Check for updates

Chromosome-level genome assembly of *Ajuga decumbens*

Yubang Gao[1,2†], Jingzhao Li[1,2†], Yuli Xie[1], Teng Zhang[1], Kai Tian[1,3], Xiaotang Li[1] and Lunguang Yao[1,2,3*]

[1]School of Life Sciences and Agricultural Engineering, Nanyang Normal University, Nanyang, Henan, China, [2]Henan Province Artemisia Argyi Development and Utilization Engineering Technology Research Center, Nanyang Normal University, Nanyang, Henan, China, [3]Henan Field Observation and Research Station of Headwork Wetland Ecosystem of the Central Route of South-to-North Water Diversion Project, School of Life Sciences and Agricultural Engineering, Nanyang Normal University, Nanyang, Henan, China

Introduction

Ajuga decumbens, a critical species in the Ajuga genus, is a flowering plant widely used in traditional Chinese medicine (Figure 1A). The Ajuga genus encompasses more than 300 plant species. It is chiefly found in temperate regions across Europe, Asia, Australia, North America, and Africa (Atay et al., 2016; Park et al., 2017). Prominent species in this genus are recognized for their medicinal properties (Qing et al., 2017). Phytochemical studies have identified a range of bioactive compounds in Ajuga, such as phytosterols, diterpenoids, triterpenoids, sesquiterpenoids, and iridoids (Qing et al., 2017; Dong et al., 2020). Pharmacological studies have shown that Ajuga has anticancer (Graham et al., 2000), antipyretic (Shafi et al., 2004; Debell et al., 2005), anti-inflammatory (Korkina et al., 2007; Dong et al., 2020; Liu et al., 2020), antioxidant (Chenni et al., 2007; Bouderbala et al., 2008), anti-malarial (Kuria et al., 2002), antimicrobial (Chen et al., 1999; Kariba, 2001), anti-arthritic (Ono et al., 2008), antitumor (Wessner et al., 1992; Cárdenas et al., 1994), anti-tussive (College, 1977), hypoglycemic (El Hilaly and Lyoussi, 2002; Chenni et al., 2007), and insecticidal properties (Jbilou et al., 2006, 2008).

Genomics plays a significant role in uncovering medicinal plants' biological traits, chemical synthesis mechanisms, gene-assisted breeding, and synthetic biology. The development of a high-quality genome for the medicinal plant *Artemisia annua* aids in elucidating the biosynthesis of artemisinin and in studying genes highly related to its content, thereby facilitating the breeding of *Artemisia annua* and the development of drugs to combat antimalarial resistance (Shen et al., 2018; Liao et al., 2022). By analyzing the *Panax schinseng* genome, further insights into the diversity of ginsenosides have been elucidated (Kim et al., 2018). The assembly of the *Salvia miltiorrhiza* genome revealed that tanshinones, which accumulate primarily in the roots and most abundantly in the hairy roots. Studies on the biosynthesis of phenolic acids have advanced the breeding of new varieties and biochemical research (Xu et al., 2016; Song et al., 2020; Yang et al., 2022). Therefore, obtaining high-quality genomes of medicinal plants and conducting gene
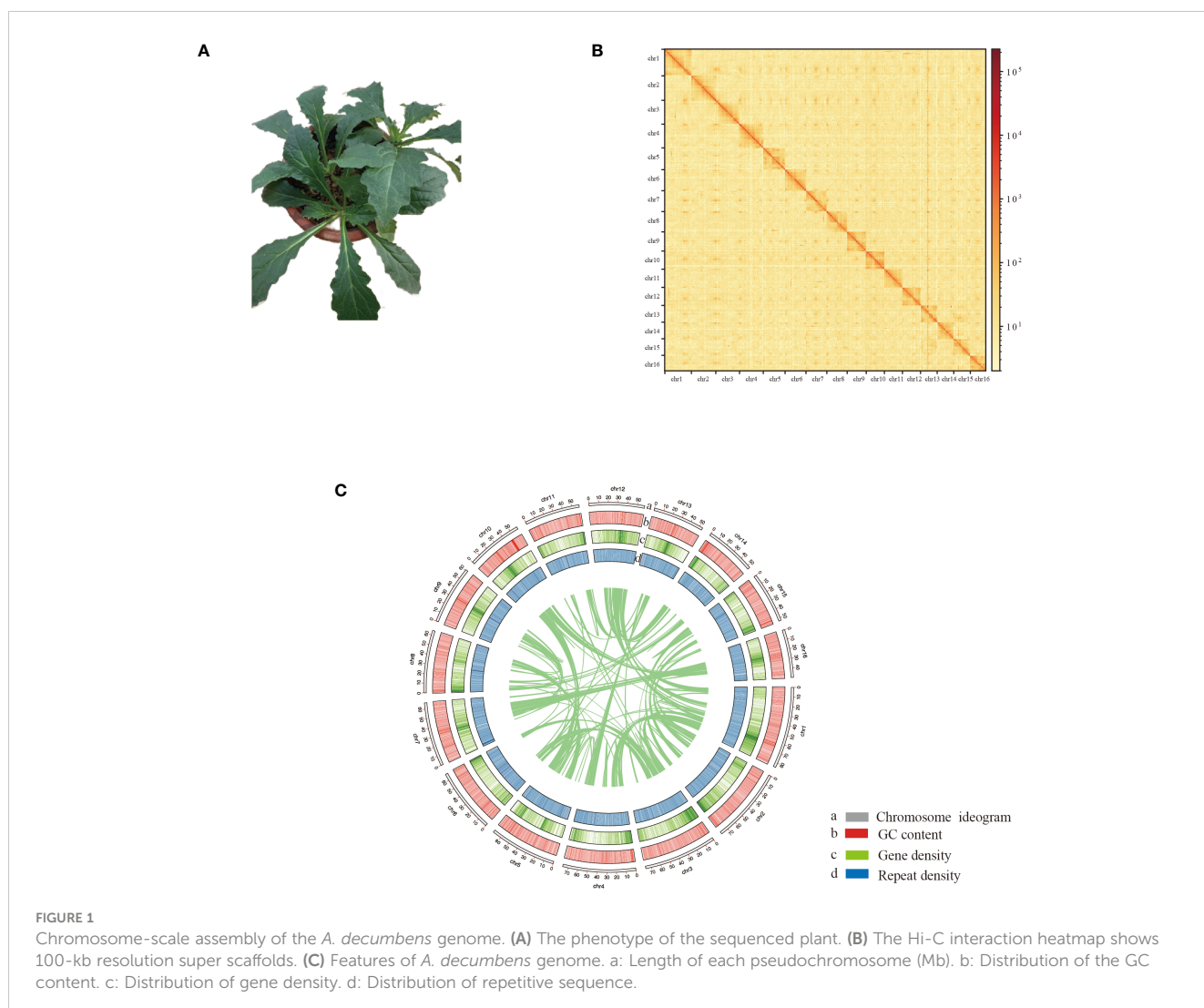
mining and synthetic pathway analysis for active ingredients, coupled with strategies such as functional genomics, are crucial for the rapid advancement of medicinal plant biology.

Only the chromosome-level genome data of *Ajuga chamaepitys* is available from NCBI (GCA_958295605.1), but it lacks corresponding gene information. The widely used *A. decumbens* still lacks equivalent genome data. The high-quality genome of *A. decumbens* plays a crucial role in unraveling the genetic underpinnings of significant traits, driving genetic advancements, and aiding in synthesizing pharmacological compounds. PacBio HiFi sequencing can produce high-quality contigs, while Hi-C sequencing is instrumental in sorting and orienting these contigs (Dudchenko et al., 2017; Cheng et al., 2021). Hi-C sequencing has been successfully applied to generate high-quality genome sequences in complex organisms (Chang et al., 2023; Nakandala et al., 2023). In this study, we performed PacBio HiFi sequencing and Hi-C sequencing of *A. decumbens*, examining its high-quality genomic information at the chromosome level. This genome resource will significantly advance the exploration of medicinal resources and the cultivation of new varieties in the Ajuga genus.

Results

Genome sequencing and genome assembly

Previous studies estimated a genome size of 1.1 Gb (Choi et al., 2019). PacBio HiFi long reads and Hi-C sequencing data were generated (Supplementary Table S1). Our sequencing produced approximately 17.8 Gb of HiFi reads, resulting in a 1.1 Gb genome assembly. The assembly contains 1666 contigs with an N50 contig size of 2.09 Mb. Using the Hi-C data, we effectively organized the contigs into 16 clusters, corresponding to the 16 chromosomes of *A. decumbens*. Pseudochromosome lengths varied from 49.8 to 84.6 Mb. The Hi-C data contact matrix heatmap (Figure 1B) effectively demonstrated the contigs' clustering, arrangement, and orientation. The final chromosome-level assembly of the *A. decumbens* genome was 970 Mb (Supplementary Table S2). 99.89% of HiFi reads were aligned to the genome. The completeness of the genome assembly was further validated by BUSCO (Benchmarking Universal Single-Copy Orthologs), indicating 99.44% coverage of essential conserved

FIGURE 1
Chromosome-scale assembly of the *A. decumbens* genome. **(A)** The phenotype of the sequenced plant. **(B)** The Hi-C interaction heatmap shows 100-kb resolution super scaffolds. **(C)** Features of *A. decumbens* genome. a: Length of each pseudochromosome (Mb). b: Distribution of the GC content. c: Distribution of gene density. d: Distribution of repetitive sequence.

plant proteins. The k-mer QV reached 56.02%. The k-mer error rate was 2.55e-06. These results indicated that the genome was highly accurate. The genome's raw LAI was 18.19, and the adjusted LAI was 12.56, meeting the reference genome standards. These results suggest that the genome assembly is of good quality.

Genome annotation

Further annotation of the assembled *A. decumbens* genome revealed a high repeat content of about 76.1%, with retrotransposons (Class I elements) making up 56.92% (Supplementary Table S3). Long terminal repeat (LTR) sequences, including 31.54% Gypsy-type and 12.19% Copia-type, comprised 46.73% of the genome. It is similar to many other plant genomes where LTR-retrotransposons are predominant.

Additionally, we annotated protein-coding genes by integrating homology-based searches and RNA-Seq data. This hybrid approach identified 32,452 gene models in the *A. decumbens* genome. These models have an average coding sequence (CDS) length of 1,184 bp and an average of 4.59 exons per gene (Supplementary Table S4). Among these genes, 25,347 (78.11%) showed homology with Arabidopsis proteins, and 31,927 (98.38%) were functionally annotated using multiple public databases. Figure 1C demonstrates the typical complementary relationship between the density of genes and the repetitive elements in the genome. The analysis subsequently predicted non-coding RNAs in the genome, including 20,493 rRNAs, 5,029 tRNAs, 106 miRNAs, and 9,568 snRNAs.

Phylogenetic analysis

To examine the evolutionary history of *A. decumbens*, we assessed 11 angiosperm species, including a basal angiosperm (*A. trichopoda*), a monocot (*O. sativa*), and nine dicots (Supplementary Table S5). Analysis of these species yielded 307,695 proteins, from which 28,699 gene families were identified. Of these, 3,039 families were common across the species, indicative of a shared angiosperm ancestor. Among these gene families, 182 single-copy orthologous genes were found (Supplementary Table S6), aiding in constructing a phylogenetic tree. This tree showed *A. decumbens* and *L. japonicus* as closely related, having diverged 38.26 million years ago (Figure 2A). We pinpointed 897 gene families unique to *A. decumbens*, comprising 3,375 genes (Figure 2B). Of these, 2,980 genes (88.30%) were annotated. Notably, *A. decumbens* showed gene family expansion (161 families, 1,299 genes) and contraction (184 families, 232 genes), with the expanded genes being predominantly associated with salt stress resistance (31 genes).
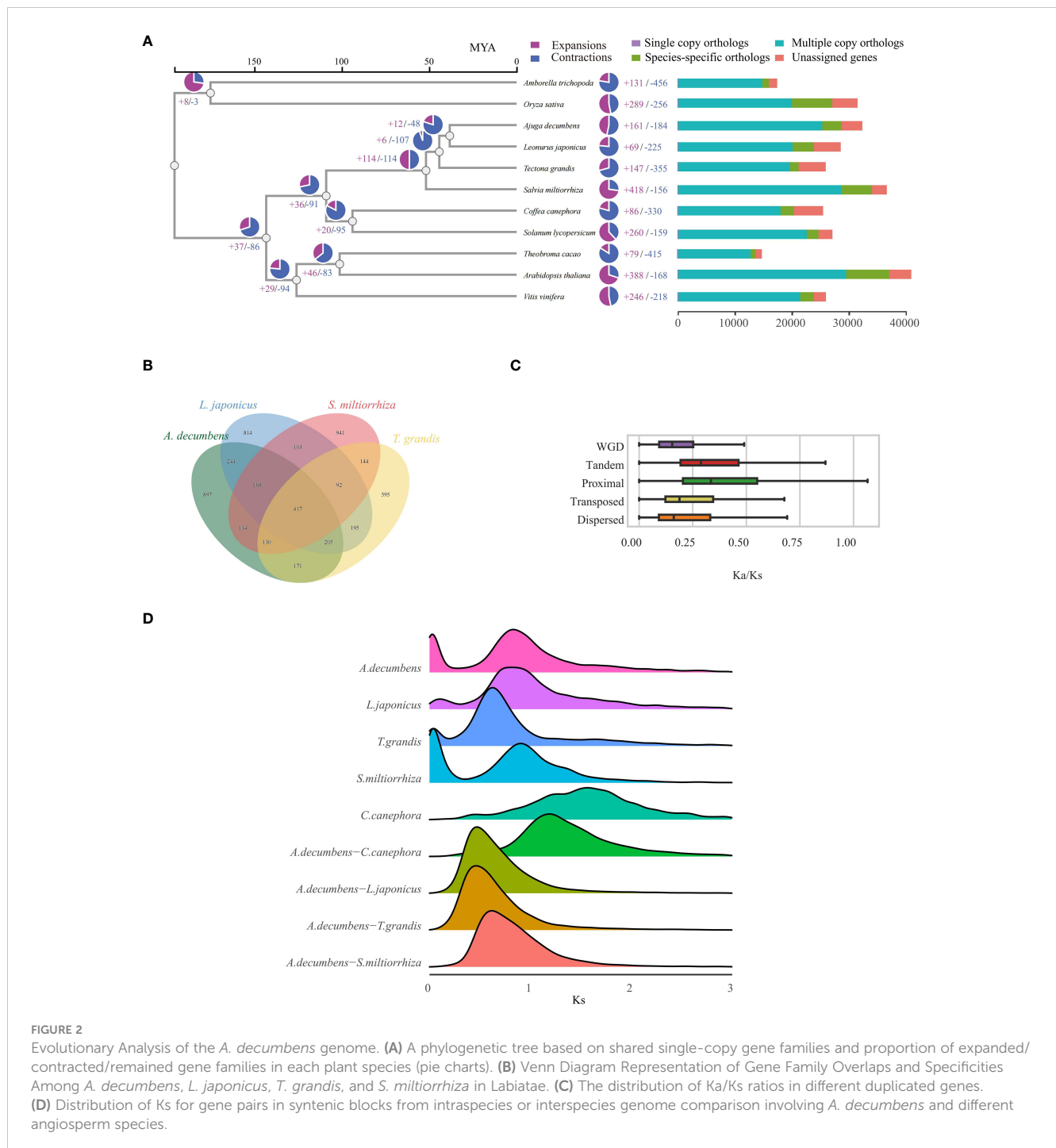
Our comparative study of *A. decumbens*, *L. japonicus*, *T. grandis*, and *S. miltiorrhiza* identified 417 common gene families among these species. 897 gene families (3,375 genes) were unique to *A. decumbens*. We discovered 15,017 LTR retrotransposons in *A. decumbens*, with a notable burst of retrotransposition around 100,000 years ago (Supplementary Figure S1). The genome showed predominant Dispersed Duplication (DSD) genes, with

Tandem Duplication (TD) genes also being significant (Supplementary Figure S2). We calculated non-synonymous (Ka) and synonymous (Ks) substitution rates and their ratios (Ka/Ks) for gene pairs from various duplication types. This revealed distinct trends in Ka/Ks ratios across duplication modes (Figure 2C), where TD and Proximal Duplication (PD) genes had higher ratios compared to others, especially Whole Genome Duplication (WGD) genes with the lowest ratios. The Ks distribution indicates a significant whole genome duplication (WGD) in the *A. decumbens* genome, with a prominent Ks peak at 0.8 (Figure 2D). Compared to *A. decumbens*, *C. canephora* lacks the γ event of whole genome duplication. Dot plots, collinearity maps, and depth charts (Supplementary Figures S3-S5) demonstrate a 2:1 collinearity relationship between *A. decumbens* and *C. canephora*. This event is widespread in the Lamiaceae family (Figure 2D). Additionally, PD and TD genes were significantly enriched in diterpenoid and terpenoid metabolic processes (Supplementary Figures S6, S7), underscoring their importance in *A. decumbens*' terpene metabolism.

Discussion

The genus Ajuga holds significant medicinal and economic value, yielding a diverse array of compounds, including phytosteroids, new iridoid and diterpene compounds, triterpenes, sterols, anthocyanin-glucosides, iridoid glycosides, solanesol, flavonoids, triglycerides, and essential oils. In recent decades, most research has centered on the pharmacology of *A. decumbens*, with genomics of the species receiving minimal attention.

The genomic research on *A. decumbens* has laid the foundation for future comparative genomics studies. It has opened new avenues for enhancing our understanding of the genetic diversity, evolutionary relationships, and functional genomics of *A. decumbens*. *A. decumbens*' genome data assists researchers in understanding the genetic diversity within the genus, uncovering unique genetic traits that enable adaptation to diverse environments, and further exploring how these genetic variations promote phenotypic diversity and ecological adaptability. According to the CCDB database (Rice et al., 2015), the chromosome numbers in the genus Ajuga include 8, 14, 15, 16, 18, 31, and 32, et al. The phenotypes of different species within this genus vary, particularly in flower color and morphology. Investigating how variations in chromosome numbers and corresponding genome sequences within this genus lead to the formation of different species is a valuable research question at the comparative genomic level. Comparing the genomes of different species from the Ajuga genus allows tracing their evolutionary paths. It clarifies the evolutionary history of traits related to medicinal properties, including divergence times and evolutionary pressures. Furthermore, comparative genomics helps identify genes responsible for synthesizing specific bioactive compounds and understanding the biosynthetic pathways of key medicinal compounds. This knowledge can enhance the production of desired compounds in *A. decumbens* plants or model organisms through metabolic engineering. Insights from genomic research can

FIGURE 2

Evolutionary Analysis of the *A. decumbens* genome. **(A)** A phylogenetic tree based on shared single-copy gene families and proportion of expanded/contracted/remained gene families in each plant species (pie charts). **(B)** Venn Diagram Representation of Gene Family Overlaps and Specificities Among *A. decumbens*, *L. japonicus*, *T. grandis*, and *S. miltiorrhiza* in Labiatae. **(C)** The distribution of Ka/Ks ratios in different duplicated genes. **(D)** Distribution of Ks for gene pairs in syntenic blocks from intraspecies or interspecies genome comparison involving *A. decumbens* and different angiosperm species.

also be used for genetic modification of *A. decumbens* to enhance its stress tolerance and medicinal value. Additionally, genomic data lays the groundwork for synthetic biology, enabling the synthesis of valuable compounds in microbial or other plant systems, offering environmentally sustainable production methods. This research has generated high-quality genomic and transcriptomic data for *A. decumbens*. In-depth studies of the *A. decumbens* genome have deepened our understanding of plant genetics and biochemistry and created tremendous potential for advances in pharmacology, plant biology, biotechnology, and practical applications in medicine and agriculture.

Recent advancements in high-throughput sequencing technologies and decreased costs have shifted the focus from metabolite-based studies to in-depth whole-genome research in various medicinal plants (Chen et al., 2021). Many medicinal plant genomes are highly repetitive or heterozygous, complicating high-quality genome assembly. We employed a hybrid sequencing strategy to achieve the first genome assembly of *A. decumbens*. This genome is substantial (1143 Mb) with 16 pseudochromosome sequences. We annotated 32,452 genes in *A. decumbens*, of which 78.11% were found to be homologous to Arabidopsis sequences, leaving 22.89% as potentially unique to the species or lineage,

meriting further investigation. In summary, the assembled *A. decumbens* genome from this study is a valuable resource for genetic research in the Ajuga genus, aiding in identifying significant pharmacologically relevant metabolites.

Materials and methods

Plant materials and sequencing

Samples of wild *A. decumbens* for this study were obtained from Nanyang, Henan (33.293° N, 112.024° E) and cultivated under laboratory conditions of 25°C, 3000 lx, and a 16-hour light: 8-hour dark photoperiod. Young leaves were used for genomic DNA extraction, while stems, roots, and leaves were utilized for transcriptome sequencing. These samples were collected from the same individual plant. Immediately after collection, samples were plunged into liquid nitrogen, transferred to the laboratory, and kept at -80°C. High-molecular-weight genomic DNA was prepared using the cetyltrimethylammonium bromide (CTAB) method and purified with the Qiagen genomic kit (Qiagen, 13343). This DNA was used to construct PacBio HiFi sequencing libraries. A 20 kb insert library was created using the SMRTbell template preparation kit (Pacific Biosciences) and sequenced on the PacBio Revio platform in CCS mode. In the Hi-C sequencing procedure, fresh leaf samples underwent formaldehyde treatment for DNA-protein crosslinking. Chromatin was cleaved with MboI, and 5' overhangs were filled with biotinylated residues. After re-ligation, DNA was broken into roughly 350-bp pieces by sonication. The Hi-C library, made according to standard methods, was sequenced on the DNBSEQ platform in PE150 mode. We prepared a paired-end DNBSEQ library for RNA sequencing with 350-bp inserts using the NextEra DNA Flex Library Prep Kit (Illumina). This was subsequently sequenced on the DNBSEQ-T7 platform (MGI Tech).

Genome assembly

The *A. decumbens* genome assembly at the contig level was performed using Hifiasm version 0.19.5 (Cheng et al., 2021) with default settings. Raw Hi-C sequencing data were processed using fastp (Chen et al., 2018) version 0.20.1. Cleaned Hi-C paired-end reads were aligned to the assembled genome with Bowtie2 (Langmead and Salzberg, 2012) version 2.3.2 using parameters "–end-to-end –very-sensitive, -L 30". Hi-C reads alignment to the scaffolds utilized the Juicer pipeline, then processed with 3D-DNA (Dudchenko et al., 2017) using "–editor-repeat-coverage 20" as a parameter. The Juicebox tool (Durand et al., 2016) facilitated assembly visualization. Genome completeness and annotation were assessed with BUSCO (Manni et al., 2021) using the "embryophyta_odb10" configuration file in "genome" mode. Assembly continuity was assessed by calculating the Long Terminal Repeat Assembly Index (LAI) with LTR_retriever (Ou and Jiang, 2018) using default settings.

Genome annotation

Before gene annotation, EDTA (Ou et al., 2019) version 1.8.4 was used to predict repetitive sequences on the chromosome-level assembly. LTR_Finder (Xu and; Wang, 2007) and LTRharvest (Ellinghaus et al., 2008) were employed to identify LTRs. LTR_retriever (Ou & Jiang, 2018) was used to integrate the results and estimate insertion times. Repeated DNA sequences were predicted using TIR-Learner (Su et al., 2019) and HelitronScanner (Xiong et al., 2014). The final repetitive sequences were annotated using RepeatMasker (Nishimura, 2000).

In gene prediction experiments across 11 species' genomes, BRAKER3 outperformed Funannotate and MAKER2 (Gabriel et al., 2023a; Gabriel et al., 2023b). BRAKER3 has been successfully applied in gene prediction for approximately 25 species. We harnessed the BRAKER3 pipeline (Gabriel et al., 2023b), integrating RNA-Seq and homology-based approaches for predicting protein-coding genes. First, clean RNA-Seq reads (roots, stems, and leaves) were mapped to the genome using HISAT2 (Kim et al., 2019) (version 2.10.2) to obtain transcriptome mapping data. Protein sequences from OrthoDB (Kriventseva et al., 2019) version 10.0 were downloaded and aligned to the genome assembly using ProtHint version 2.6.0. Then, we used GeneMark-EP+ (Brůna et al., 2020) version 4.65 to integrate the two types of data. Functional annotation of protein-coding genes was conducted via BLASTP (Mahram and Herbordt, 2015), targeting an E-value of < 1e−5 and aligning to prominent public databases such as NCBI NR and Swiss-Prot. Following this, protein domains were identified utilizing InterProScan (Jones et al., 2014) version 4.8. Additionally, EggNOG-mapper (Huerta-Cepas et al., 2019) version 5.0 was employed to determine clusters of orthologous groups (COG).

Non-coding RNAs include rRNAs, tRNAs, snRNAs, and miRNAs. tRNA is predicted using tRNAscan-SE (Chan et al., 2021) version 1.3.1. Due to the high homology of rRNA across different species, rRNA homology searches are performed using Blastn with Arabidopsis rRNA sequences. miRNAs and snRNAs in the genome are identified using INFERNAL (Nawrocki and Eddy, 2013) and PFAM database.

Phylogenetic analysis

Download the protein sequences of *A. trichopoda*, *O. sativa*, *V. vinifera*, *T. cacao*, *A. thaliana*, *S. lycopersicum*, *C. canephora*, *T. grandis*, *L. japonicus*, *S. miltiorrhiza*, and *A. decumbens* for the following analysis. Orthologous, phylogenetic, and gene family analyses were conducted using OrthoVenn3 (Sun et al., 2023). We used a BLASTP (Mahram and Herbordt, 2015) E-value cutoff of 1E^{-5} and OrthoMCL (Li et al., 2003) Markov clustering to determine pairwise sequence similarity. The phylogenetic tree construction employed FastTree2 (Price et al., 2010) utilizing the JTT+CAT model via the maximum likelihood method, supplemented by the SH test for node reliability assessment. We used single-copy genes and fossil evidence to construct the

divergence tree. r8s (Sanderson, 2003) was used to estimate divergence times for *A. thaliana* with *T. cacao*, *S. lycopersicum* with *C. canephora*, *A. thaliana* with *V. vinifera*, *A. trichopoda* with *V. vinifera*, and *L. japonicus* with *T. grandis*. Gene family expansion and contraction were determined using CAFE (Mendes et al., 2020) version 5 by comparing differences in cluster sizes between ancestors and each species. A random birth and death model was used to assess changes in gene families for each lineage in the phylogenetic tree. Conditional likelihood was used as a test statistic, and a P-value equal to or less than 0.01 was considered significant.

Whole-genome duplication analysis

Protein sequences from *A. decumbens* were analyzed against its protein sequences to identify syntenic blocks. Comprehensive BLASTP analysis, with an e-value threshold of 10^{-10}, was employed, followed applying the DupGen_finder pipeline (Qiao et al., 2019) to delineate inter-species syntenic blocks under standard settings. Rates of non-synonymous (Ka) and synonymous (Ks) substitutions, along with their ratio (Ka/Ks), for duplicated gene pairs were computed using the YN model in KaKs_Calculator 2.0 (Wang et al., 2010), following amino acid to codon alignment conversion via PAL2NAL (Suyama et al., 2006) v14.

Data availability statement

The data presented in the study are deposited in the NCBI repository (PRJNA1042970) and Figshare database (https://doi.org/10.6084/m9.figshare.24596520.v1).

Author contributions

YG: Writing – review & editing, Writing – original draft. JL: Writing – review & editing, Writing – original draft. YX: Writing – review & editing, Writing – original draft. TZ: Writing – review & editing, Writing – original draft. KT: Writing – review & editing,

Writing – original draft. XL: Writing – review & editing, Writing – original draft. LY: Writing – review & editing, Writing – original draft.

Funding

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.2024.1413468/full#supplementary-material

References

Atay, I., Kirmizibekmez, H., Kaiser, M., Akaydin, G., Yesilada, E., and Tasdemir, D. (2016). Evaluation of in *vitro* antiprotozoal activity of Ajuga laxmannii and its secondary metabolites. *Pharm. Biol.* 54, 1808–1814. doi: 10.3109/13880209.2015.1129542

Bouderbala, S., Lamri-Senhadji, M., Prost, J., Lacaille-Dubois, M., and Bouchenak, M. (2008). Changes in antioxidant defense status in hypercholesterolemic rats treated with Ajuga iva. *Phytomedicine* 15, 453–461. doi: 10.1016/j.phymed.2007.10.001

Brůna, T., Lomsadze, A., and Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics Bioinf.* 2, lqaa026. doi: 10.1093/nargab/lqaa026

Cárdenas, J., Esquivel, B., Gupta, M., Ray, A., Rodríguez-Hahn, L., Rodríguez-Hahn, L., et al. (1994). Clerodane diterpenes in labiatae. *Fortschr. der Chemie Organischer Naturstoffe/Progress Chem. Organic Natural Prod.* 63, 107–196.

Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49, 9077–9096. doi: 10.1093/nar/gkab688

Chang, Y., Zhang, R., Ma, Y., and Sun, W. (2023). A haplotype-resolved genome assembly of Rhododendron vialii based on PacBio HiFi reads and Hi-C data. *Sci. Data* 10, 451. doi: 10.1038/s41597-023-02362-1

Chen, H., Liu, D. Q., Zhang, L. X., Xia, Z. H., Yang, L., Liu, Z. L., et al. (1999). Two new clerodane diterpenes with antibacterial activity from Ajuga lupulina. *Indian J. Chem* B1999, 38, 743–745

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560

Chen, Y., Nie, F., Xie, S.-Q., Zheng, Y.-F., Dai, Q., Bray, T., et al. (2021). Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* 12, 60. doi: 10.1038/s41467-020-20236-7

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18, 170–175. doi: 10.1038/s41592-020-01056-5

Chenni, A., Yahia, D. A., Boukortt, F., Prost, J., Lacaille-Dubois, M., and Bouchenak, M. (2007). Effect of aqueous extract of Ajuga iva supplementation on plasma lipid

profile and tissue antioxidant status in rats fed a high-cholesterol diet. *J. Ethnopharmacol.* 109, 207–213. doi: 10.1016/j.jep.2006.05.036

Choi, B., Yang, S., Song, J. H., and Jang, T. S. (2019). Karyotype and genome size variation in Ajuga L.(Ajugoideae–Lamiaceae). *Nordic J. Bot.* 37, e02337. doi: 10.1111/njb.02337

College, J. N. M. (1977). Dictionary of Chinese materia medica. *Journal.*

Debell, A., Makonnen, E., Zerihun, L., Abebe, D., and Teka, F. (2005). *In-vivo* antipyretic studies of the aqueous and ethanol extracts of the leaves of Ajuga remota and Lippia adoensis. *Ethiopian Med. J.* 43, 111–118.

Dong, B., Yang, X., Liu, W., An, L., Zhang, X., Tuerhong, M., et al. (2020). Anti-inflammatory neo-clerodane diterpenoids from Ajuga pantantha. *J. Natural Prod.* 83, 894–904. doi: 10.1021/acs.jnatprod.9b00629

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

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., 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

El Hilaly, J., and Lyoussi, B. (2002). Hypoglycaemic effect of the lyophilised aqueous extract of Ajuga iva in normal and streptozotocin diabetic rats. *J. Ethnopharmacol.* 80, 109–113. doi: 10.1016/S0378-8741(01)00407-X

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinf.* 9, 1–14. doi: 10.1186/1471-2105-9-18

Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., et al. (2023a). BRAKER3: Fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *Biorxiv* 30, 2023.06.10.544449. doi: 10.1101/2023.06.10.544449

Gabriel, L., Hoff, K. J., Bruna, T., Lomsadze, A., Borodovsky, M., and Stanke, M. (2023b). "The BRAKER3 genome annotation pipeline," in *Plant and Animal Genomes Conference.* 30.

Graham, J., Quinn, M., Fabricant, D., and Farnsworth, N. (2000). Plants used against cancer–an extension of the work of Jonathan Hartwell. *J. Ethnopharmacol.* 73, 347–377. doi: 10.1016/S0378-8741(00)00341-X

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Jbilou, R., Amri, H., Bouayad, N., Ghailani, N., Ennabili, A., and Sayah, F. (2008). Insecticidal effects of extracts of seven plant species on larval development, α-amylase activity and offspring production of Tribolium castaneum (Herbst)(Insecta: Coleoptera: Tenebrionidae). *Bioresource Technol.* 99, 959–964. doi: 10.1016/j.biortech.2007.03.017

Jbilou, R., Ennabili, A., and Sayah, F. (2006). Insecticidal activity of four medicinal plant extracts against Tribolium castaneum (Herbst)(Coleoptera: Tenebrionidae). *Afr. J. Biotechnol.* 5 (10).

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Kariba, R. M. (2001). Antifungal activity of Ajuga remota. *Fitoterapia* 72, 177–178. doi: 10.1016/S0367-326X(00)00280-X

Kim, N. H., Jayakodi, M., Lee, S. C., Choi, B. S., Jang, W., Lee, J., et al. (2018). Genome and evolution of the shade-requiring medicinal herb Panax ginseng. *Plant Biotechnol. J.* 16, 1904–1917. doi: 10.1111/pbi.12926

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Korkina, L., Mikhal'Chik, E., Suprun, M., Pastore, S., and Dal Toso, R. (2007). Molecular mechanisms underlying wound healing and anti-inflammatory properties of naturally occurring biotechnologically produced phenylpropanoid glycosides. *Cell Mol. Biol.* 53, 84–91.

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., et al. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* 47, D807–D811. doi: 10.1093/nar/gky1053

Kuria, K. A., Chepkwony, H., Govaerts, C., Roets, E., Busson, R., De Witte, P., et al. (2002). The Antiplasmodial Activity of Isolates from Ajuga r emota. *J. Natural Prod.* 65, 789–793. doi: 10.1021/np0104626

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liao, B., Shen, X., Xiang, L., Guo, S., Chen, S., Meng, Y., et al. (2022). Allele-aware chromosome-level genome assembly of Artemisia annua reveals the correlation between ADS expansion and artemisinin yield. *Mol. Plant* 15, 1310–1328. doi: 10.1016/j.molp.2022.05.013

Liu, W., Song, Z., Wang, H., Yang, X., Joubert, E., Zhang, J., et al. (2020). Diterpenoids as potential anti-inflammatory agents from Ajuga pantantha. *Bioorg. Chem.* 101, 103966. doi: 10.1016/j.bioorg.2020.103966

Mahram, A., and Herbordt, M. C. (2015). NCBI BLASTP on high-performance reconfigurable computing systems. *ACM Trans. Reconfigurable Technol. Syst. (TRETS)* 7 (4), 1–20. doi: 10.1145/2629691

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38, 4647–4654. doi: 10.1093/molbev/msab199

Mendes, F. K., Vanderpool, D., Fulton, B., and Hahn, M. W. (2020). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518. doi: 10.1093/bioinformatics/btaa1022

Nakandala, U., Masouleh, A. K., Smith, M. W., Furtado, A., Mason, P., Constantin, L., et al. (2023). Haplotype resolved chromosome level genome assembly of Citrus australis reveals disease resistance and other citrus specific genes. *Hortic. Res.* 10, uhad058. doi: 10.1093/hr/uhad058

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509

Nishimura, D. (2000). RepeatMasker. *Biotech. Softw. Internet Rep.* 1, 36–39. doi: 10.1089/152791600319259

Ono, Y., Fukaya, Y., Imai, S., and Yamakuni, T. (2008). Beneficial effects of Ajuga decumbens on osteoporosis and arthritis. *Biol. Pharm. Bull.* 31, 1199–1204. doi: 10.1248/bpb.31.1199

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., Su, W., Liao, Y., Chougule, K., Agda, J. R., Hellinga, A. J., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20, 1–18. doi: 10.1186/s13059-019-1905-y

Park, H. Y., Kim, D. H., and Sivanesan, I. (2017). Micropropagation of Ajuga species: a mini review. *Biotechnol. Lett.* 39, 1291–1298. doi: 10.1007/s10529-017-2376-4

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS One* 5, e9490. doi: 10.1371/journal.pone.0009490

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, 1–23. doi: 10.1186/s13059-019-1650-2

Qing, X., Yan, H.-M., Ni, Z.-Y., Vavricka, C. J., Zhang, M.-L., Shi, Q.-W., et al. (2017). Chemical and pharmacological research on the plants from genus Ajuga. *Heterocyclic Commun.* 23, 245–268. doi: 10.1515/hc-2017-0064

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., et al. (2015). The Chromosome Counts Database (CCDB)–a community resource of plant chromosome numbers. *New Phytol.* 206, 19–26. doi: 10.1111/nph.13191

Sanderson, M. J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19, 301–302. doi: 10.1093/bioinformatics/19.2.301

Shafi, N., Khan, G. A., Arfan, M., and Gilani, N. (2004). Isolation and pharmacological screening of 8-O-acetyl harpagide from Ajuga bracteosa Wall. *Biol. Sci. PJSIR* 47, 176–179.

Shen, Q., Zhang, L., Liao, Z., Wang, S., Yan, T., Shi, P., et al. (2018). The genome of Artemisia annua provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. *Mol. Plant* 11, 776–788. doi: 10.1016/j.molp.2018.03.015

Song, Z., Lin, C., Xing, P., Fen, Y., Jin, H., Zhou, C., et al. (2020). A high-quality reference genome sequence of Salvia miltiorrhiza provides insights into tanshinone synthesis in its red rhizomes. *Plant Genome* 13, e20041. doi: 10.1002/tpg2.20041

Su, W., Gu, X., and Peterson, T. (2019). TIR-Learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. *Mol. Plant* 12, 447–460. doi: 10.1016/j.molp.2019.02.008

Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., and Wang, Y. (2023). OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* 51 (W1), W397–W403. doi: 10.1093/nar/gkad313

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

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

Wessner, M., Champion, B., Girault, J.-P., Kaouadji, N., Saidi, B., and Lafont, R. (1992). Ecdysteroids from Ajuga iva. *Phytochemistry* 31, 3785–3788. doi: 10.1016/S0031-9422(00)97527-7

Xiong, W., He, L., Lai, J., Dooner, H. K., and Du, C. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc. Natl. Acad. Sci.* 111, 10263–10268. doi: 10.1073/pnas.1410068111

Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., et al. (2016). Analysis of the genome sequence of the medicinal plant Salvia miltiorrhiza. *Mol. Plant* 9, 949–952. doi: 10.1016/j.molp.2016.03.010

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

Yang, H., Chen, H., Ni, Y., Li, J., Cai, Y., Ma, B., et al. (2022). *De novo* hybrid assembly of the salvia miltiorrhiza mitochondrial genome provides the first evidence of the multi-chromosomal mitochondrial DNA structure of salvia species. *Int. J. Mol. Sci.* 23, 14267. doi: 10.3390/ijms232214267