finally, we performed telomere patching using TeloClip v0.3.2 (https://github.com/Adamtaranto/teloclip) along with raw long reads, yielding all 20 telomeres that were enriched with seven-base telomere repeats (5'-TTTAGGG-3') (Supplementary Fig. S5; Supplementary Table S4). The total size of the pseudochromosomes in the final assembly was 1.13 Gb (Fig. 1a, b and Supplementary Table S2), which is consistent with the estimated genome size or that of Cmo_YZ (1.14 Gb)[11]. Collinearity analysis with the Cmo_YZ assembly revealed that the current assembly showed high collinearity with it, though there are several chromosomal inversions (Fig. 1c).

To characterize the centromeres, we performed a comprehensive identification of tandem repeats across the genome using TRF[49] and identified two predominant repeat units (118 and 180 bp of monomer lengths, respectively; Supplementary Fig. S6). The 118-bp unit (*Cmon118*) was the potentially primary centromeric satellite repeat. Although this repeat dominated the centromeric regions of seven chromosomes, it was absent on chromosomes 5, 9, and 10 (Supplementary Fig. S7). This may be because the centromeres of these three chromosomes are maintained through the massive accumulation of LTR/*Gypsy* retrotransposons[50]. We also discovered another highly abundant 180-bp repetitive sequence (*Cmon180*). Compared with *Cmon118*, this unit was highly conserved and was predominantly distributed in the terminal regions of the chromosomes, indicating that it was likely a subtelomeric satellite repeat[51].

To assess the quality of the assembly, we first evaluated its completeness using BUSCO analysis with embryophyta_odb10 and

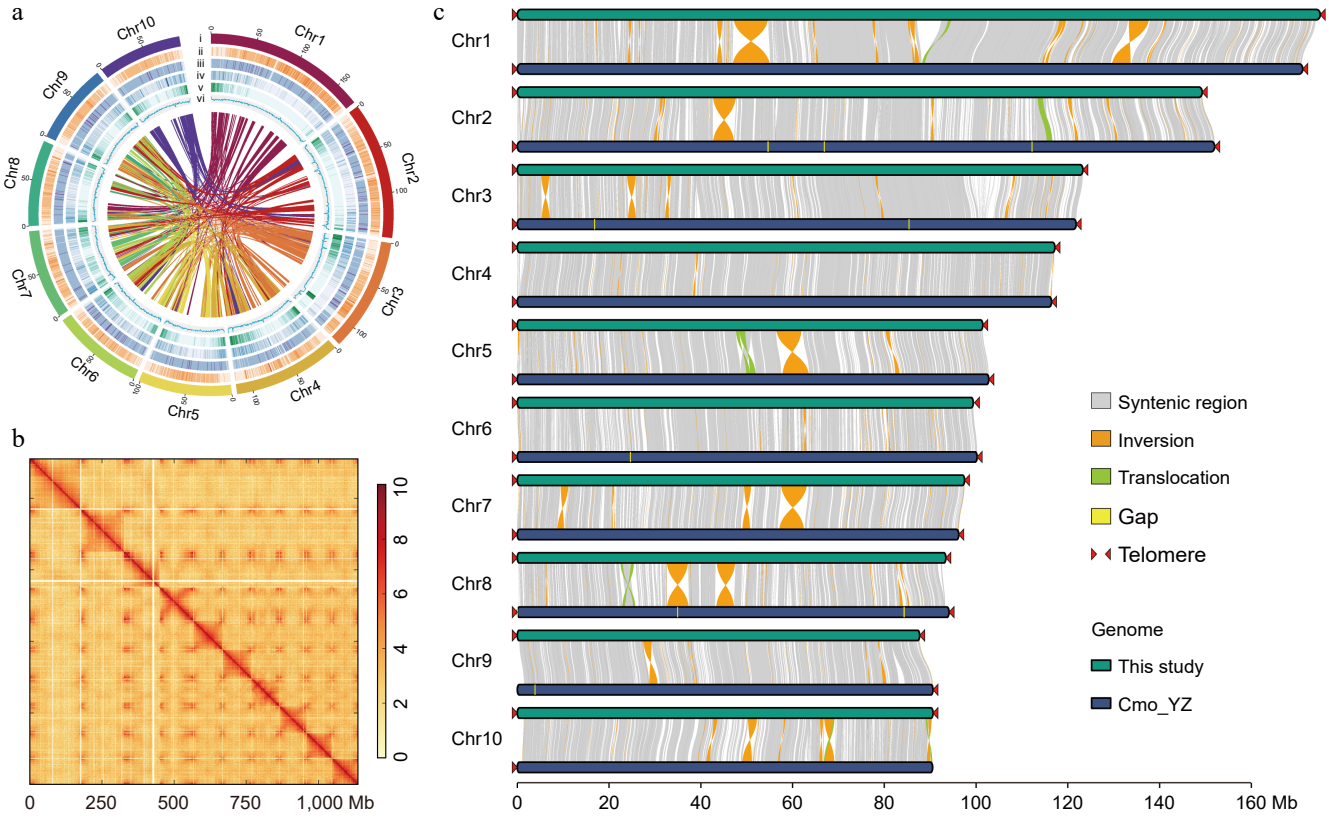

Fig. 1 Assembly of the *C. monnieri* genome. (a) The circos plot of features across the chromosomes of *C. monnieri*. Tracks i–vi represent the chromosomes, LTR/*Gypsy*, LTR/*Copia*, DNA-TEs, gene, and GC content, respectively. (b) Heatmap displaying whole-genome Hi-C interaction frequencies for *C. monnieri*. The intensity of the interactions is represented by colors shading from yellow (low) to red (high). (c) Collinearity mapping between this study (top) and Cmo_YZ[11] (bottom). Gray ribbons connect collinear regions between the two genomes. Gap regions are represented by yellow blocks in Cmo_YZ assembly. The triangles at the ends of chromosomes indicate the identified telomeric repeat sequences.

https://doi.org/10.48130/gcomm-0026-0014

viridiplantae_odb10, of which 98.2% (1,584/1,614) and 98.8% (420/425) complete genes were present in the assembly, respectively (Supplementary Table S5). Additionally, we estimated the contiguity and repeat region assembly quality using the LAI[20], resulting in a value of 20.94 (Supplementary Table S2), indicating that the current assembly had high completeness. We also mapped the PacBio HiFi reads, Illumina reads, and RNA-seq reads to the assembly, yielding mean mapping rates of 99.94%, 99.78%, and 94.27%, respectively. Further evaluation using Merqury[22] showed that the average QV score of the chromosomes was 55.86 (ranging from 54.94 to 57.29 for each chromosome), indicating high (> 99.999%) assembly accuracy (Supplementary Table S6). Additionally, the Hi-C interaction matrices also showed that the assembly had high consistency (Fig. 1b and Supplementary Fig. S7). Compared with the previously published *C. monnieri* genome (Cmo_YZ[11]), our assembly achieved a T2T status. Consequently, our assembly exhibits a higher contig N50 (100.9 Mb vs. 78.45 Mb) and chromosome anchoring rate (95.9% vs. 93.9%). Overall, our gapless assembly represents a contiguous and complete genome of *C. monnieri*.

Annotation of TEs and protein-coding genes in the *C. monnieri* genome

In total, 884.86 Mb of repeat elements was annotated, constituting 74.76% of the *C. monnieri* genome (Supplementary Table S7). Among the annotated transposable elements (TEs), LTR retrotransposons were the most dominant TE type, covering 51.45% of the genome sequence (Fig. 2a, Supplementary Table S7). In contrast, the content of other TEs (e.g., DNA transposons, long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), and rolling-circle (RC)/*Helitron*) was low (< 3%) (Fig. 2a and Supplementary Table S7). Analysis of the distribution of sequence divergence across the TE superfamilies revealed that younger TEs (with lower divergence), particularly for LTR retrotransposons, had higher proportions, indicating the LTR retrotransposons had recently been active in the genome (Fig. 2a). Notably, we found a strong positive correlation between genome size and TE content in Apiaceae species ($R = 0.98$, $p = 3.35 \times 10^{-6}$; Supplementary Fig. S8).

Subsequently, we predicted the protein-coding genes from the repeat-masked genome using Braker[24], which integrates *ab initio*

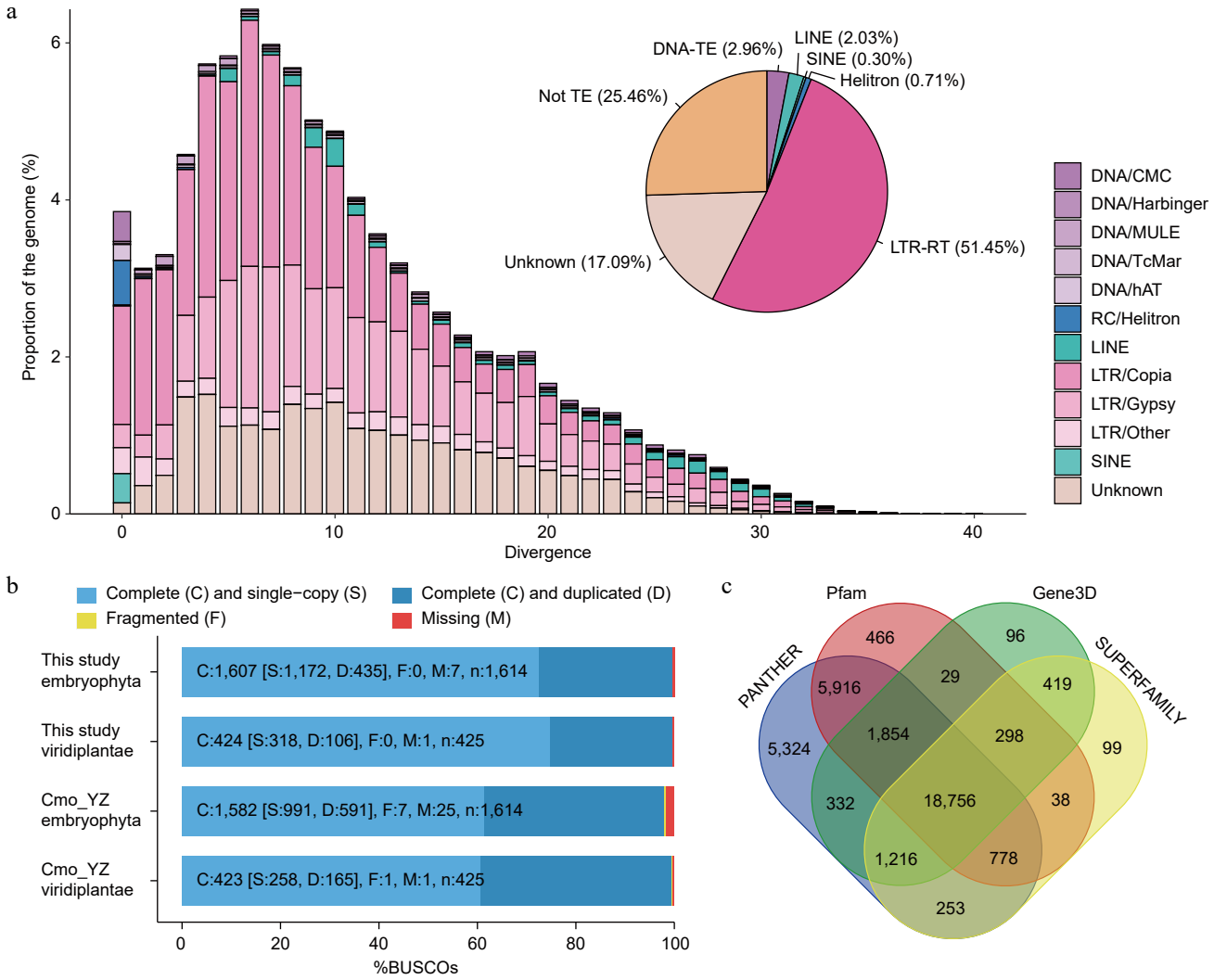

Fig. 2 Characterization of TEs and protein-coding genes in the *C. monnieri* genome. (a) A repeat landscape of the *C. monnieri* genome. The *x*-axis (divergence) represents the Kimura 2-parameter distance between individual repeat elements and their corresponding consensus sequences. The inset pie chart illustrates the percentage of TEs within the whole genome. (b) BUSCO completeness for the different datasets. (c) Venn diagram of the annotation results from different databases.

gene predictions, transcriptome data, and homologous protein evidence. In total, 39,218 protein-coding genes were annotated in the *C. monnieri* genome, of which 99.06% were located on the chromosomes. We then evaluated the completeness of gene annotation using BUSCO[19], resulting in 99.6% (1,607/1,614 in embryophyta_odb10) and 99.7% (424/425 in viridiplantae_odb10) of all BUSCO genes being contained in the annotated gene set (Fig. 2b, Supplementary Table S8). Of all the annotated genes, 97.48% (38,231/39,218) had functional annotations from the InterPro database[29], and 87.79%, 71.74%, 58.65%, and 56.06% contained functional terms in the databases of PANTHER, Pfam, Gene3D, and SUPERFAMILY, respectively (Fig. 2c, Supplementary Table S9).

Genome evolution of Apiaceae

To resolve the evolutionary history of Apiaceae and elucidate the chromosomal diversification patterns within the clade, we reconstructed a phylogenetic tree using 321 single-copy orthologous genes across the Apiaceae species with *A. thaliana* and *V. vinifera* as the outgroups (Fig. 3a). The resulting maximum likelihood tree was used to infer the divergence times via fossil-calibrated molecular clock methods. Consistent with previous studies, species belonging to the subfamily Apioideae formed a well-supported monophyletic group, and *C. monnieri* was resolved as the sister to *P. praeruptorum* and *S. divaricata*[37]. *C. monnieri*, along with *P. praeruptorum* and *S. divaricata*, diverged from *C. sativum* 11 million years ago (Mya) (Fig. 3a).

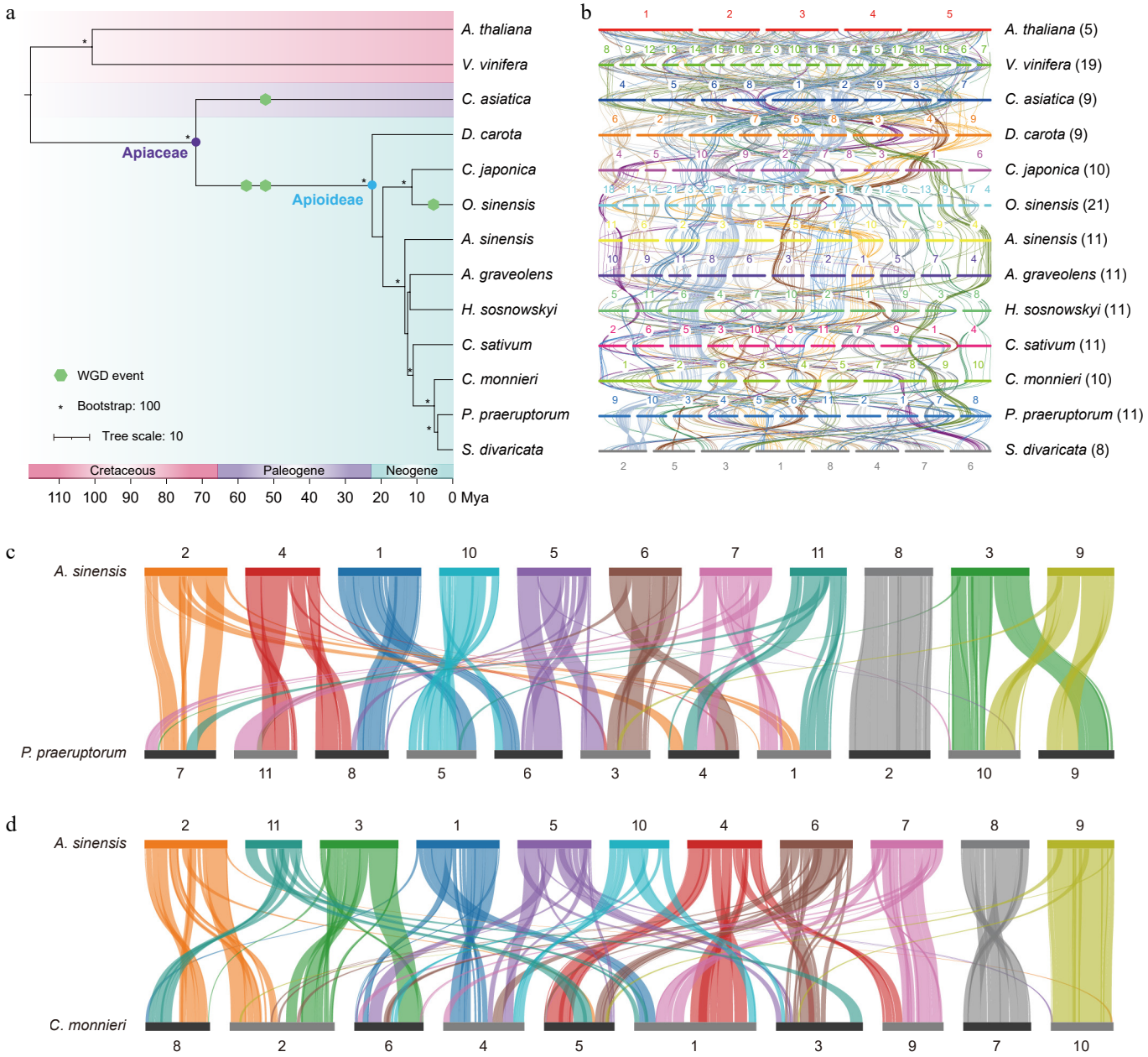

Fig. 3 Genome evolution of *C. monnieri* and Apiaceae. (a) Phylogenetic analysis of *C. monnieri* and other Apiaceae lineages. (b) Syntenic analysis of the Apiaceae species. (c) Synteny blocks between *A. sinensis* and *P. praeruptorum*. (d) Synteny blocks between *A. sinensis* and *C. monnieri*. For (c) and (d), the colored ribbons represent syntenic blocks identified by MCScanX and visualized using SynVisio. Numbers indicate chromosome identification numbers (IDs).

https://doi.org/10.48130/gcomm-0026-0014

To investigate the patterns of chromosomal evolution, we conducted a genome-wide synteny analysis based on single-copy ortholog anchors across Apiaceae species, revealing the uneven conservation of structure of chromosomes in Apiaceae (Fig. 3b). Species with 11-chromosome karyotypes in Apioideae, including *A. sinensis*, *H. sosnowskyi*, *A. graveolens*, *C. sativum*, and *P. praeruptorum*, exhibited highly conserved syntenic blocks (Fig. 3b). The detailed synteny analyses between *A. sinensis* (as the reference) and other 11-chromosome species further resolved fine-scale chromosomal dynamics (Supplementary Fig. S9). *A. sinensis* consistently displayed a low degree of chromosomal structural rearrangement across the comparisons, indicating it preserved one of ancestral and conserved karyotypes within the Apioideae. For example, the genome of *A. graveolens* showed only limited rearrangement relative to *A. sinensis*, with chromosomal rearrangements detected only for chromosomes 9, 10, and 11 (Supplementary Fig. S9a). As the phylogenetic distance increased, the number of structural rearrangements generally rose; however, even in more distantly related 11-chromosome species, certain chromosomes remained highly conserved, such as chromosome 2 of *P. praeruptorum* (Fig. 3c), representing conserved linkage blocks that had persisted throughout diversification of the Apioideae. This is consistent with prior studies indicating that the ancestral chromosome number of Apioideae was 11[34,52], suggesting that these species retain the plesiomorphic chromosomal architecture of the subfamily. In contrast, species with deviant chromosome counts exhibited markedly reduced collinearity and increased interchromosomal rearrangements (Supplementary Fig. S10). For instance, *C. monnieri* (n = 10) underwent extensive chromosomal rearrangements relative to 11-chromosome species like *A. sinensis*, and all chromosomes except 7 and 10 experienced extensive chromosomal rearrangements, leading to its reduced chromosome count (Fig. 3d).

This pattern was further reflected in the comparative quantification of interchromosomal rearrangements by measuring how many chromosomal segments from other species contributed to a target chromosome. Species with 11 chromosomes (e.g., *A. sinensis* and *A. graveolens*) had fewer such segments, indicating lower rearrangement frequencies (Supplementary Fig. S11). A notable feature was the genome of *O. sinensis* (n = 21), which underwent an additional whole-genome duplication (WGD)[32]. A detailed chromosome-to-chromosome synteny comparison showed that most chromosomes of *O. sinensis* matched two counterparts in *A. sinensis*, providing strong evidence for a recent WGD event. Nevertheless, one set of its chromosomes still maintained the 11-chromosome structure, which explains why it had the fewest interchromosomal rearrangements (Supplementary Fig. S10d).

These results demonstrated that WGD, conservation of the ancestral 11-chromosome architecture, and lineage-specific rearrangements collectively drove the genomic evolution of Apiaceae. Species with 11 chromosomes in the Apioideae represent a conserved lineage, while *C. monnieri* and other taxa with deviant chromosome numbers have undergone extensive rearrangements. The observed genomic dynamics shed light on the adaptive evolution and taxonomic diversification of the Apiaceae.

Discussion

In this study, we generated a gapless T2T genome assembly of *C. monnieri*, achieving a significant improvement in continuity and accuracy over previously reported assemblies[11]. The high QV score, nearly complete BUSCO coverage, and the identification of 20 telomeres highlighted the high quality and structural completeness of the genome. This T2T genome provides a robust foundation for downstream research on medicinal compound biosynthesis, functional genomics, and evolutionary biology. However, limitations inherent to the haplotype-collapsed assembly strategy remain. For example, chromosome 7 displayed the signatures of potential chimeric assembly, a phenomenon most likely caused by the extreme sequence divergence between the two haplotypes of this chromosome. Generating a completely phased diploid T2T genome to definitively resolve these complex allelic variations will be the primary focus of future research.

The repeat composition of the *C. monnieri* genome revealed a substantial accumulation of LTR retrotransposons, which is consistent with patterns reported in several Apiaceae species[36–38]. The relatively high proportion of LTRs, particularly *Gypsy* and *Copia* elements, suggests that TE expansion has played a major role in shaping variations in genome size within the family, supporting the hypothesis that differences in LTR proliferation and deletion rates may underlie the pronounced genome size disparities observed among closely related genera[53].

Comparative genomic analyses further clarified the evolutionary placement of *C. monnieri* within the Apiaceae. Using 321 conserved single-copy orthologs, we reconstructed a robust phylogeny that positioned *C. monnieri* as sister to *P. praeruptorum* and *S. divaricata*. The estimated divergence times among these species provide new insights into lineage diversification patterns within the tribe and highlight the value of high-quality genomic datasets for reconstructing Apiaceae's evolution. Our comparative genomic analyses illuminated substantial differences in chromosomal evolution across the Apiaceae. Species retaining the ancestral 11-chromosome karyotype, such as *A. sinensis* and *A. graveolens*[33,34], showed high syntenic conservation, whereas *C. monnieri* (n = 10) displayed extensive chromosomal rearrangements involving both fissions and fusions. These lineage-specific structural changes may underlie unique metabolic traits, ecological functions, and adaptation strategies.

Overall, our study offers a significantly improved genomic resource for *C. monnieri*, facilitating future multidisciplinary research. First, from a pharmacological perspective, *C. monnieri* is rich in active secondary metabolites, particularly coumarins and volatile oils. High-continuity genomic data are crucial for the precise identification of biosynthetic gene clusters and provide a reliable foundation for elucidating the complete biosynthetic pathways of these medicinal compounds. Second, from an ecological standpoint, *C. monnieri* mediates triple-trophic plant–aphid–predator interactions by releasing specific HIPVs such as β-myrcene. The complete genome sequence will greatly facilitate the discovery of the genetic mechanisms underlying these ecological traits and volatile emissions. Finally, integrating this T2T assembly with future pan-genomic analyses and transcriptomic profiling will serve as a robust resource for population genomics and marker-assisted breeding. This will ultimately accelerate the cultivation of superior *C. monnieri* varieties with enhanced medicinal yields and ecological pest management capabilities, further advancing both the biological and applied value of this species.

Author contributions

The authors confirm their contributions to the paper as follows: conceived and designed the study: Ge F, Chen J; prepared the materials: Zhao L, Li Z, Zhang X, Liang X; performed genome assembly and annotation: Liu Y, Zhao L; performed the evolutionary analyses: Liu

Y, Zhao L, Zhang J, Ju Q, Fan X; wrote and revised the paper: Liu Y, Ge F, Chen J. All authors reviewed the results and approved the final version of the manuscript.

Data availability

We deposited the raw sequence data in the National Genomics Data Center (NGDC), Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn), under the BioProject accession number PRJCA052818. The raw sequences of PacBio HiFi reads (CRR2409651), Nanopore ultra-long reads (CRR2409652), whole-genome sequence short reads (CRR2409657 and CRR2409658), RNA-seq reads (CRR2409659–CRR2409663), and Hi-C reads (CRR2409653–CRR2409656) have been deposited in the Genome Sequence Archive (GSA) of the NGDC under the accession number CRA034706. The genome assembly has been deposited in the National Center for Biotechnology Information (NCBI) Genome (PRJNA1392685). The assembled genome sequences, gene and TE annotations are available on Zenodo (doi:10.5281/zenodo.20540880). All study data are included in the main article and Supplementary Materials.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant number 2023YFD1400800) and the Initiative Scientific Research Program, Institute of Zoology, Chinese Academy of Sciences (2024IOZ0106, 2023IOZ0203, 2025IOZ09, and SKLA2506). The authors thank Jinshuai Zhao and Liqiang Xie from the Institute of Zoology, Chinese Academy of Sciences, for their support in the data analyses.

Conflicts of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper online at: https://doi.org/10.48130/gcomm-0026-0014.

Dates

Received 9 May 2026; Revised 4 June 2026; Accepted 11 June 2026; Published online 20 July 2026

References

[1] Li YM, Jia M, Li HQ, Zhang ND, Wen X, et al. 2015. *Cnidium monnieri*: a review of traditional uses, phytochemical and ethnopharmacological properties. *The American Journal of Chinese Medicine* 43:835–877

[2] An J, Yang H, Zhang Q, Liu C, Zhao J, et al. 2016. Natural products for treatment of osteoporosis: the effects and mechanisms on promoting osteoblast-mediated bone formation. *Life Sciences* 147:46–58

[3] Lee TH, Chen YC, Hwang TL, Shu CW, Sung PJ, et al. 2014. New coumarins and anti-inflammatory constituents from the fruits of *Cnidium monnieri*. *International Journal of Molecular Sciences* 15:9566–9578

[4] Li QN, Liang NC, Wu T, Wu Y, Xie H, et al. 1994. Effects of total coumarins of Fructus cnidii on skeleton of ovariectomized rats. *Acta Pharmacologica Sinica* 15:528–532

[5] Shin E, Choi KM, Yoo HS, Lee CK, Hwang BY, et al. 2010. Inhibitory effects of coumarins from the stem barks of *Fraxinus rhynchophylla* on adipocyte differentiation in 3T3-L1 cells. *Biological and Pharmaceutical Bulletin* 33:1610–1614

[6] Sun Y, Yang AWH, Lenon GB. 2020. Phytochemistry, ethnopharmacology, pharmacokinetics and toxicology of *Cnidium monnieri* (L) Cusson. *International Journal of Molecular Sciences* 21:1006

[7] Tamura S, Fujitani T, Kaneko M, Murakami N. 2010. Prenylcoumarin with Rev-export inhibitory activity from Cnidii Monnieris Fructus. *Bioorganic & Medicinal Chemistry Letters* 20:3717–3720

[8] Su W, Ouyang F, Li Z, Yuan Y, Yang Q, et al. 2023. *Cnidium monnieri* (L.) Cusson flower as a supplementary food promoting the development and reproduction of ladybeetles *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae). *Plants* 12:1786

[9] Yang Q, Men X, Zhao W, Li C, Zhang Q, et al. 2021. Flower strips as a bridge habitat facilitate the movement of predatory beetles from wheat to maize crops. *Pest Management Science* 77:1839–1850

[10] Jiang X, Zhao L, Sergers A, Chang C, Zhang X, et al. 2025. Herbivore-induced plant volatiles (HIPVs) from companion plant could enhance predator recruitment and biocontrol of cereal aphids. *Entomologia Generalis* 45:1067–1077

[11] Wang Z, He J, Qi Q, Wang K, Tang H, et al. 2024. Chromosome-level genome assembly of *Cnidium monnieri*, a highly demanded traditional Chinese medicine. *Scientific Data* 11:667

[12] Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating *k*-mer statistics. *Bioinformatics* 33:2759–2761

[13] Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33:2202–2204

[14] Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18:170–175

[15] Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, et al. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biology* 25:107

[16] Zhang X, Zhang S, Zhao Q, Ming R, Tang H. 2019. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants* 5:833–845

[17] Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, et al. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology* 16:259

[18] Jain C, Rhie A, Hansen NF, Koren S, Phillippy AM. 2022. Long-read mapping to repetitive reference sequences using Winnowmap2. *Nature Methods* 19:705–710

[19] Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* 38:4647–4654

[20] Ou S, Chen J, Jiang N. 2018. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research* 46:e126

[21] Ou S, Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology* 176:1410–1422

[22] Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* 21:245

[23] Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* 117:9451–9457

[24] Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, et al. 2024. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research* 34:769–777

[25] Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120

[26] Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37:907–915

[27] Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, et al. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40:D1202–D1210

[28] Coe K, Bostan H, Rolling W, Turner-Hissong S, Macko-Podgórni A, et al. 2023. Population genomics identifies genetic signatures of carrot

domestication and improvement and uncovers the origin of high-carotenoid orange carrots. *Nature Plants* 9:1643−1658
[29] Jones P, Binns D, Chang HY, Fraser M, Li W, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236−1240
[30] Pootakham W, Naktang C, Kongkachana W, Sonthirod C, Yoocha T, et al. 2021. *De novo* chromosome-level assembly of the *Centella asiatica* genome. *Genomics* 113:2221−2228
[31] Liu H, Zhang JQ, Zhang RR, Zhao QZ, Su LY, et al. 2024. The high-quality genome of *Cryptotaenia japonica* and comparative genomics analysis reveals anthocyanin biosynthesis in Apiaceae. *The Plant Journal* 118:717−730
[32] Liu JX, Liu H, Tao JP, Tan GF, Dai Y, et al. 2023. High-quality genome sequence reveals a young polyploidization and provides insights into cellulose and lignin biosynthesis in water dropwort (*Oenanthe sinensis*). *Industrial Crops and Products* 193:116203
[33] Han X, Li C, Sun S, Ji J, Nie B, et al. 2022. The chromosome-level genome of female ginseng (*Angelica sinensis*) provides insights into molecular mechanisms and evolution of coumarin biosynthesis. *The Plant Journal* 112:1224−1237
[34] Song X, Sun P, Yuan J, Gong K, Li N, et al. 2021. The celery genome sequence reveals sequential paleo-polyploidizations, karyotype evolution and resistance gene reduction in Apiales. *Plant Biotechnology Journal* 19:731−744
[35] Schelkunov MI, Shtratnikova VY, Klepikova AV, Makarenko MS, Omelchenko DO, et al. 2024. The genome of the toxic invasive species *Heracleum sosnowskyi* carries an increased number of genes despite absence of recent whole-genome duplications. *The Plant Journal* 117:449−463
[36] Song X, Wang J, Li N, Yu J, Meng F, et al. 2020. Deciphering the high-quality genome sequence of coriander that causes controversial feelings. *Plant Biotechnology Journal* 18:1444−1456
[37] Huang XC, Tang H, Wei X, He Y, Hu S, et al. 2024. The gradual establishment of complex coumarin biosynthetic pathway in Apiaceae. *Nature Communications* 15:6864
[38] Wang ZH, Liu X, Cui Y, Wang YH, Lv ZL, et al. 2024. Genomic, transcriptomic, and metabolomic analyses provide insights into the evolution and development of a medicinal plant *Saposhnikovia divaricata* (Apiaceae). *Horticulture Research* 11:uhae105
[39] Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463−467
[40] Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* 16:157
[41] Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772−780
[42] Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972−1973
[43] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268−274
[44] Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301−302
[45] Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, et al. 2022. TimeTree 5: an expanded resource for species divergence times. *Molecular Biology and Evolution* 39:msac174
[46] Tang H, Krishnakumar V, Zeng X, Xu Z, Taranto A, et al. 2024. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta* 3:e211
[47] Wang Y, Tang H, DeBarry JD, Tan X, Li J, et al. 2012. *MCScanX*: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40:e49
[48] Bandi V GC. 2020. Interactive exploration of genomic conservation. *Proceedings of Graphics Interface 2020, Toronto, Canada, 28–29 May, 2020*. Canada: Canadian Human-Computer Communications Society. pp. 74–83 doi: 10.20380/GI2020.09
[49] Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27:573−580
[50] Xie L, Huang Y, Huang W, Shang L, Sun Y, et al. 2025. Genetic diversity and evolution of rice centromeres. *Nature Genetics* 57:2808−2818
[51] Horáková L, Bačovský V, Čegan R, Janoušek B, Patzak J, et al. 2025. Contrasting pattern of subtelomeric satellites in the Cannabaceae family. *Frontiers in Plant Science* 16:1631369
[52] Walker JW, Doyle JA. 1975. The bases of angiosperm phylogeny: palynology. *Annals of the Missouri Botanical Garden* 62:664−723
[53] Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Research* 110:91−107