



## OPEN ACCESS

## EDITED BY

Cristina Vettori,  
National Research Council (CNR), Italy

## REVIEWED BY

Ruirui Huang,  
University of San Francisco, United States  
Hukam C. Rawal,  
University of Nevada, United States

## \*CORRESPONDENCE

Yiqiang Wang  
✉ wangyiqiang12@csuoft.edu.cn  
Meng Li  
✉ limeng0422@csuoft.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

RECEIVED 21 February 2025

ACCEPTED 17 July 2025

PUBLISHED 11 August 2025

## CITATION

Yao Z, Feng Z, Wu F, Zhang P, Wang Q,  
Ai B, Wang Y and Li M (2025) The  
near-complete genome assembly of  
*Ampelopsis grossedentata* provides  
insights into its origin, evolution, and  
the regulation of flavonoid biosynthesis.  
*Front. Plant Sci.* 16:1580779.  
doi: 10.3389/fpls.2025.1580779

## COPYRIGHT

© 2025 Yao, Feng, Wu, Zhang, Wang, Ai, Wang  
and Li. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). 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.

# The near-complete genome assembly of *Ampelopsis grossedentata* provides insights into its origin, evolution, and the regulation of flavonoid biosynthesis

Zhi Yao<sup>1,2†</sup>, Zhi Feng<sup>1,2†</sup>, Fuwen Wu<sup>1,2</sup>, Peiling Zhang<sup>1,2</sup>,  
Qiyue Wang<sup>3</sup>, Binling Ai<sup>4</sup>, Yiqiang Wang<sup>1,2\*</sup> and Meng Li<sup>1,2\*</sup>

<sup>1</sup>Key Laboratory of Forestry Biotechnology of Hunan Province, Central South University of Forestry and Technology, Changsha, China, <sup>2</sup>Uelushan Laboratory Carbon Sinks Forests Variety Innovation Center, Central South University of Forestry and Technology, Changsha, China, <sup>3</sup>College of Biological, Hunan Normal University, Changsha, China, <sup>4</sup>Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

*Ampelopsis grossedentata*, native to southern China, is renowned for its therapeutic and nutritional benefits, often called the “king of flavonoids” due to its high dihydromyricetin content. The dried stems, leaves, and shoot tips, known as “vine tea,” are consumed as a health beverage and traditional remedy for colds and fever. In this study, we assembled a near-complete reference genome of *A. grossedentata* spanning 555.42 Mb, where Hi-C-based correction resolved 18 out of its 20 chromosomes into gap-free assemblies. The genome, anchored to 20 chromosomes, comprises 44 contigs with an N50 of 21.93 Mb and 28 scaffolds with an N50 of 30.45 Mb, containing 25,999 protein-coding genes and 62.62% repetitive sequences. The *A. grossedentata* experienced two whole-genome duplication (WGD) events: a whole-genome triplication event shared by the core angiosperms and a WGD event shared with Vitaceae family. Through transcriptome-metabolome integrated analysis, *AgF3H1* gene was identified as playing a crucial role in the biosynthesis of dihydromyricetin (a flavanone) in *A. grossedentata*. The *AgF3H* gene is essential for converting pentahydroxy flavones to dihydromyricetin within the flavonoid biosynthesis pathway in *A. grossedentata*, as confirmed by molecular docking results. Thus, we postulate that *AgF3H1* serves as a pivotal regulatory gene in the dihydromyricetin biosynthetic pathway of *A. grossedentata*. These insights offer valuable genetic resources for the molecular breeding of *A. grossedentata* and enhance our comprehension of Vitaceae genomic evolution and flavonoid biosynthesis regulation in medicinal and nutritional plants.

## KEYWORDS

*Ampelopsis grossedentata*, reference genome, WGD, WGCNA, *AgF3H1*

# 1 Introduction

*Ampelopsis grossedentata*, a unique Vitaceae family in China, is an ancient medicinal and edible plant mainly found in southern Yangtze River such as Guangdong, Hunan and Hubei (Gu et al., 2020; Cao et al., 2023). According to the “National Compilation of Chinese Herbal Medicines,” the whole plant is used for clearing heat, detoxifying, lowering liver and lowering blood pressure to treat cold, fever and hepatitis (Ma et al., 2019; Li et al., 2022). Using young stems and leaves, it is processed by enzyme inactivation, rolling and drying to make “vine tea,” also known as “Maoyan terry tea,” which is helpful to relieve cough, eliminate phlegm and dispel wind and dampness (Carneiro et al., 2021; Luo et al., 2023). Vine tea was first noted in the “Classic of Tea” and was traditionally consumed by Zhuang and Yao ethnic minorities before gaining popularity among other minorities like the Tujia, Dong, and Hakka (Wu et al., 2023).

Flavonoids are secondary metabolites widely existing in plants and play an important role in plant defense and development, while also possessing significant health and medicinal values, making them a focus of attention (Shen et al., 2022; Liu et al., 2023). Contemporary pharmacological research indicates that *A. grossedentata*’s tender stems, leaves, and shoot tips predominantly contain flavonoids (Zeng et al., 2023). With its total flavonoid content reaching 35%–45%, it is the plant with the highest known concentration of flavonoids, providing it significant commercial potential and vast market opportunities (Zhang et al., 2016; Zhang X. et al., 2019). Dihydromyricetin is a naturally occurring dihydroflavonol compound, primarily derived from *A. grossedentata* (Zeng et al., 2023). In the extract of *A. grossedentata*, dihydromyricetin accounts for about 35% of the total flavonoids, making them the most abundant flavonoid monomer compounds in the plant (Zhang et al., 2018; Hu et al., 2020). Dihydromyricetin has significant pharmacological effects such as antihyperglycemic (Chen et al., 2016; Ran et al., 2019), antioxidation (Ye et al., 2015; Xie et al., 2019), anti-tumor (Zhou et al., 2014; Guo et al., 2019), anti-inflammatory (Hou et al., 2015; Zhang et al., 2022), and antibacterial properties (Wu et al., 2017; Xiong et al., 2021). Therefore, *A. grossedentata* is commonly used as dietary supplements such as teas, beverages, and lozenges (Carneiro et al., 2020; Zhang et al., 2021). Additionally, *A. grossedentata* extract can inhibit melanin production and is widely used in skin-whitening products (Huang et al., 2016). Enhancing the flavonoid content in *A. grossedentata* is crucial for augmenting its medicinal and nutritional benefits.

The flavonoid production process involves anthocyanin, isoflavonoid, flavone, and flavonol pathways, initiated by enzymes chalcone synthase, chalcone isomerase, and flavanone 3-hydroxylase (F3H) (Zeng et al., 2013; Ni et al., 2020). The *F3H* gene encodes a key enzyme for flavonol biosynthesis, which catalyzes the conversion of flavanones into dihydrokaempferol, dihydroquercetin, and dihydromyricetin (Prescott and John, 1996). To date, *F3H* genes have been cloned from various plant species, including *Malus pumila* (Davies, 1993), *Vitis vinifera* (Sparvoli et al., 1994), *Arabidopsis thaliana* (Pelletier and Shirley, 1996), and *Glycine max* (Zabala and Vodkin, 2005). Current

research on *A. grossedentata* primarily focuses on its pharmacological effects, antioxidant activity, physiological and biochemical characteristics, transcriptome sequencing, and chloroplast genome analysis (Ye et al., 2015; Huang et al., 2016; Gu et al., 2020; Luo et al., 2023; Wu et al., 2023). Based on transcriptome sequencing, Li et al. (2020) and Yu et al. (2021) predicted key genes involved in the flavonoid and dihydromyricetin biosynthetic pathways in *A. grossedentata*, respectively. Zhang et al. (2022), discovered an *AgF3H* gene through RNA-seq, showed the full-length CDS from leaf cDNA, and confirmed its expression in *Saccharomyces cerevisiae*, providing insights into dihydromyricetin hydroxylation in *A. grossedentata*. Although multiple genes involved in flavonoid biosynthesis have been identified in *A. grossedentata*, the accumulation patterns of flavonoids in different tissues and molecular regulatory mechanisms remain unclear.

In this study, we utilized wild *A. grossedentata* samples from Yongshun County, Hunan Province (PZY009, Figure 1a). Using Illumina, PacBio, and ONT ultra-long reads combined with Hi-C technology, we constructed the first near-complete reference genome for *A. grossedentata*. We performed transcriptomic and metabolomic analyses on the roots, shoot tips, stems, and leaves of PZY009 (Figure 1b) at the same developmental stage to identify key enzyme genes involved in the flavonoid biosynthesis pathway. This research provides valuable insights into the evolution, molecular-assisted breeding, and chemical diversity of *A. grossedentata*’s bioactive compounds.

## 2 Materials and methods

### 2.1 Plant materials and genome sequencing

The roots, stems, leaves, and shoot tips of wild germplasm (PZY009) (Figure 1b) were collected from the *A. grossedentata* garden in Yongshun County, Xiangxi Autonomous Prefecture, Hunan Province (29°16′42″N, 109°53′17″E). Samples were flash-frozen with liquid nitrogen and stored at -80°C.

High-quality genomic DNA was extracted from *A. grossedentata* leaves using a modified CTAB method (Allen et al., 2006). DNA integrity and purity were verified by agarose gel electrophoresis and spectrophotometer. Illumina short-read genome libraries (2×150 bp) were prepared according to Illumina standard protocols and sequenced on the Illumina NovaSeq platform. ONT ultra-long read sequencing was performed on the PromethION sequencer. PacBio sequencing included generating long-read libraries from genomic DNA fragments up to 15kb, which were sequenced on the PacBio Sequel II platform to produce HiFi (CCS) reads. Hi-C libraries were constructed using the HindIII restriction enzyme and sequenced on the Illumina NovaSeq chromosome assembly platform.

For comprehensive transcriptome analysis, RNA was isolated from roots, shoot tips, stems, and leaves using a plant RNA extraction kit according to the standard protocol. Following Oxford Nanopore Technologies’ strand-switching protocol,

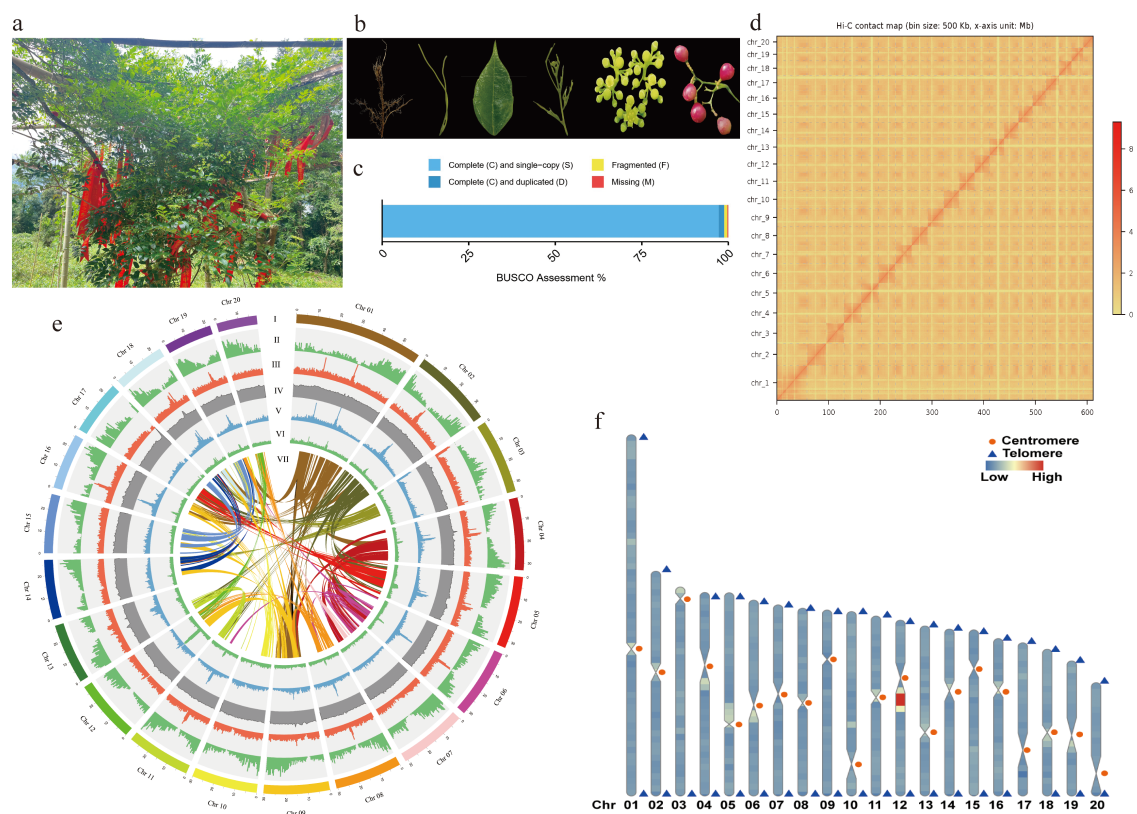


FIGURE 1

Comprehensive genome assembly of *A. grossedentata*. (a) Botanical overview of *A. grossedentata*. The red element is the silk used by local residents for blessings. (b) Microscopic images of *A. grossedentata* tissues. From left to right: root, stem, leaf, shoot tip, flower, and fruit. (c) BUSCO analysis of the *A. grossedentata* genome. (d) Hi-C heatmap of the assembled genome. Horizontal coordinate: Length of the chromosome; Vertical coordinate: The name of the chromosome; Legend: The frequency of interaction. The size of bin is 500kb. (e) Chromosomal features: (I) chromosome length; (II) gene density; (III) TE density; (IV) GC content; (V) LTR density; (VI) repeat coverage; (VII) synteny. All features are displayed at 500 kb resolution. (f) Telomere Identification: Triangles and circles denote telomeres and centromeres, with warm and cool tones indicating regions of high and low TE density, respectively.

mRNA was enriched to synthesize cDNA. PCR-amplified cDNA was sequenced using the PromethION sequencer.

## 2.2 Genome assembly and assessment

We evaluated genome size, heterozygosity, and repeat content using 17-mer k-mer distribution from Illumina short reads. Raw PacBio subreads were filtered and corrected via the PacBio circular consensus sequencing (pbccs) pipeline (<https://github.com/PacificBiosciences/ccs>) and assembled *de novo* using hifiasm software (v0.16.1-r375) (Cheng et al., 2021). Pilon software was used for primary contig correction. Genome assembly quality was assessed using BWA-MEM (v0.7.17) (Li and Durbin, 2009), CEGMA (Parra et al., 2007), and BUSCO (v5.2.2) (Simão et al., 2015). Hi-C chromosome conformations were captured using a DNase-based method and sequenced in 150 bp paired-end mode on Illumina NovaSeq (Ramani et al., 2020).

When processing Illumina DNA sequencing data, we utilized fastp (v0.21.0) software to perform filtering, removing low-quality reads and adapter sequences (Chen et al., 2018). Then, we utilized

K-mer-based analysis methods (Marçais and Kingsford, 2011) using Jellyfish (v2.2.7, <https://academic.oup.com/bioinformatics/article/27/6/764/234905?login=true>) and Genome Characteristics Estimation (GCE) software (Liu et al., 2013) to estimate genome size and heterozygosity rate. To confirm whether the sequencing data contained contamination, we used blast+ (Camacho et al., 2009) to extract the first 50,000 reads and compared them with the NT nucleotide sequence database. Finally, we used MEGAN (Huson et al., 2016) for species classification.

We used Filtlong (v 0.2.1, <https://biocontainers.pro/tools/filtlong>) and Porechop (v0.2.4, <https://github.com/rrwick/Porechop>) to remove short reads (<10 kb) and adapter sequences, then used NextDenovo for preliminary assembly of ONT ultra-long reads. Subsequently, we corrected the ONT draft genome using Racon (<https://github.com/isovic/racon>) and ONT ultra-long sequencing data, as well as Pilon (v1.24, Release Pilon version 1.24 · broadinstitute/pilon) and Illumina sequencing data. For the PacBio HiFi draft genome assembly, we used CCS reads to filter out low-quality sequences, then assembled the genome with hifiasm (v0.16.1-r375).

We adopted Purge\_dups software ([https://link.zhihu.com/?target=https%3A/github.com/dfguan/purge\\_dups](https://link.zhihu.com/?target=https%3A/github.com/dfguan/purge_dups)) (Guan et al.,



2020) to clear haplotypes, and used minimap2 (v2.28) (Li, 2018) for mitochondrial and chloroplast sequence alignment, filtering out sequences with over 50% base pair alignment. Furthermore, we eliminated bacterial contamination using the BLAST refseq database, while also removing poorly supported contigs (McGinnis and Madden, 2004). In this step, the use of fastp (Chen et al., 2018) allowed us to filter the raw Hi-C sequencing data to obtain purer data. Then, HICUP (Wingett et al., 2015) mapped the clean data to the genome assembly, thereby removing unmapped reads, invalid pairs, and duplicates.

In the process of generating the genome draft, we used ALLHiC software (v 0.9.8) (Zhang X. T. et al., 2019) and successfully generated a 2n karyotype genome draft through agglomerative hierarchical clustering. Additionally, by utilizing 3D-DNA (Dudchenko et al., 2017) and Juicer (v1.5) (Durand et al., 2016b), we converted the interactions of contigs into specific binary files. This process was visualized using Juicebox (Durand et al., 2016a), guiding the manual ordering and orientation of contigs. Based on this, we manually removed redundant contigs according to interaction relationships, while filling gaps with 100 Ns. We also used HiCExplorer (Wolff et al., 2020) to plot the interaction strength and positional relationship of contigs.

## 2.3 Identification of telomeres and centromeres

To achieve reference genome assembly, we used winnowmap (v1.11) (Jain et al., 2020) with parameters  $k=15$  and  $-MD$  to align the fill-in data with the genome gap regions, in order to fill the gaps in the genome. If the alignment spans both ends of a gap, we select the longest and best alignment to replace the gap region. Next, we used Winnowmap2 (parameters:  $k=15$ , dissimilarity  $> 0.9998$ ,  $-MD$ ,  $ax\ map-pb$ ) (Jain et al., 2022) to align the revised genome containing filled gaps with HiFi reads  $\geq 10$  kbp in size. Telomere repeat sequences (AAACCCCT at the 5' and 3' ends) of all reads were searched, with the most abundant reads marked as reference and the rest as queries. Then, we used medaka\_consensus to assemble these reference and query sequences. Subsequently, we used nucmer (v3.1) (Kurtz et al., 2004) to replace the terminal sequences on each pseudo-chromosome. Finally, we used Racon and Pacbio HiFi reads to perform error correction on the near-complete reference genome assembly. Centromere and telomere identification were carried out using CentIER (v3.0) (Xu et al., 2024) and A Telomere Identification toolKit (tidk, v0.2.63) (Brown et al., 2025) with default parameters, respectively.

## 2.4 Genome annotation

We initially annotated tandem repeat sequences using Genome-wide Microsatellite Analyzing Toward Application (GMATA, v21, <https://github.com/XuewenWangUGA/GMATA.git>) and Tandem Repeats Finder (v4.10, TRF) (Benson, 1999). We integrated ab initio and homology-based methods to annotate transposable elements

(TEs) within the *A. grossedentata* genome. Specifically, we employed MITE-hunter (Han and Wessler, 2010) and RepeatModeler2 (v1.0.11) (Flynn et al., 2020) with default parameters to predict the ab initio repeat library of *A. grossedentata*. We then developed an LTR-RT library using LTRharvest (Ellinghaus et al., 2008) and LTR\_Finder (Ou and Jiang, 2019) with default settings, and created a non-redundant LTR-RT library through LTR\_retriever (Ou and Jiang, 2018). These libraries were compared with the TEclass repbase (v20170127; <https://www.girinst.org/>) (Zhuo and Feschotte, 2015) database to classify the repeat families. Finally, the LTR\_retriever, MITE-Hunter, and RepeatModeler2 (v1.0.11) libraries were merged and input into RepeatMasker (v4.0.7) (Chen, 2004) to annotate repetitive elements in each assembled genome. LTR, Copia, and Gypsy insertion times were estimated using LTR\_retriever with default parameters.

Gene structure was annotated using homology search, *de novo* prediction, and reference-guided transcriptome assembly. In the homology prediction process, we used blast+ (Camacho et al., 2009) to locate protein sequences on the reference genome, and then used Exonerate to predict transcripts and coding regions (Slater and Birney, 2005). Additionally, genes predicted by BUSCO were incorporated into the homology prediction results, which was done during genome quality assessment (Manni et al., 2021). For *de novo* gene prediction, we relied on Augustus (v3.3) (Stanke et al., 2008) and GlimmerHMM (Delcher et al., 2007) with default parameters, operating through training sets. For RNA-seq reads, we used fastp (Chen et al., 2018) for filtering and HISAT2 (Kim et al., 2019) for genome alignment. The alignment results were then used as input for Stringtie to obtain transcripts (Kovaka et al., 2019), followed by prediction using TransDecoder (<https://github.com/TransDecoder/TransDecoder>). For Nanopore RNA-seq reads, we used NanoFilt (v2.8.0, <https://github.com/wdecoster/nanofilt>) for filtering and then Pychopper (<https://github.com/epi2me-labs/pychopper>) to determine full-length sequences. Post error correction using racon, these full-length sequences were aligned to the genome using minimap (Li, 2016).

The alignment results were fed into Stringtie to obtain transcripts (Kovaka et al., 2019). Finally, all predicted gene sets were combined into one gene set through MAKER (Holt and Yandell, 2011) and further optimized to obtain the final gene set. Lastly, we used BUSCO (Simão et al., 2015) to verify the completeness of the genome annotation to ensure the reliability and accuracy of our work.

Protein functions were predicted by comparing protein sequences against multiple public databases using DIAMOND (Buchfink et al., 2015). Databases utilized included non-redundant database (Deng et al., 2006), Swiss-Prot (Boeckmann et al., 2003), eggNOG (<http://eggno5.embl.de/>), Gene Ontology (GO, <https://www.geneontology.org/>), and Kyoto Encyclopedia of Genes and Genomes (KEGG, <https://www.genome.jp/kegg/>) (Kanehisa and Goto, 2000). This comparative analysis aimed to identify associated gene functions, conserved motifs, and protein domains. Annotation was performed via KOBAS (Xie et al., 2011), and putative domains and GO terms of genes were identified using



InterProScan with default settings (Blum et al., 2021). BLAST+ (Camacho et al., 2009) was used to compare the EvidenceModeler-integrated (Haas et al., 2008) protein sequences against the four major public protein databases with an E value cutoff of  $1e-05$ , retaining results with the lowest E value.

Non-coding RNAs (ncRNAs) were classified into categories such as miRNAs, rRNAs, tRNAs, snoRNAs, and snRNAs. To identify ncRNAs, we employed two strategies: database searching and model prediction. tRNAs were predicted using tRNAscan-SE with eukaryote parameters (Chan et al., 2021). MicroRNAs, rRNAs, snRNAs, and snoRNAs were detected using Infernal cmscan (Nawrocki and Eddy, 2013) against the Rfam database. (<https://rfam.xfam.org/>). rRNAs and their subunits were predicted using RNAmmer (Lagesen et al., 2007).

## 2.5 Comparative genomic analysis

In comparative genomic studies, to ensure research accuracy, we selected species with high-quality genome assemblies that are phylogenetically closely related to the target species (*A. grossedentata*). Accordingly, three *Vitis* species (*V. vinifera*, *V. rotundifolia*, and *V. amurensis*) along with *Cissus rotundifolia* from the Vitaceae family were prioritized. Additionally, we included model plants (*Arabidopsis thaliana* and *Oryza sativa*) and multiple species with elevated flavonoid contents: *Glycine max*, *Papaver somniferum*, *Salvia miltiorrhiza*, *Solanum lycopersicum*, *Dendrobium officinale*, and *Scutellaria baicalensis*. *Nymphaea colorata* was designated as the outgroup based on its phylogenetic distance as a monocotyledonous plant from *A. grossedentata*, a eudicot species (Supplementary Table S13). Genetic family clustering of these 14 plant species was conducted using BLAST+ (Camacho et al., 2009) and OrthoFinder (Emms and Kelly, 2019). Gene families were annotated using the Panther database (Mi et al., 2019). Unique gene families for each species were identified through GO and KEGG enrichment analysis, facilitated by clusterProfiler (Yu et al., 2012). Single-copy orthologous genes were extracted and aligned using MUSCLE (Edgar, 2004), with alignment results filtered by TrimAl (Capella-Gutiérrez et al., 2009) and consolidated into a supermatrix alignment.

A Maximum-likelihood (ML) phylogenetic tree was constructed via RAxML employing the PROTGAMMAWAG model (Stamatakis, 2014). Divergence times between species were estimated using the MCMCTree program in PAML (Yang, 2007), with burn-in=10,000, sample-number=100,000, and sample-frequency=2, using calibration times from the TimeTree database (Kumar et al., 2022). Including: *N. colorata* – *O. sativa*: 168–191 Mya; *D. officinale* – *O. sativa*: 108–123 Mya; *P. somniferum* – *O. sativa*: 142–163 Mya; *P. somniferum* – *A. thaliana*: 126–136 Mya; *V. vinifera* – *A. thaliana*: 109–124 Mya; *G. max* – *A. thaliana*: 102–112 Mya; *S. lycopersicum* – *A. thaliana*: 111–123 Mya; *S. lycopersicum* – *S. miltiorrhiza*: 75–96 Mya; *S. baicalensis* – *S. miltiorrhiza*: 33–72 Mya; *V. vinifera* – *C. rotundifolia*: 31–96 Mya; *V. vinifera* – *V. rotundifolia*: 4–14 Mya; *V. vinifera* – *V. amurensis*: 5–40 Mya; *V. vinifera* – *C. rotundifolia*: 31–96 Mya. In brief, an all-against-all

BlastP search was performed on the 14 proteomes using DIAMOND (Buchfink et al., 2015) with a cutoff e-value of 10<sup>-5</sup>. HOGs were obtained using PhyloMCL (Zhou et al., 2020) with default parameters. For each HOG, PASTA (Tang and Riva, 2013) was used for multiple sequence alignment, and protein alignments were converted to nucleotide alignments. A maximum likelihood tree was reconstructed for each HOG using IQ-TREE2 (Minh et al., 2020) with 100 bootstrap replicates. The GDs on each gene tree were estimated using a previously described strategy (Ren et al., 2018) if nodes on the GF tree had >50% bootstrap support. Patterns of duplicate retention for GD candidates were counted for further evaluation.

Gene family evolution was modeled as a random birth and death process, with expansion and contraction rates of one gene per million years. CAFE software (Han et al., 2013) was used to predict gene family changes in *Ampelopsis cordata* relative to its ancestors, with a p-value threshold of 0.05 to identify significant size changes. Phylogenetic tree topology and branch lengths informed the significance of these changes.

Single-copy orthologous genes were aligned via MUSCLE (Edgar, 2004), and positive selection was analyzed using PAML CodeML (Yang, 2007), considering *A. grossedentata* as the foreground branch. P-values were determined using  $\chi^2$  statistics, with FDR correction for multiple testing.

Collinearity analysis involved using DIAMOND (Buchfink et al., 2015) to identify similar gene pairs between species ( $e < 1E-5$ , C-score > 0.5, filtered by JCVI software) (Tang et al., 2024). Adjacent similar gene pairs of chromosomes were determined based on the gff3 file, and collinear blocks were identified using MCScanX (parameters: -a -e 1e-5 -s 5) (Wang et al., 2012), with circular plots generated via the R package circlize (Gu et al., 2014). Syntenic blocks were identified using 'jvarkit.catalog.orthologs - cscore=0.7' (Tang et al., 2024), and genes from all collinear blocks were obtained.

To identify whole-genome duplications (WGD) in the *Ampelopsis cordata* genome, we utilized a comprehensive WGD and intra-genome collinearity detection tool, along with Ks estimation and peak fitting (Sun et al., 2022). The combined use of 4DTv and Ks values of syntenic regions is a widely accepted method for detecting WGD events. For *A. grossedentata*, WGD events were identified using the WGD software (Zwaenepoel and Van de Peer, 2019).

## 2.6 Transcriptome analysis

Adapters were initially filtered from the raw RNA short-reads, followed by the removal of poly(A) tails and low-quality reads ( $Q < 20$ ). The remaining high-quality reads were used to determine Q20, Q30, and GC content. These clean reads were then mapped to the reference genome and full-length transcript using HISAT (Kim et al., 2019). Reads with perfect matches or a single mismatch were utilized to reconstruct transcripts via StringTie (Kovaka et al., 2019). Expressed genes were identified based on mapping results. If reads aligned with annotated gene sequences, the gene was classified as existing and coded

accordingly. Otherwise, if reads aligned with the full-length transcript but not with any annotated sequence, the gene was considered novel and recorded as a new identification.

Gene expression was quantified using fragments per kilobase of transcript per million mapped reads (FPKM). Read counts from the sequenced library were normalized using a scaling factor in edgeR (Robinson et al., 2010). Differential expression of dgps paralogs across roots, stems, leaves, and shoot tips was analyzed with EBSeq, while DESeq2 (Anders and Huber, 2010) was used for cross-tissue comparisons. Significant differential expression was determined with  $FDR < 0.05$  and  $|\log_2(\text{foldchange})| \geq 2$ . GO and KEGG enrichment analyses of DEGs were performed via clusterProfiler (Wu et al., 2021). The PPI network was analyzed using NetworkAnalyst and STRING.

## 2.7 Widely targeted metabolomic analysis

Raw data were converted to mzXML format using MSConvert from the ProteoWizard software suite (Rasmussen et al., 2022) and processed in R with XCMS for feature detection (Navarro-Reig et al., 2015), retention time correction, and alignment. Metabolites were identified by matching accurate mass and MS/MS data with HMDB (Wishart et al., 2007), MassBank (Horai et al., 2010), Knapsack, ReSpec, LipidMaps (Sud et al., 2007), KEGG (<https://www.genome.jp/kegg/>), and a proprietary database from Panomix Biomedical Tech Co., Ltd. (Suzhou, China). Metabolite molecular weights were determined by the m/z ratios of parent ions. Molecular formulas were predicted using ppm and adduct ions and matched with databases for MS identification. MS/MS data was concurrently matched with fragment ions and database information for identification.

We employed two multivariate statistical analysis models, unsupervised (PCA) and supervised (PLS-DA, OPLS-DA), to differentiate groups using the R ropls package (Thévenot et al., 2015). Statistical significance was determined by P-value from group comparisons. Biomarker metabolites were filtered based on P-value, VIP (variable importance in projection from OPLS-DA), and fold change. Metabolites with  $P < 0.05$  and  $VIP > 1$  were deemed significantly differentially expressed.

Differential metabolites underwent pathway analysis via MetaboAnalyst (Xia and Wishart, 2011), integrating pathway enrichment and topology analyses. The identified metabolites were mapped to KEGG pathways for biological interpretation, and visualizations were created using the KEGG Mapper tool.

## 2.8 Weighted gene co-expression network analysis

For co-expression network analysis to detect high gene correlation modules, we used the WGCNA package in R (Zhang and Horvath, 2005). Modules associated with phenotypic traits were identified by converting the adjacency matrix to a topological overlap matrix and filtered using the WGCNA goodGenes

function. The hierarchical gene clustering tree was pruned with the cutreeDynamic function, and modules with correlation coefficients ( $r$ ) above 0.75 were merged. The gene co-expression network was constructed using the blockwiseModules function with an unsigned TOMType. Module eigengenes were computed via the WGCNA's module eigengenes function, and their association with phenotypic traits was evaluated using Pearson correlation analysis. Hub genes were identified using the CytoHubba plugin (Chin et al., 2014) in Cytoscape (Shannon et al., 2003).

## 2.9 Molecular docking of AgF3H genes

AlphaFold3 (Abramson et al., 2024) predicted the crystal structure of the AgF3H protein in *A. grossedentata*. The crystal structure was processed using the Protein Preparation Wizard module in Schrödinger for preprocessing, native ligand state regeneration, H-bond optimization, energy minimization, and water removal (Abramson et al., 2024). The 2D sdf files of pentahydroxyflavanone, naringenin, and eriodictyol were converted into 3D chiral conformations using the LigPrep module in Schrödinger. The SiteMap module pinpointed the optimal binding site, while the Receptor Grid Generation module configured the most appropriate enclosing box for this site, thereby defining the active site of the AgF3H proteins. Pentahydroxyflavanone, naringenin, and eriodictyol were docked to the active sites of AgF3H1 and AgF3H2 proteins using high-precision XP docking. MM-GBSA calculations provided the binding free energy (dG Bind) between the ligands and proteins, where lower values indicate more stable binding.

Per the manufacturer's guidelines, the M-MLV reverse transcriptase kit was utilized to synthesize the first-strand cDNA for qRT-PCR analysis. The qRT-PCR was conducted with the iTaq Universal SYBR Green super mix and recorded by the ABI 7500 PCR system. The procedure was replicated three times, with each iteration including standards and negative controls. The qRT-PCR protocol entailed a 30s denaturation at 95°C, followed by 40 cycles of 5s denaturation at 95°C, 30s annealing at 60°C, and a 20s extension at 60°C. Each sample was run thrice, and the qRT-PCR outcome was averaged from three replicate applications. The standard gene GAPDH of *A. grossedentata* served as the internal reference gene (Xu, 2017), with Ct values determining the relative expression of AgF3H1 (Livak and Schmittgen, 2001) (Supplementary Table S21).

## 3 Results

### 3.1 Sequencing and assembly of the *A. grossedentata* genome

We generated 70.8 Gb of high-quality paired-end reads on the Illumina platform for  $k$ -mer ( $k=17$ ) analysis to estimate the genome size of *A. grossedentata* (Table 1; Supplementary Figure S1; Supplementary Table S1). The final assembled genome size of *A. grossedentata* was 555.42 Mb, with a GC content of 31.98%, a repeat

sequence proportion of 62.62%, and a heterozygosity rate of 1.48% (Table 1 and Supplementary Table S1), indicating a highly heterozygous and repetitive genome.

We evaluated the genome assembly quality using sequence consistency and BUSCO metrics. Sequence consistency revealed a 99.02% alignment rate and 99.29% coverage rate of short reads to the *A. grossedentata* genome, indicating high consistency (Supplementary Table S2). BUSCO analysis showed 98.8% completeness for the 425 single-copy orthologous genes, confirming the high integrity of the assembly (Figure 1c).

The proportions of A, T, G, and C in the *A. grossedentata* genome were within normal ranges, with an N content of 0.00%, which is within the acceptable range (<10%) (Supplementary Table S3). The heterozygous SNP proportion was 0.2953%, and the homozygous SNP proportion was 6.7836e-05% (Supplementary Table S4), indicating high single-base accuracy. These results demonstrate that the *A. grossedentata* genome sequence has high consistency, accuracy, and completeness.

### 3.2 Hi-C technology assisted the assembly of near-complete reference genome of *A. grossedentata*

To achieve chromosome-level assembly, we utilized high-throughput chromosome conformation capture (Hi-C) sequencing technology, generating 69.8 Gb of 235.36 million paired-end Hi-C reads (Supplementary Table S1). Using Allhic software, we anchored 28 scaffolds, totaling 608.41 Mb of sequences, to the *A. grossedentata* genome (Supplementary Table S5). With the error correction and assembly assistance of Hi-C technology, we obtained 20 chromosome-level sequences, achieving a genome anchoring rate of 99.89% (Supplementary Table S6). Each chromosome contains at least one scaffold, with lengths ranging from 18.57 Mb (Chr 20) to 59.11 Mb (Chr 1) (Supplementary Table S7). The Hi-C interaction matrix heatmap demonstrated higher interaction intensity among adjacent sequences, with 20 pseudochromosomes aligned along the diagonal (Figure 1d). A circos map was drawn based on grape genome data (Figure 1e). Using 7-base telomere repeat sequences (AAACCCT) as queries, we identified 38 telomeres on 20 pseudochromosomes (except Chr 03 and Chr 17, each missing one telomere) and located potential centromeres on each chromosome. Detailed regions are listed in Supplementary Tables S8 and S9. The assembly is deemed a high-quality, near-complete genome (Figure 1f).

### 3.3 Repeat sequence prediction and genome annotation

Eukaryotic genomes' repeat sequences play a crucial role in evolution, inheritance, and life variation, making them vital for comprehensive analysis of gene expression control, genome structure, and species evolution. Using homology, *de novo*, and transcriptome predictions, we foresaw a total of 25,756 genes in *A. grossedentata*,

exceeding *N. colorata* (19,299), but trailing *V. vinifera* (29,591), *V. rotundifolia* (26,742), and *V. amurensis* (29,168) (Supplementary Figure S2a). The average gene length in *A. grossedentata* is 7895 bp, with an average exon length of 351 bp, an average intron length of 1241 bp, and an average coding region length of 1615 bp (Supplementary Table S10). Furthermore, 9848 genes (37.36%) and 25,990 genes (98.60%) had homologous gene predictions in the eggNOG and NR databases, respectively (Supplementary Table S11; Supplementary Figure S2b). Additionally, we identified 2,077 non-coding RNAs in the *A. grossedentata* genome, including 849 rRNAs, 497 tRNAs, 122 miRNAs, and 609 snoRNAs (Supplementary Table S12).

### 3.4 Comparative genomic analysis

We analyzed the genes in the *A. grossedentata* genome against 12 other species (*V. vinifera*, *V. rotundifolia*, *V. amurensis*, *C. rotundifolia*, *A. thaliana*, *G. max*, *D. officinale*, *P. somniferum*, *S. miltiorrhiza*, *S. lycopersicum*, *O. sativa*, and *S. baicalensis*) and one outgroup species (*N. colorata*) (Supplementary Table S13). We utilized the genomes of these species for homologous gene identification, gene family clustering, and analysis of single-copy gene enrichment in *A. grossedentata* (Figure 2a). The analysis revealed 27,473 gene families across the 14 species, with 6,738 gene families being conserved, including 175 single-copy gene families shared among all species (Figure 2a). We extracted clustering data for *C. rotundifolia*, *A. thaliana*, *V. vinifera*, *V. rotundifolia*, *V. amurensis*, *A. grossedentata*, and *S. lycopersicum* to create an upset plot (Figure 2b). The plot indicated that *A. grossedentata* possesses 193 unique gene families (comprising 1,075 genes) relative to other species. Enrichment analyses demonstrated that these unique gene families are predominantly involved in "metabolism" and "transport" pathways (Supplementary Figure S3).

Based on 175 single-copy homologous genes from 14 species, we constructed a high-confidence phylogenetic tree and estimated divergence times using the Bayesian relaxation molecular clock method (Figure 2c). Among these 14 species, *A. grossedentata* is most closely related to *C. rotundifolia*, diverging approximately 49.0 Mya. The lineage of *A. grossedentata* and the genus *Vitis* (*V. vinifera*, *V. rotundifolia*, *V. amurensis*) diverged from a common ancestor around 35.0 Mya (Figure 2c).

The expansion and contraction of gene families are pivotal in developing plant-specific traits and phenotypic diversity. Expanded gene families may acquire new functions, enhancing environmental adaptability. Analysis revealed that the MRCA (most recent common ancestor) had 27,183 gene families (Figure 2c). Compared to its closest ancestor, *A. grossedentata* significantly expanded 186 gene families (including 1,097 genes) and contracted 47 gene families (including 127 genes) (Figure 2c). GO enrichment analysis indicated that contracted gene families were enriched in "response to stimulus" and "methylation" (Supplementary Figure S4a), while expanded families were enriched in "secondary metabolite biosynthetic process" and "secondary metabolic process" pathways (Supplementary Figure S4b).



**TABLE 1** Genome assembly and annotation statistics of *A. grossedentata*.

Genome information	Value
Revised genome size (Mbp)	555.42
Chromosome number	20
Contig number	44
Scaffold number	28
Contig N50 (bp)	21,931,686
Scaffold N50 (bp)	30,449,182
GC content (%)	31.98%
Heterozygous ratio (%)	1.48%
Repeats (%)	62.62%
<i>K</i> -mer	17
Number of genes	26,359
Number of miRNAs	122
Number of rRNAs	849
Number of tRNAs	497
Number of snoRNAs	609
Complete BUSCOs (%)	98.8%

### 3.5 Whole-genome duplications of *A. grossedentata*

The 1:1 ratio of syntenic depth between the representative Vitaceae species (*C. rotundifolia*, *V. vinifera*, *V. rotundifolia*, and *V. amurensis*) and *A. grossedentata*, coupled with conserved syntenic patterns, indicates that both lineages experienced both recent and ancient whole-genome duplication (WGD) events (Figures 3a–d). Application of Tree2GD to the 14 genomes revealed 2 polyploidization events in the ancestor of *Vitales* Juss. ex Bercht. & J. Presl (11,681GDs) and that of *Rosales* Bercht. & J. Presl (15,800 GDs). And ratio of (AB)(AB) of the ancestor of *Vitales* (50.17%) and that of *Rosales* (51.14%) revealed strong phylogenomic signals for WGD. Tree2GD analysis was conducted on the genomes of 14 species, and it was found that *A. grossedentata* identified a total of 20,029 GDs (10.93%), indicating that *A. grossedentata* experienced two polyploid events (Figure 2c).

The distributions of 4DTv and *K*s revealed two distinct peaks, indicating the occurrence of two WGD events (Figures 3e, f). The most recent WGD event aligns with those in *C. rotundifolia*, *V. vinifera*, *V. rotundifolia*, and *V. amurensis*, indicating a common WGD event in Vitaceae species. The other WGD event corresponds to the whole-genome triplication (WGT,  $\gamma$  event) common to core eudicots (Jaillon et al., 2007).

To understanding the chromosome evolution and phylogenetic relationships among Vitaceae species, we conducted genomic collinearity analysis of *C. rotundifolia*, *V. vinifera*, *V. rotundifolia*,

*V. amurensis*, and *A. grossedentata* (Figure 3g). We observed fewer scattered dots in the comparison between *A. grossedentata* and *C. rotundifolia*, indicating a closer phylogenetic relationship between these two species compared to others (Figure 3g). Additionally, recombination and gene fragment rearrangement events were detected on chromosome 9 in *A. grossedentata*, including inversions and translocations (Figure 3g), which may have led to the high content of flavonoid compounds in *A. grossedentata*. Overall, these findings provide new insights into the chromosome evolution of *A. grossedentata* and will offer scientific evidence beyond the genus for studying important agronomic traits in *Vitis* species.

### 3.6 Integrated metabolome and transcriptome analyses

Flavonoids are widely present in Vitaceae plants, including *A. grossedentata*, and play various roles in secondary metabolism. Previous research identified 138 flavonoid-related genes and isoforms, partially elucidating the flavonoid biosynthesis pathway (Li et al., 2020) (Figure 4a). We conducted metabolomics analyses on the roots, shoot tips, stems, and leaves of *A. grossedentata* to determine the correlation between flavonoids and gene expression. The reproducibility of the sequencing data was confirmed by PCA, OPLS-DA, and PLS-DA results for three biological duplicates (Supplementary Figure S5). Utilizing high-resolution LC-MS/MS analysis, 1,526 metabolites were detected across four tissues, including flavonoids (70), terpenoids (41), and carboxylic acids and their derivatives (37) (Supplementary Figure S6; Supplementary Table S14). Half of the 70 flavonoids (35) were abundant in the flavonoid biosynthesis pathway (Ko00941, Supplementary Table S15). These flavonoids were mainly expressed in shoot tips except for four metabolites found in roots (8-Prenylaringenin, Pinocembrin, Chrysoplenol D, Apigenin) (Figure 4b). Earlier research indicated flavonoids synthesis mainly happens through dihydromyricetin accumulation, involving compounds like naringenin, dihydromyricetin, myricetin, and delphinidin (Yu et al., 2021). The differential expression metabolites in pairwise comparisons between the shoot tips and roots were most enriched in the flavonoid biosynthesis pathway (Supplementary Figure S7), with dihydromyricetin having the highest content in the shoot tips and showing the most significant differential expression between the shoot tips and roots (Figure 4c). Utilizing the near-complete of *A. grossedentata*, we performed transcriptome sequencing on the roots, stems, leaves, and shoot tips of PZY009 to identify key genes and transcription factors (TFs) involved in dihydromyricetin synthesis. The results of PCA for three biological replicates indicate high reproducibility of the sequencing data (Supplementary Figure S8). Our gene expression analysis discovered 10,566 genes (5,993 upregulated and 4,573 downregulated) when comparing shoot tips and roots (Supplementary Table S16; Supplementary Figure S9). GO and KEGG enrichment annotations revealed these genes' functions were chiefly enriched in metabolic and secondary metabolite

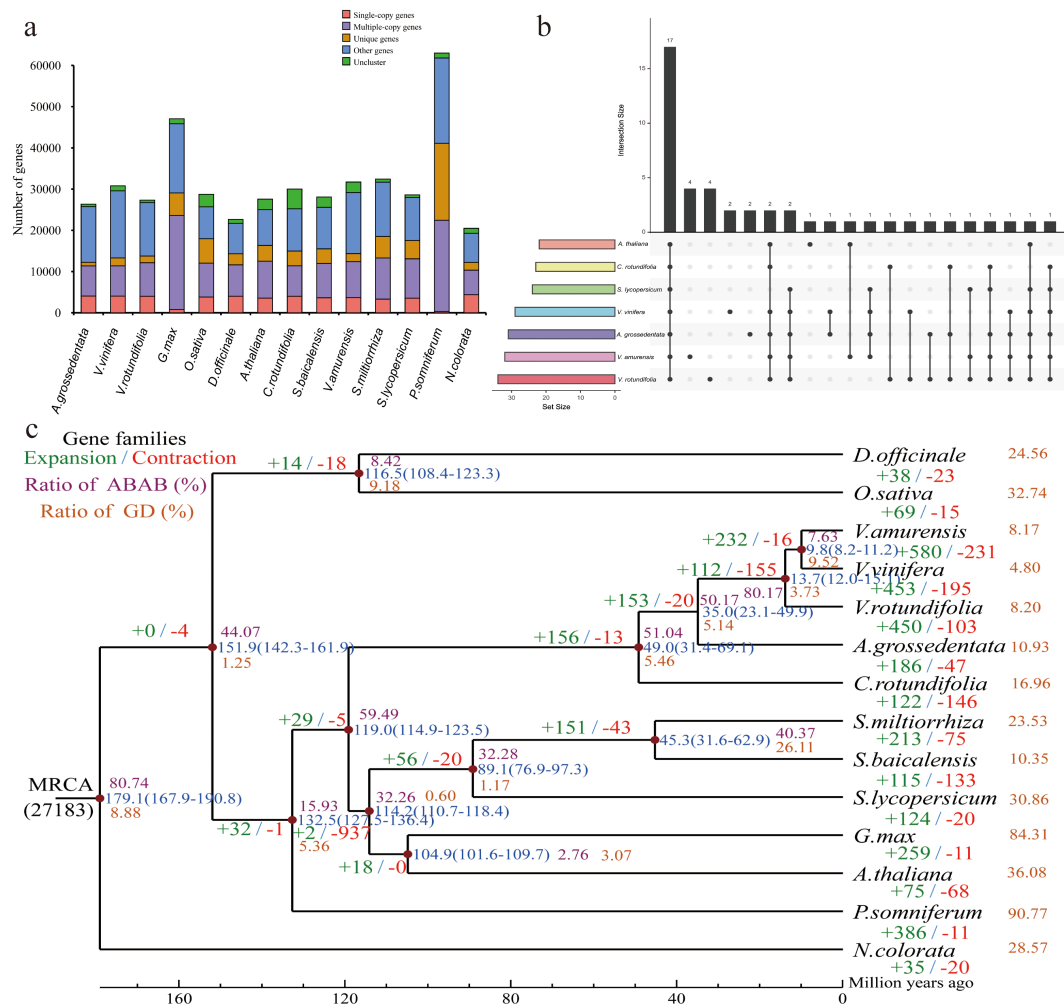


FIGURE 2

Gene family cluster and evolutionary analysis. (a) Gene family copy number across 14 species (*V. vinifera*, *V. rotundifolia*, *V. amurensis*, *C. rotundifolia*, *A. thaliana*, *G. max*, *D. officinale*, *P. somniferum*, *S. miltiorrhiza*, *S. lycopersicum*, *O. sativa*, *S. baicalensis*, *N. colorata* and *A. grossedentata*). (b) Upset diagram of shared and unique gene families among *A. grossedentata*, *V. vinifera*, *V. amurensis*, *V. rotundifolia*, *A. thaliana*, *C. rotundifolia* and *S. lycopersicum*. (c) Phylogenetic tree with divergence times and gene family expansion/contraction in 14 species. GD events in different eudicot lineages identified by reconciliation of gene trees and the species tree. Node numbers indicate estimated divergence times (Mya) with error ranges in blue parentheses; red and green numbers denote contracted and expanded gene families, respectively.

biosynthesis pathways, particularly flavonoid biosynthesis and isoflavonoid biosynthesis (Supplementary Figure S10).

### 3.7 Screening of key genes in the flavonoid compound biosynthesis pathway

Of 35 metabolites enriched in the flavonoid synthesis pathway (Ko00941), four key metabolites were identified: naringenin, dihydromyricetin, myricetin, and delphinidin. Therefore, we conducted WGCNA analysis using naringenin, dihydromyricetin, myricetin, and delphinidin with transcriptome data. Through WGCNA analysis, 17,085 genes from 12 samples were clustered into 9 modules (Figure 4d; Supplementary Figure S11; Supplementary Table S17). We focused on the yellow module (2,707 genes), brown module (3,234 genes), turquoise module (4,263 genes), and blue module (4,012

genes) (Figure 4d; Supplementary Table S17). Correlation analysis between genes and transcription factors in each module, using correlation coefficients  $\geq 0.9$  or  $\leq -0.9$ , identified 10 key enzyme genes and 364 transcription factors in the flavonoid biosynthesis pathway: yellow module (6 genes, 64 transcription factors), brown module (2 genes, 99 transcription factors), turquoise module (1 gene, 77 transcription factors), and blue module (1 gene, 124 transcription factors) (Supplementary Tables S18, S19). Tissue-specific expression profiling of *A. grossedentata* revealed pronounced spatial expression heterogeneity among ten screened key genes. Notably, the *AgF3H1* gene exhibited strong tissue-specific expression characteristics in shoot tips, demonstrating significantly higher transcriptional levels (FPKM values) compared to other examined tissues ( $P < 0.05$ ), with statistically significant differences. Systematic analysis of key structural genes in secondary metabolic pathways further uncovered marked tissue expression preferences: *AgFLS1* and *AgDFR* genes displayed high-

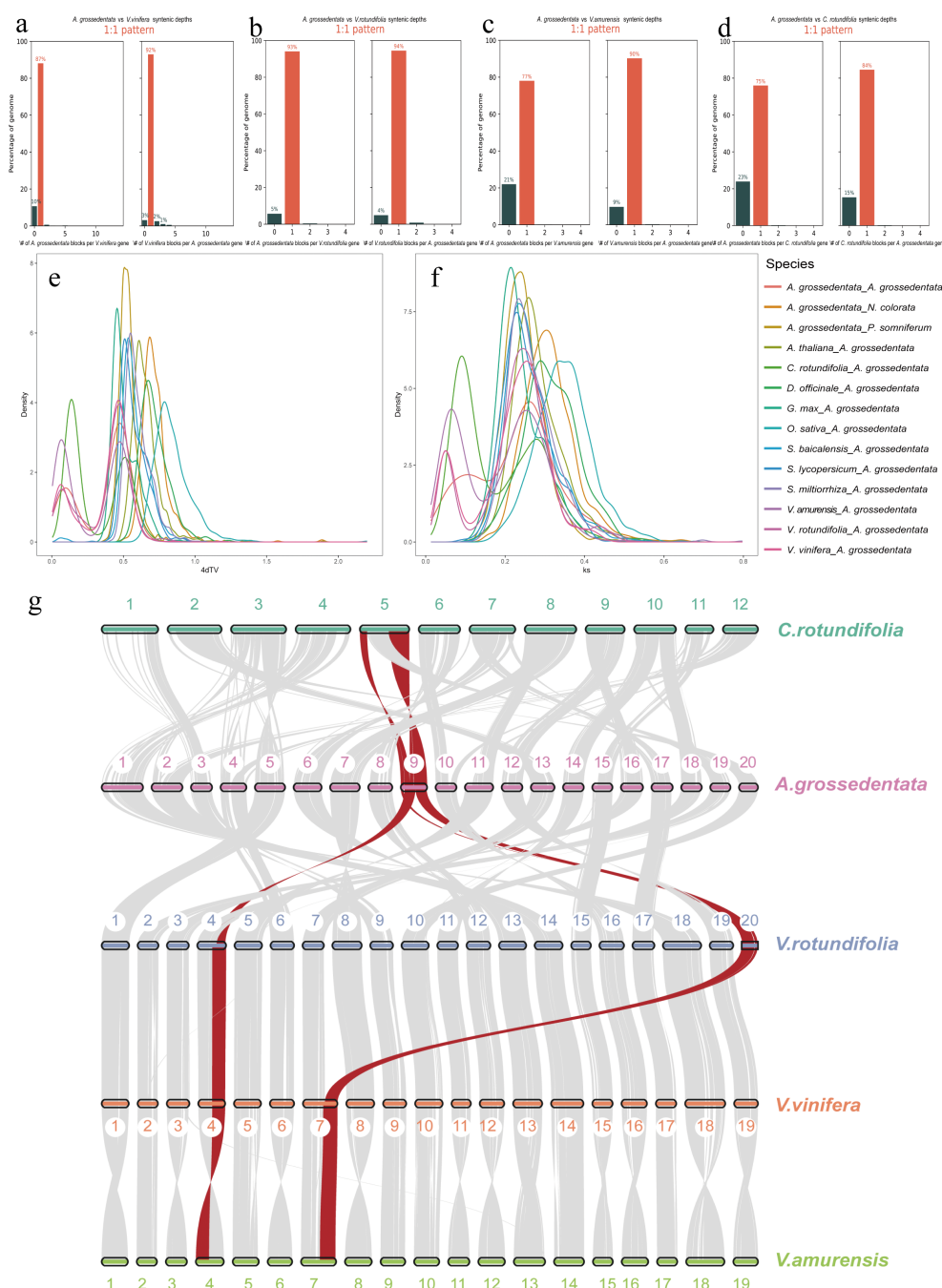


FIGURE 3

Comparative genomic and evolutionary analysis of *A. grossedentata*. (a–d) Syntenic depth between the representative Vitaceae species (*C. rotundifolia*, *V. vinifera*, *V. rotundifolia*, and *V. amurensis*) and *A. grossedentata*. (e, f) Distribution of 4dtv and Ks in 14 species. The dotted lines indicate the peaks of the *A. grossedentata* 4dtv and Ks. 4dtv, four-fold synonymous third-codon transversion position; Ks, synonymous substitution. (g) Synteny between *C. rotundifolia*, *V. rotundifolia*, *V. vinifera*, *V. amurensis* and *A. grossedentata*. The gray lines represent collinear blocks among species, while the red lines denote chromosomal fragment translocations between *A. grossedentata* and *C. rotundifolia*, *V. rotundifolia*, *V. vinifera*, and *V. amurensis*.

abundance expression patterns in roots, while *AgPAL3* and *AgFLS3* genes showed significant tissue-specific overexpression in stems (Supplementary Figure S12). These tissue-specific expression patterns were subsequently validated through qRT-PCR, with experimental data demonstrating high consistency with transcriptomic analysis results (Supplementary Figure S13). Then, inter-group correlation analysis was

performed on the 10 genes and 364 TFs, and 240 TFs were screened using correlation coefficients  $\geq 0.95$  or  $\leq -0.95$  (Supplementary Table S18). The 10 genes and 240 TFs were used to calculate degree values using Cytoscape, and 21 TFs with degree values  $\geq 5$  were selected (Figure 4e; Supplementary Table S19). *A. grossedentata* is the plant with the highest flavonoid content, and the variation trend of



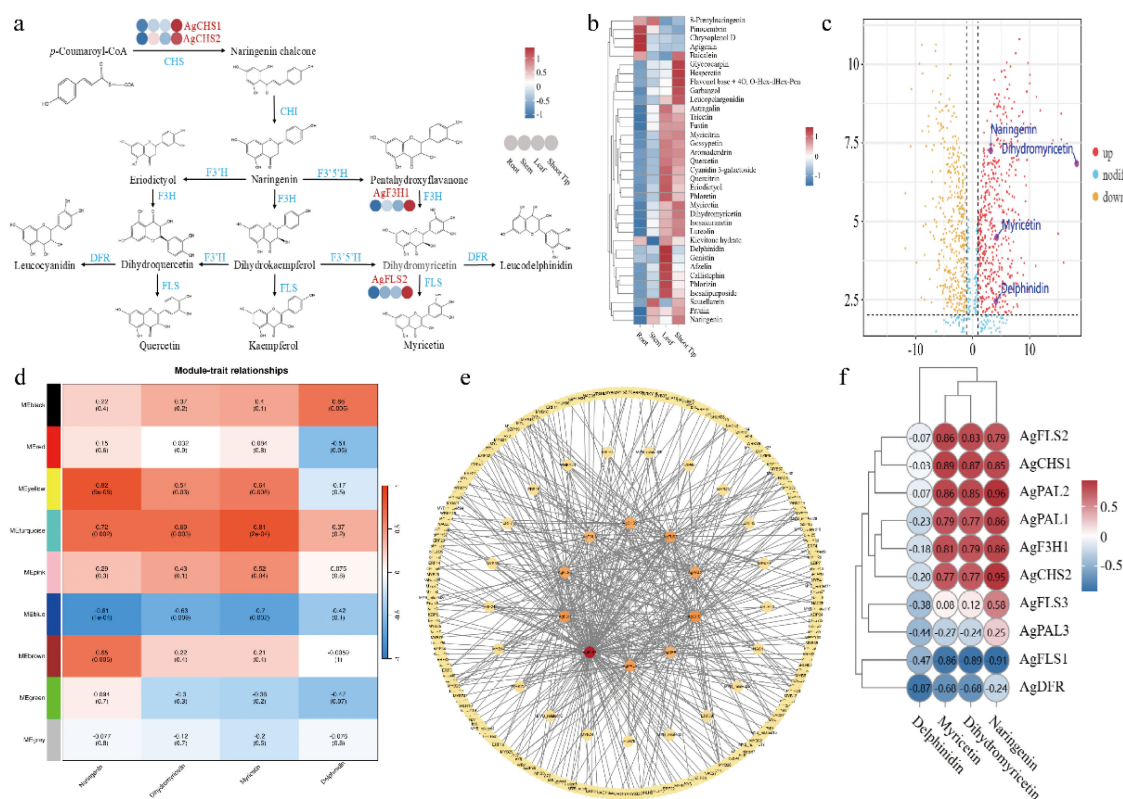


FIGURE 4

Metabolite levels, gene expression clustering, and flavonoid biosynthesis pathway. **(a)** Simplified flavonoid biosynthetic pathways in *A. grossedentata*. Enzymes in blue typeface; gene expression in rounded boxes. **(b)** Heatmap showing expression of 35 metabolites in the flavonoid synthesis pathway across different tissues at identical developmental stages. Normalized (z-score) FPKM values color-coded by expression level. **(c)** Volcano plot of differential metabolites in roots and buds. x-axis: Log2 fold change; y-axis: -log10 P-value. Dot size represents VIP value; red: upregulated, yellow: downregulated, blue: non-significant. Labeled metabolites: naringenin, dihydromyricetin, myricetin, delphinidin. **(d)** A heatmap displaying the correlations between gene expression modules and flavonoid metabolites shows that rows represent distinct modules (ME), and columns denote flavonoid compounds. **(e)** Interaction networks of 10 key hub genes and 240 transcription factors in brown, yellow, blue, and turquoise modules mapped using Cytoscape. **(f)** Heat maps showing correlation between naringenin, dihydromyricetin, myricetin, and delphinidin with 10 key genes identified by WGCNA.

dihydromyricetin basically reflects the changes in total flavonoid content. The inter-group correlation heatmap of 10 key genes with naringenin, dihydromyricetin, myricetin, and delphinidin showed that *AgF3H1*, *AgFLS2*, *AgCHS1*, *AgCHS2*, *AgPAL1*, and *AgPAL2* had the highest correlation with naringenin, dihydromyricetin, and myricetin (Figure 4f). In *A. grossedentata*, dihydromyricetin is synthesized by the *AgF3'5'H* enzyme gene (catalyzing dihydroquercetin or dihydrokaempferol) and the *AgF3H* enzyme gene (catalyzing pentahydroxyflavanone) (Figure 4a). Through WGCNA analysis, a *AgF3H1* gene was screened out, so we speculate that *AgF3H1* is the key gene for synthesizing dihydromyricetin.

### 3.8 Molecular docking of *AgF3H* genes

The *F3H* enzyme belongs to the 2-oxoglutarate-dependent dioxygenase (2-ODD) family, featuring a conserved N-terminal region and a C-terminal similar to the 4-hydroxylase alpha subunit (Cheng et al., 2014; Kawai et al., 2014). Within the flavonoid biosynthesis pathway of *A. grossedentata*, the *AgF3H* gene catalyzes

the conversion of eriodictyol, naringenin, and pentahydroxyflavanone into dihydroquercetin, dihydrokaempferol, and dihydromyricetin, respectively (Figure 4a). Analysis of the genome sequence of *A. grossedentata* revealed two *F3H* genes (*AgF3H1* and *AgF3H2*). Subsequently, molecular docking was performed using the amino acid sequences of *AgF3H1* and *AgF3H2* with eriodictyol, naringenin, and pentahydroxyflavanone. The XP docking and MM-GBSA analysis showed that eriodictyol and pentahydroxyflavanone had docking scores of -8.032 and -6.121 with *AgF3H1*, and MM-GBSA binding free energies of -32.96 and -41.49 kcal/mol, respectively, indicating stable binding (Supplementary Table S20). Eriodictyol deeply penetrates the *AgF3H1* active pocket, forming hydrophobic interactions with ALA123 and VAL124, and hydrogen bonds with VAL124, GLU122, ARG21, and ASP330 (Figure 5a). Pentahydroxyflavanone also penetrates the *AgF3H1* active pocket, forming hydrophobic interactions with ALA334 and LEU333, and hydrogen bonds with ASP330 and GLN276, along with a  $\pi$ -Cation bond with LYS215 (Figure 5b). For *AgF3H2*, pentahydroxyflavanone and naringenin had docking scores of -6.747 and -6.025, and MM-GBSA binding free energies of -35.60 and -30.49 kcal/mol, respectively, indicating stable

binding. Naringenin deeply penetrates the *AgF3H2* active pocket, forming hydrophobic interactions with PHE319, TYR323, PRO220, and LEU214, and hydrogen bonds with LYS196, SER117, and ARG128 (Figure 5c). Pentahydroxyflavanone penetrates the *AgF3H2* active pocket, forming hydrophobic interactions with PRO220 and TYR323, hydrogen bonds with TYR323 and LYS215, and additional hydrogen bonds with ASP330 and ARG128, as well as a  $\pi$ -Cation bond and a  $\pi$ - $\pi$  bond with HIP217 (Figure 5d). In summary, the *AgF3H1* and *AgF3H2* genes in *A. grossedentata* are more likely to bind to pentahydroxyflavanone, and after binding, *AgF3H1* has higher stability with pentahydroxyflavanone than *AgF3H2*. Therefore, combined with the WGCNA results, we speculate that the *AgF3H1* gene is the key enzyme gene catalyzing the synthesis of dihydromyricetin from pentahydroxyflavanone. *AgF3H1* is located on chromosome 9, and interspecies collinearity analysis shows that chromosome 9 of *A. grossedentata* underwent gene rearrangement during evolution (Figure 3c). Local collinearity results show that the *AgF3H1* gene of *A. grossedentata* corresponds to the *F3H* genes on chromosome 4 of *Vitis* species and chromosome 5 of *C. rotundifolia* (Figure 5e), which is consistent with the interspecies collinearity results (Figure 3c). Through real-time fluorescence quantitative experiments, it was found that the expression level of the *AgF3H1* gene in different tissues is consistent with the transcriptome data (Figure 5f). In conclusion, we speculate that *AgF3H1* is the key gene for dihydromyricetin biosynthesis in *A. grossedentata*.

## 4 Discussion

The exclusive Chinese species, *Ampelopsis grossedentata*, is characterized by young stems and leaves rich in flavonoids, which are commonly used in functional health beverages and folk medicines. In this research, we utilized Illumina and PacBio sequencing in conjunction with high-throughput Hi-C technology to assemble a near-complete reference genome for *A. grossedentata*. The scaffold N50 is 30.45 Mb, and the contig N50 is 21.93 Mb, significantly surpassing Vitaceae species such as *Tetrastigma hemsleyanum* (contig N50: 2.15 Mb, scaffold N50: 86 Mb) (Zhu et al., 2023) and the medicinal plant *Neolamarckia cadamba* (contig N50: 0.82 Mb, scaffold N50: 29.20 Mb) (Zhao et al., 2022). The revised genome size of *A. grossedentata* is 555.42 Mb, anchored to 20 pseudochromosomes, marking the first near-complete reference genome of *Ampelopsis*. Its genome size is comparable to other Vitaceae species such as *V. amurensis* Rupr. (approximately 522.28 Mb) (Wang et al., 2024), *V. amurensis* (604.56 Mb) (Wang et al., 2021), *V. vinifera* (494.87 Mb) (Shi et al., 2023) and *V. rotundifolia* (413.91 Mb) (Huff et al., 2023). The *A. grossedentata* genome exhibits high heterozygosity (1.48%) and a large number of repetitive sequences (62.62%), which are higher than those in *Vitis* species such as *V. zhejiang-adstricta* (heterozygosity: 0.845%; repetitive sequences: 47.49%) (Li H. Y. et al., 2024), *V. amurensis* Rupr. (heterozygosity: 1.20%; repetitive sequences: 59.21%) (Wang et al., 2024), and *C. rotundifolia* (heterozygosity: 1.19%; repetitive sequences: 47.41%) (Xin et al., 2022). The short-read coverage reaches 99.29%, and the BUSCO evaluation shows that 98.80% of

the genome is completely assembled, outperforming Vitaceae species *V. amurensis* Rupr. (BUSCO: 97.50%) (Wang et al., 2024), *V. amurensis* (reads: 98.58%; BUSCO: 94.60%) (Wang et al., 2021) and *V. vinifera* (BUSCO: 98.50%) (Shi et al., 2023). These findings indicate *A. grossedentata* genome is superior to other Vitaceae plants in sequence consistency, assembly accuracy and completeness, which provides a solid foundation for phylogenetic, gene function and molecular breeding research.

Throughout plant evolution, many species have experienced one or more ancient genome polyploidization events (Blanc and Wolfe, 2004; Jiao et al., 2011; Soltis and Soltis, 2016). Whole-genome duplication (WGD) has significantly contributed to plant speciation and the development of valuable traits (Rensing, 2014; Song et al., 2024). Due to their close relationship with the common ancestor of angiosperms, *Vitis* species and even Vitaceae plants are widely used in evolutionary analyses (Xin et al., 2022). Phylogenetic studies reveal that the divergence between *Cissus* and *Vitis* occurred approximately 49.0 million years ago (range: 31.4–69.1 million years ago), confirming previous estimates based on whole-genome data for the divergence of *C. rotundifolia* (68.41 million years ago, range: 44.1 to 89.8 million years ago) and *C. quadrangularis* (range: 60.19 to 84.68 million years ago) from *Vitis* (Xin et al., 2022; Li Q. Y. et al., 2024). However, there is also evidence suggesting that the divergence between *Cissus* and *Vitis* may have occurred around 38.07 million years ago (range: 21.38 to 67.28 million years ago) (Li et al., 2024). Therefore, the high-quality genome data of *A. grossedentata* provides strong support for resolving the evolutionary relationships and developmental positions among Vitaceae species. Ks and 4dtv show two peaks, indicating that *A. grossedentata* experienced the WGT- $\gamma$  event common to angiosperms and a Vitaceae-specific whole-genome duplication event during its evolution (Jaillon et al., 2007; Tang et al., 2008). WGD events not only double the genome size but also facilitate the acquisition and loss of gene copies (Van de Peer et al., 2009). Previous studies have identified useful genes from WGDs associated with plant growth and metabolic pathways (Chae et al., 2014; Chakraborty, 2018). Collinearity analysis shows that the chromosome collinearity pattern among Vitaceae species is chaotic, and chromosome 9 of *A. grossedentata* has undergone chromosome reorganization and/or gene fragment rearrangement events during evolution. *A. grossedentata* expanded/contracted/lost a large number of genes in two WGD events, resulting in ancestral genes being scattered across multiple chromosomes. Gene family expansion has been recognized as a key driving factor in the formation of various plant species adapting to natural variations, and these expanded gene families increase plant adaptability to biotic and abiotic stresses (Renny-Byfield and Wendel, 2014). There are 1,114 expanded gene families in *A. grossedentata*, enriched in “secondary metabolite biosynthetic process” and “secondary metabolic process” pathways. In conclusion, we propose that after experiencing two WGD events and gene recombination events on chromosome 9, *A. grossedentata* accumulated key genes for synthesizing flavonoid compounds.

The mining of genetic resources and the screening of candidate genes for key traits enable researchers to identify crucial genetic variations and environmental adaptability. Through the integration

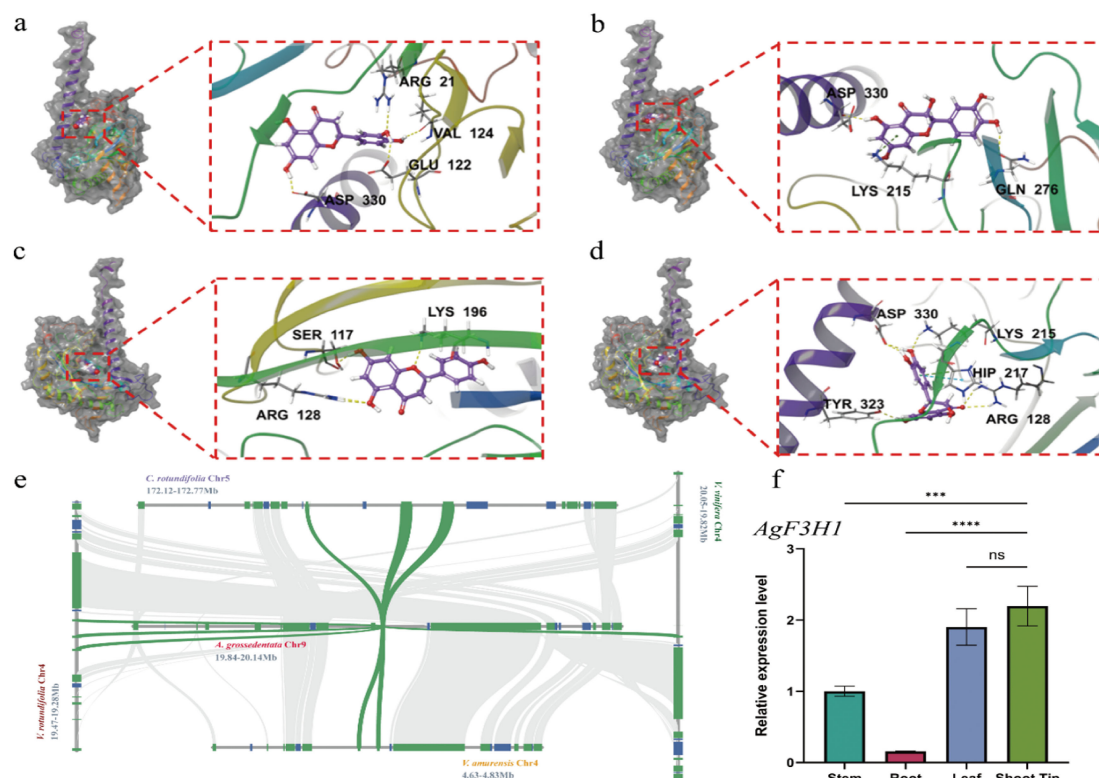


FIGURE 5

Bioinformatics analysis of *AgF3H1* and *AgF3H2* genes in *A. grossedentata*. (a) Molecular docking of *AgF3H1* with Eriodictyol. (b) Molecular docking of *AgF3H1* with pentahydroxyflavanone. (c) Molecular docking of *AgF3H2* with naringenin. (d) Molecular docking of *AgF3H2* with pentahydroxyflavanone. Yellow: hydrogen bonds; green:  $\pi$ -cation bonds; blue:  $\pi$ - $\pi$  bonds. (e) Local collinearity analysis of chromosome 9 in *A. grossedentata* with corresponding chromosomes in *V. vinifera*, *V. amurensis*, *V. rotundifolia*, *C. rotundifolia* and *C. sinensis*. Green: homology of *AgF3H1*. (f) Quantitative RT-PCR analysis of *AgF3H1* expression across various tissues of *A. grossedentata*. Data are presented as mean  $\pm$  SD from three independent replicates. Statistical significance was assessed using Student's t-test: \*\*\* $P$  < 0.001, \*\*\*\* $P$  < 0.0001; n.s. indicates non-significant differences.

of gene co-expression and flavonoid metabolomics analysis, we delineated the flavonoid biosynthetic pathway and its regulatory network in *A. grossedentata*. Dihydromyricetin is the most abundant flavonoid monomer compound in *A. grossedentata*, and its variation trend generally reflects the changes in total flavonoid content. In *A. grossedentata*, *AgF3H* and *AgF3'5'H* are key catalytic genes for synthesizing dihydromyricetin, which can catalyze pentahydroxyflavanone, dihydroquercetin, and dihydrokaempferol into dihydromyricetin. By analyzing the correlation between genes and metabolites using the WGCNA package, 10 key genes highly interconnected with flavonoid compound synthesis were screened in yellow, brown, blue, and turquoise modules. These genes include *AgPAL3*, *AgPAL2*, *AgPAL1*, *AgFLS3*, *AgFLS2*, *AgFLS1*, *AgF3H1*, *AgDFR*, *AgCHS2*, and *AgCHS1*. Transcriptome and qRT-PCR experiments showed that *AgF3H1* had the highest expression in shoot tips, followed by decreasing expression in leaves, stems, and roots, suggesting that high expression of the *AgF3H1* gene may lead to increased dihydromyricetin content in *A. grossedentata*, consistent with previous research results. Based on the near-complete genome sequence, we identified *AgF3H1* and *AgF3H2* in the *A. grossedentata* genome. Molecular docking showed that the *AgF3H* genes has a

higher binding affinity with pentahydroxyflavanone, but *AgF3H1* (-41.49 kcal/mol) is more stable than *AgF3H2* (-30.49 kcal/mol) when bound to pentahydroxyflavanone. High expression of *F3H* can increase the flavonoid content in plant tissues. Studies have found that the *S. lycopersicum* *SIF3H* gene was transferred into *Nicotiana tabacum*, and the results showed that the flavonoid content in *N. tabacum* overexpressing the *SIF3H* gene was about 30% higher than in the wild type (Meng et al., 2015). The *CsF3Hs* gene from the tea plant was transferred into *A. thaliana*, and it was found that the content of most flavonol glycosides and oligomeric proanthocyanidins in the seeds significantly increased (20-40%), indicating that *CsF3Hs* plays a key role in flavonoid biosynthesis in *C. sinensis* (Han et al., 2017). Totally, the regulation of the *F3H* gene significantly impacts flavonoid metabolism and synthesis, underscoring the pivotal role of *AgF3H* in the biosynthesis of flavonoid compounds and derivatives in *A. grossedentata*. Therefore, combined with the WGCNA results, we speculate that the *AgF3H1* gene is a key enzyme gene catalyzing the synthesis of dihydromyricetin from pentahydroxyflavanone. In subsequent research phases, comprehensive functional validation of the *AgF3H1* gene will be performed through both homologous and heterologous



systems. Systematic investigations will include overexpression analysis in native host species alongside heterologous expression in model organisms, complemented by targeted gene knockout experiments using CRISPR/Cas9-mediated genome editing. These multi-platform approaches will elucidate the gene's regulatory mechanisms in flavonoid biosynthesis pathways and its pleiotropic effects on plant physiological processes. The resulting data will establish a theoretical foundation for molecular breeding programs aimed at enhancing phytochemical profiles in crops, while concurrently providing technical parameters for developing nutraceutical-enriched agricultural products through metabolic engineering strategies.

In summary, we present the first near-complete genome assembly of *A. grossedentata*, providing comprehensive genomic data crucial for in-depth studies of this species and other medicinal and edible plants. Comparative genomic evolutionary analysis sheds light on the evolutionary trajectory of Vitaceae. The discovery of candidate genes involved in flavonoid biosynthesis paves the way for future genetic enhancement of *A. grossedentata*.

## Data availability statement

The genome assembly and raw sequencing data for *Ampelopsis grossedentata* have been submitted to NCBI under project ID PRJNA1117225.

## Author contributions

ZY: Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. ZF: Formal Analysis, Methodology, Software, Writing – original draft. FW: Methodology, Software, Writing – review & editing. PZ: Formal Analysis, Methodology, Software, Writing – original draft. QW: Data curation, Validation, Writing – original draft. BA: Formal Analysis, Supervision, Validation, Writing – original draft. YW: Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. ML: Investigation, Supervision, Validation, Writing – review & editing.

## References

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. doi: 10.1038/s41586-024-07487-w
- Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106. doi: 10.1186/gb-2010-11-10-r106
- 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
- Blanc, G., and Wolfe, K. H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16, 1667–1678. doi: 10.1105/tpc.021345
- Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–d354. doi: 10.1093/nar/gkaa977
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31, 365–370. doi: 10.1093/nar/gkg095
- Brown, M. R., Manuel Gonzalez de La Rosa, P., and Blaxter, M. (2025). Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*. 41, btaf049. doi: 10.1093/bioinformatics/btaf049
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Natural Science Foundation of China (32171842), National Key Research and Development Project of China (2019YFD1100403), and Hunan Graduate Research Innovation Project (Key project, Grant No. CX20240068). We appreciate the sequencing and bioinformatics support provided by Glbizzia Biosciences Co., Ltd (Beijing).

## Conflict of interest

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

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## 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.2025.1580779/full#supplementary-material>

- Cao, L., Deng, W., Lin, Y. F., Zhu, X., Xu, X., Zhang, Z. B., et al. (2023). *Ampelopsis grossedentata* represents a new host of the 16SrI group of phytoplasma associated with yellow leaf symptoms in China. *Plant Dis.* 108, 780. doi: 10.1094/pdis-09-23-1820-pdn
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Carneiro, R. C., Wang, H. J., Duncan, S. E., and O'Keefe, S. F. (2020). Flavor compounds in vine tea (*Ampelopsis grossedentata*) infusions. *Food Sci. Nutr.* 8, 4505–4511. doi: 10.1002/fsn.3.1754
- Carneiro, R. C., Ye, L., Baek, N., Teixeira, G. H., and O'Keefe, S. F. (2021). Vine tea (*Ampelopsis grossedentata*): A review of chemical composition, functional properties, and potential food applications. *J. Funct. Foods* 76, 104317. doi: 10.1016/j.jff.2020.104317
- Chae, L., Kim, T., Nilo-Poyanco, R., and Rhee, S. Y. (2014). Genomic signatures of specialized metabolism in plants. *Science* 344, 510–513. doi: 10.1126/science.1252076
- Chakraborty, P. (2018). Herbal genomics as tools for dissecting new metabolic pathways of unexplored medicinal plants and drug discovery. *Biochim. Open* 6, 9–16. doi: 10.1016/j.biopen.2017.12.003
- Chen, J., Wu, Y., Zou, J., and Gao, K. (2016).  $\alpha$ -Glucosidase inhibition and antihyperglycemic activity of flavonoids from *Ampelopsis grossedentata* and the flavonoid derivatives. *Bioorg. Med. Chem.* 24, 1488–1494. doi: 10.1016/j.bmc.2016.02.018
- Chen, N. (2004). Using Repeat Masker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* 5, 4.10. 11–14.10. 14. doi: 10.1002/0471250953.bi0410s05
- 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
- Chen, S. F., Zhou, Y. Q., Chen, Y. R., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W., 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
- Cheng, A. X., Han, X. J., Wu, Y. F., and Lou, H. X. (2014). The function and catalysis of 2-oxoglutarate-dependent oxygenases involved in plant flavonoid biosynthesis. *Int. J. Mol. Sci.* 15, 1080–1095. doi: 10.3390/ijms15011080
- Chin, C. H., Chen, S. H., Wu, H. H., Ho, C. W., Ko, M. T., and Lin, C. Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* 8, 11. doi: 10.1186/1752-0509-8-s4-s11
- Davies, K. M. (1993). A cDNA clone for flavanone 3-hydroxylase from *Malus*. *Plant Physiol.* 103, 291. doi: 10.1104/pp.103.1.291
- Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23, 673–679. doi: 10.1093/bioinformatics/btm009
- Deng, Y. Y., Li, J. Q., Wu, S. F., Zhu, Y. P., Chen, Y. W., and He, F. C. (2006). Integrated nr database in protein annotation system and its localization. *Comput. Eng.* 32, 71–72.
- 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., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016a). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3, 99–101. doi: 10.1016/j.cels.2015.07.012
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016b). 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
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinform.* 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
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Gu, Z. G., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circline implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812. doi: 10.1093/bioinformatics/btu393
- Gu, L., Zhang, N., Feng, C., Yi, Y., and Yu, Z. W. (2020). The complete chloroplast genome of *Ampelopsis grossedentata* (Hand.-Mazz.) W. T. Wang (family: Vitaceae) and its phylogenetic analysis. *Mitochondrial DNA B. Resour.* 5, 2423–2424. doi: 10.1080/23802359.2020.1775508
- Guan, D. F., McCarthy, S. A., Wood, J., Howe, K., Wang, Y. D., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36, 2896–2898. doi: 10.1093/bioinformatics/btaa025
- Guo, Z., Guozhang, H., Wang, H., Li, Z., and Liu, N. (2019). Ampelopsin inhibits human glioma through inducing apoptosis and autophagy dependent on ROS generation and JNK pathway. *BioMed. Pharmacother.* 116, 108524. doi: 10.1016/j.biopha.2018.12.136
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., 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
- Han, Y. H., Huang, K. Y., Liu, Y. J., Jiao, T. M., Ma, G. L., Qian, Y. M., et al. (2017). Functional analysis of two flavanone-3-hydroxylase genes from *Camellia sinensis*: a critical role in flavonoid accumulation. *Genes* 8, 300. doi: 10.3390/genes8110300
- Han, M. V., Thomas, G. W., Lugo-Martínez, J., and Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* 30, 1987–1997. doi: 10.1093/molbev/mst100
- Han, Y., and Wessler, S. R. (2010). MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* 38, e199. doi: 10.1093/nar/gkq862
- Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* 12, 491. doi: 10.1186/1471-2105-12-491
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass. Spectrom.* 45, 703–714. doi: 10.1002/jms.1777
- Hou, X. L., Tong, Q., Wang, W. Q., Shi, C. Y., Xiong, W., Chen, J., et al. (2015). Suppression of inflammatory responses by dihydromyricetin, a flavonoid from *Ampelopsis grossedentata*, via inhibiting the activation of NF- $\kappa$ B and MAPK signaling pathways. *J. Nat. Prod.* 78, 1689–1696. doi: 10.1021/acs.jnatprod.5b00275
- Hu, H., Luo, F., Wang, M., Fu, Z., and Shu, X. (2020). New method for extracting and purifying dihydromyricetin from *Ampelopsis grossedentata*. *ACS Omega* 5, 13955–13962. doi: 10.1021/acsomega.0c01222
- Huang, H. C., Liao, C. C., Peng, C. C., Lim, J. M., Siao, J. H., Wei, C. M., et al. (2016). Dihydromyricetin from *Ampelopsis grossedentata* inhibits melanogenesis through down-regulation of MAPK, PKA and PKC signaling pathways. *Chem. Biol. Interact.* 258, 166–174. doi: 10.1016/j.cbi.2016.08.023
- Huff, M., Hulse-Kemp, A. M., Scheffler, B. E., Youngblood, R. C., Simpson, S. A., Babiker, E., et al. (2023). Long-read, chromosome-scale assembly of *Vitis rotundifolia* cv. Carlos and its unique resistance to *Xylella fastidiosa* subsp. *fastidiosa*. *BMC Genom.* 24, 409. doi: 10.1186/s12864-023-09514-y
- Huson, D. H., Beier, S., Flade, I., Górski, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* 12, e1004957. doi: 10.1371/journal.pcbi.1004957
- Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148
- Jain, C., Rhie, A., Hansen, N. F., Koren, S., and Phillippy, A. M. (2022). Long-read mapping to repetitive reference sequences using Winnowmap2. *Nat. Methods* 19, 705–710. doi: 10.1038/s41592-022-01457-8
- Jain, C., Rhie, A., Zhang, H. W., Chu, C., Walenz, B. P., Koren, S., et al. (2020). Weighted minimizer sampling improves long read mapping. *Bioinformatics* 36, i111–i118. doi: 10.1093/bioinformatics/btaa435
- Jiao, Y. N., 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
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kawai, Y., Ono, E., and Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *Plant J.* 78, 328–343. doi: 10.1111/tpj.12479
- 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
- Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 278. doi: 10.1186/s13059-019-1910-1
- Kumar, S., Suleski, M., Craig, J. M., Kaspric, A. E., Sanderford, M., Li, M., et al. (2022). TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* 39 (8), msac174. doi: 10.1093/molbev/msac174
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5, R12. doi: 10.1186/gb-2004-5-2-r12
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. (2007). RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Li, H. (2016). Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* 32, 2103–2110. doi: 10.1093/bioinformatics/btw152
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

- Li, X. H., Cao, M. H., Ma, W. B., Jia, C. H., Li, J. H., Zhang, M. X., et al. (2020). Annotation of genes involved in high level of dihydromyricetin production in vine tea (*Ampelopsis grossedentata*) by transcriptome analysis. *BMC Plant Biol.* 20, 1–12. doi: 10.1186/s12870-020-2324-7
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, Y., Hu, H., Yang, H., Lin, A., Xia, H., Cheng, X., et al. (2022). Vine tea (*Ampelopsis grossedentata*) extract attenuates CCl<sub>4</sub>-induced liver injury by restoring gut microbiota dysbiosis in mice. *Mol. Nutr. Food Res.* 66, e2100892. doi: 10.1002/mnfr.202100892
- Li, H. Y., Liu, Y. B., Fan, P. G., Dai, Z. W., Hao, J. C., Duan, W., et al. (2024). The Genome of *Vitis zhejiang-adstricta* strengthens the protection and utilization of the endangered ancient grape endemic to China. *Plant Cell Physiol.* 65, 216–227. doi: 10.1093/pcp/pcad140
- Li, Q. Y., Wang, Y., Zhou, H. M., Liu, Y. S., Gichuki, D. K., Hou, Y. J., et al. (2024). The *Cissus quadrangularis* genome reveals its adaptive features in an arid habitat. *Hortic. Res.* 11, uhac038. doi: 10.1093/hr/uhac038
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., et al. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. *arXiv preprint arXiv*. doi: 10.1016/S0925-4005(96)02015-1
- Liu, S., Zhang, Q., Kollie, L., Dong, J., and Liang, Z. (2023). Molecular networks of secondary metabolism accumulation in plants: current understanding and future challenges. *Ind. Crops Prod.* 201, 116901. doi: 10.1016/j.indcrop.2023.116901
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>−</sup>ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Luo, Q. J., Zhou, W. C., Liu, X. Y., Li, Y. J., Xie, Q. L., Wang, B., et al. (2023). Chemical constituents and α-glucosidase inhibitory, antioxidant and hepatoprotective activities of *Ampelopsis grossedentata*. *Molecules* 28 (24), 7956. doi: 10.3390/molecules28247956
- Ma, J. Q., Sun, Y. Z., Ming, Q. L., Tian, Z. K., Yang, H. X., and Liu, C. M. (2019). Ampelopsin attenuates carbon tetrachloride-induced mouse liver fibrosis and hepatic stellate cell activation associated with the SIRT1/TGF-β1/Smad3 and autophagy pathway. *Int. Immunopharmacol.* 77, 105984. doi: 10.1016/j.intimp.2019.105984
- 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
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27, 764–770. doi: 10.1093/bioinformatics/btr011
- McGinnis, S., and Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 32, W20–W25. doi: 10.1093/nar/gkh435
- Meng, C., Zhang, S., Deng, Y. S., Wang, G. D., and Kong, F. Y. (2015). Overexpression of a tomato flavanone 3-hydroxylase-like protein gene improves chilling tolerance in tobacco. *Plant Physiol. Biochem.* 96, 388–400. doi: 10.1016/j.plaphy.2015.08.019
- Mi, H. Y., Muruganujan, A., Ebert, D., Huang, X. S., and Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47, D419–D426. doi: 10.1093/nar/gky1038
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Navarro-Reig, M., Jaumot, J., García-Reiriz, A., and Tauler, R. (2015). Evaluation of changes induced in rice metabolome by Cd and Cu exposure using LC-MS with XCMS and MCR-ALS data analysis strategies. *Anal. Bioanal. Chem.* 407, 8835–8847. doi: 10.1007/s00216-015-9042-2
- 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
- Ni, J. B., Zhao, Y., Tao, R. Y., Yin, L., Gao, L., Strid, Å., et al. (2020). Ethylene mediates the branching of the jasmonate-induced flavonoid biosynthesis pathway by suppressing anthocyanin biosynthesis in red Chinese pear fruits. *Plant Biotechnol. J.* 18, 1223–1240. doi: 10.1111/pbi.13287
- Ou, S. J., 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. J., and Jiang, N. (2019). LTR\_FINDER\_parallel: parallelization of LTR\_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob. DNA* 10, 48. doi: 10.1186/s13100-019-0193-0
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071
- Pelletier, M. K., and Shirley, B. W. (1996). Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings. Coordinate regulation with chalcone synthase and chalcone isomerase. *Plant Physiol.* 111, 339–345. doi: 10.1104/pp.111.1.339
- Prescott, A. G., and John, P. (1996). Dioxygenases: molecular structure and role in plant metabolism. *Annu. Rev. Plant Biol.* 47, 245–271. doi: 10.1146/annurev.arplant.47.1.245
- Ramani, V., Deng, X. X., Qiu, R. L., Lee, C., Distech, C. M., Noble, W. S., et al. (2020). Sci-Hi-C: a single-cell Hi-C method for mapping 3D genome organization in large number of single cells. *Methods* 170, 61–68. doi: 10.1016/j.meth.2019.09.012
- Ran, L., Wang, X., Lang, H., Xu, J., Wang, J., Liu, H., et al. (2019). *Ampelopsis grossedentata* supplementation effectively ameliorates the glycemic control in patients with type 2 diabetes mellitus. *Eur. J. Clin. Nutr.* 73, 776–782. doi: 10.1038/s41430-018-0282-z
- Rasmussen, J. A., Villumsen, K. R., Ernst, M., Hansen, M., Forberg, T., Gopalakrishnan, S., et al. (2022). A multi-omics approach unravels metagenomic and metabolic alterations of a probiotic and synbiotic additive in rainbow trout (*Oncorhynchus mykiss*). *Microbiome* 10, 21. doi: 10.1186/s40168-021-01221-8
- 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
- Renny-Byfield, S., and Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. *Am. J. Bot.* 101, 1711–1725. doi: 10.3732/ajb.1400119
- Rensing, S. A. (2014). Gene duplication as a driver of plant morphogenetic evolution. *Curr. Opin. Plant Biol.* 17, 43–48. doi: 10.1016/j.pbi.2013.11.002
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Shen, N., Wang, T., Gan, Q., Liu, S., Wang, L., and Jin, B. (2022). Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* 383, 132531. doi: 10.1016/j.foodchem.2022.132531
- Shi, X. Y., Cao, S., Wang, X., Huang, S. Y., Wang, Y., Liu, Z. J., et al. (2023). The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. *Hortic. Res.* 10, uhad061. doi: 10.1093/hr/uhad061
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Slater, G. S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* 6, 31. doi: 10.1186/1471-2105-6-31
- 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.pbi.2016.03.015
- Song, B. X., Buckler, E. S., and Stitzer, M. C. (2024). New whole-genome alignment tools are needed for tapping into plant diversity. *Trends Plant Sci.* 29, 355–369. doi: 10.1016/j.tplants.2023.08.013
- Sparvoli, F., Martin, C., Scienza, A., Gavazzi, G., and Tonelli, C. (1994). Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). *Plant Mol. Biol.* 24, 743–755. doi: 10.1007/bf00029856
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24, 637–644. doi: 10.1093/bioinformatics/btn013
- Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., et al. (2007). LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* 35, D527–D532. doi: 10.1093/nar/gkl838
- Sun, P. C., Jiao, B. B., Yang, Y. Z., Shan, L. X., Li, T., Li, X. N., et al. (2022). WGDl: 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
- Tang, H. B., Bowers, J. E., Wang, X. Y., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917
- Tang, H. B., Krishnakumar, V., Zeng, X. F., Xu, Z. G., Taranto, A., Lomas, J. S., et al. (2024). JCVI: A versatile toolkit for comparative genomics analysis. *iMeta* 3, e211. doi: 10.1002/imt2.211
- Tang, S., and Riva, A. (2013). PASTA: splice junction identification from RNA-sequencing data. *BMC Bioinf.* 14, 116. doi: 10.1186/1471-2105-14-116
- Thévenot, E. A., Roux, A., Xu, Y., Ezan, E., and Junot, C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome Res.* 14, 3322–3335. doi: 10.1021/acs.jproteome.5b00354
- Van de Peer, Y., Maere, S., and Meyer, A. (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10, 725–732. doi: 10.1038/nrg2600
- Wang, P. F., Meng, F. B., Yang, Y. M., Ding, T. T., Liu, H. P., Wang, F. X., et al. (2024). *De novo* assembling a high-quality genome sequence of Amur grape (*Vitis amurensis* Rupr.) gives insight into *Vitis* divergence and sex determination. *Hortic. Res.* 11, uhac117. doi: 10.1093/hr/uhac117



- Wang, Y. P., Tang, H. B., DeBarry, J. D., Tan, X., Li, J. P., Wang, X. Y., et al. (2012). MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49–e49. doi: 10.1093/nar/gkr1293
- Wang, Y., Xin, H. P., Fan, P. G., Zhang, J. S., Liu, Y. B., Dong, Y., et al. (2021). The genome of *Shanputao* (*Vitis amurensis*) provides a new insight into cold tolerance of grapevine. *Plant J.* 105, 1495–1506. doi: 10.1111/tjp.15127
- Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., et al. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res* 4, 1310. doi: 10.12688/f1000research.7334.1
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., et al. (2007). HMDB: the human metabolome database. *Nucleic Acids Res.* 35, D521–D526. doi: 10.1093/nar/gkl193
- Wolff, J., Rabbani, L., Gilsbach, R., Richard, G., Manke, T., Backofen, R., et al. (2020). Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 48, W177–W184. doi: 10.1093/nar/gkaa220
- Wu, Y. P., Bai, J. R., Zhong, K., Huang, Y. N., and Gao, H. (2017). A dual antibacterial mechanism involved in membrane disruption and DNA binding of 2R,3R-dihydromyricetin from pine needles of *Cedrus deodara* against *Staphylococcus aureus*. *Food Chem.* 218, 463–470. doi: 10.1016/j.foodchem.2016.07.090
- Wu, T. Z., Hu, E. Q., Xu, S. B., Chen, M. J., Guo, P. F., Dai, Z. H., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2 (3), 11. doi: 10.1016/j.xinn.2021.100141
- Wu, R. R., Li, X., Cao, Y. H., Peng, X., Liu, G. F., Liu, Z. K., et al. (2023). China medicinal plants of the *Ampelopsis grossedentata*-A review of their botanical characteristics, use, phytochemistry, active pharmacological components, and toxicology. *Molecules* 28. doi: 10.3390/molecules28207145
- Xia, J. G., and Wishart, D. S. (2011). Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat. Protoc.* 6, 743–760. doi: 10.1038/nprot.2011.319
- Xie, K., He, X., Chen, K., Chen, J., Sakao, K., and Hou, D. X. (2019). Antioxidant properties of a traditional vine tea, *Ampelopsis grossedentata*. *Antioxid. (Basel)* 8 (8). doi: 10.3390/antiox8080295
- Xie, C., Mao, X. Z., Huang, J. J., Ding, Y., Wu, J. M., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, W316–W322. doi: 10.1093/nar/gkr483
- Xin, H. P., Wang, Y., Li, Q. Y., Wan, T., Hou, Y. J., Liu, Y. S., et al. (2022). A genome for *Cissus illustrata* features underlying its evolutionary success in dry savannas. *Hortic. Res.* 9, uhac208. doi: 10.1093/hr/uhac208
- Xiong, Y., Zhu, G. H., Zhang, Y. N., Hu, Q., Wang, H. N., Yu, H. N., et al. (2021). Flavonoids in *Ampelopsis grossedentata* as covalent inhibitors of SARS-CoV-2 3CL (pro): Inhibition potentials, covalent binding sites and inhibitory mechanisms. *Int. J. Biol. Macromol.* 187, 976–987. doi: 10.1016/j.ijbiomac.2021.07.167
- Xu, M. (2017). Screening and validation of reference genes for quantitative RT-PCR analysis in *Ampelopsis grossedentata*. *Chin. Traditional. Herbal. Drugs*, 48 (6), 1192–1198. doi: 10.7501/j.issn.0253-2670.2017.06.023
- Xu, D., Yang, J. B., Wen, H. M., Feng, W. L., Zhang, X. H., Hui, X. Q., et al. (2024). CentIER: Accurate centromere identification for plant genomes. *Plant Commun.* 5 (10). doi: 10.1016/j.xplc.2024.101046
- Yang, Z. H. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Ye, L., Wang, H., Duncan, S. E., Eigel, W. N., and O'Keefe, S. F. (2015). Antioxidant activities of vine tea (*Ampelopsis grossedentata*) extract and its major component dihydromyricetin in soybean oil and cooked ground beef. *Food Chem.* 172, 416–422. doi: 10.1016/j.foodchem.2014.09.090
- Yu, G. C., Wang, L. G., Han, Y. Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yu, Z. W., Zhang, N., Jiang, C. Y., Wu, S. X., Feng, X. Y., and Feng, X. Y. (2021). Exploring the genes involved in biosynthesis of dihydroquercetin and dihydromyricetin in *Ampelopsis grossedentata*. *Sci. Rep.* 11, 15596. doi: 10.1038/s41598-021-95071-x
- Zabala, G., and Vodkin, L. O. (2005). The wp mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. *Plant Cell* 17, 2619–2632. doi: 10.1105/tpc.105.033506
- Zeng, S. H., Liu, Y. L., Hu, W. M., Liu, Y. L., Shen, X. F., and Wang, Y. (2013). Integrated transcriptional and phytochemical analyses of the flavonoid biosynthesis pathway in *Epimedium*. *Plant Cell. Tissue Organ. Cult.* 115, 355–365. doi: 10.1007/s11240-013-0367-2
- Zeng, T., Song, Y., Qi, S., Zhang, R., Xu, L., and Xiao, P. (2023). A comprehensive review of vine tea: Origin, research on Materia Medica, phytochemistry and pharmacology. *J. Ethnopharmacol.* 317, 116788. doi: 10.1016/j.jep.2023.116788
- Zhang, J., Chen, Y., Luo, H., Sun, L., Xu, M., Yu, J., et al. (2018). Recent update on the pharmacological effects and mechanisms of dihydromyricetin. *Front. Pharmacol.* 9. doi: 10.3389/fphar.2018.01204
- Zhang, S., Gao, S., Chen, Y., Xu, S., Yu, S., and Zhou, J. (2022). Identification of hydroxylation enzymes and the metabolic analysis of dihydromyricetin synthesis in *Ampelopsis grossedentata*. *Genes (Basel)* 13 (12). doi: 10.3390/genes13122318
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, Article17. doi: 10.2202/1544-6115.1128
- Zhang, H., Xie, G., Tian, M., Pu, Q., and Qin, M. (2016). Optimization of the ultrasonic-assisted extraction of bioactive flavonoids from *Ampelopsis grossedentata* and subsequent separation and purification of two flavonoid aglycones by high-speed counter-current chromatography. *Molecules* 21 (8). doi: 10.3390/molecules21081096
- Zhang, X., Xu, Y., Xue, H., Jiang, G. C., and Liu, X. J. (2019). Antioxidant activity of vine tea (*Ampelopsis grossedentata*) extract on lipid and protein oxidation in cooked mixed pork patties during refrigerated storage. *Food Sci. Nutr.* 7, 1735–1745. doi: 10.1002/fsn3.1013
- Zhang, X. T., Zhang, S. C., Zhao, Q., Ming, R., and Tang, H. B. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* 5, 833–845. doi: 10.1038/s41477-019-0487-8
- Zhang, Q. L., Zhao, Y. F., Zhang, M. Y., Zhang, Y. L., Ji, H. F., and Shen, L. (2021). Recent advances in research on vine tea, a potential and functional herbal tea with dihydromyricetin and myricetin as major bioactive compounds. *J. Pharm. Anal.* 11, 555–563. doi: 10.1016/j.jpba.2020.10.002
- Zhao, X. L., Hu, X. D., OuYang, K. X., Yang, J., Que, Q. M., Long, J. M., et al. (2022). Chromosome-level assembly of the *Neolamarckia cadamba* genome provides insights into the evolution of cadambine biosynthesis. *Plant J.* 109, 891–908. doi: 10.1111/tjp.15600
- Zhou, S., Chen, Y., Guo, C., and Qi, J. (2020). PhyloMCL: Accurate clustering of hierarchical orthogroups guided by phylogenetic relationship and inference of polyploidy events. *Methods Ecol. Evol.* 11, 943–954. doi: 10.1111/2041-210X.13401
- Zhou, Y., Shu, F., Liang, X., Chang, H., Shi, L., Peng, X., et al. (2014). Ampelopsin induces cell growth inhibition and apoptosis in breast cancer cells through ROS generation and endoplasmic reticulum stress pathway. *PloS One* 9, e89021. doi: 10.1371/journal.pone.0089021
- Zhu, S. S., Zhang, X. Y., Ren, C. Q., Xu, X. H., Comes, H. P., Jiang, W. M., et al. (2023). Chromosome-level reference genome of *Tetrastigma hemsleyanum* (Vitaceae) provides insights into genomic evolution and the biosynthesis of phenylpropanoids and flavonoids. *Plant J.* 114, 805–823. doi: 10.1111/tjp.16169
- Zhuo, X., and Feschotte, C. (2015). Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PloS Pathog.* 11, e1005279. doi: 10.1371/journal.ppat.1005279
- Zwaenepoel, A., and Van de Peer, Y. (2019). wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35, 2153–2155. doi: 10.1093/bioinformatics/bty915