BMC Genomics

RESEARCH Open Access

The whole-genome map of *Astragalus mongholicus* provides ideas for the synthesis of glutathione and flavonoid

Miaozhuo Yao[1†], Fanshu Gong[1†], Yunhui Liang[1], Huixia Gao[1], Pengfei Zhang[2] and Yaling Liu[1*]

Abstract

Background The legume *Astragalus mongholicus* (*A. mongholicus*) has a wide variety of medicinal properties. Based on the genome of *A. mongholicus*, this study intended to investigate the phylogenetic position of *A. mongholicus* and excavate the specific genes of *A. mongholicus*.

Results In this study, we used Oxford Nanopore technology to assemble and annotate a high-quality chromosomal-scale *A. mongholicus* genome. Its total size was 1.60 Gb, containing eight chromosomes. Phylogenetic and comparative genomic analyses on *A. mongholicus* indicating that *A. mongholicus* and other legumes were involved in the WGD event of Papilionaceae evolvement. We identified that 33 genes were significantly associated with GST enzymes in the glutathione pathway in *A. mongholicus*. Moreover, 23 genes were significantly associated with UDP-glucosyltransferase family protein, cyanidin3-O-galactoside 2″-O-xylosyltransferase FGGT1, caffeoyl-coa-o-methyltransferase, phenylcoumaran benzylic ether reductase POP1 enzymes in flavonoid biosynthesis (isoflavones, flavonols and flavonoids), which promoted the synthesis of a series of flavonoids in *A. mongholicus*. The functional versatility of glucosyltransferase synthesis genes may play a key role in the diversity of flavonoid-related biosynthesis.

Conclusion This study provides a valuable model genome for genetic and applied research on *A. mongholicus*.

Keywords *Astragalus Membranaceus* var. *Mongholicus*, Genome sequences, Glutathione, Flavonoid

Introduction

Astragalus, known as "Huangqi" in China, is a beneficial herb with useful medicinal properties [1]. *Astragalus* as medication was first reported in Agriculture God's Canon of Materia Medica [2] and has since been used for over 2,000 years [3]. Modern medical research has shown that *Astragalus* mainly contains saponins, flavonoids, and polysaccharides [4–6]. Medicinal *Astragalus* exhibits immunomodulatory, hypoglycemic, anti-inflammatory, antioxidant, and antiviral activities [6]. Traditionally, the herb is used to treat weakness, wounds, anemia, fever, allergies, chronic fatigue, appetite loss, uterine bleeding, and uterine prolapse [7, 8].

Among the diverse *Astragalus* species, *A. mongholicus* is particularly valued for its medicinal efficacy, yet its genomic basis remains underexplored. As one of the largest genera in Leguminosae, *Astragalus* plays an important phylogenetic role in extant flowering plants. Leguminosae are one of the most common angiosperms in global ecosystems, with a wide distribution ranging

†Miaozhuo Yao and Fanshu Gong are co-first authors.

*Correspondence:
Yaling Liu
lylzpf@126.com
[1]College of Life Science, Shanxi Agricultural University, Taigu, Shanxi Province 030801, China
[2]College of Horticulture, Shanxi Agricultural University, Taigu, Shanxi Province 030801, China

from tropical woody lianas to desert shrubs and a correspondingly diverse range of morphology. Leguminosae include Papilionoideae, Caesalpinioideae, Detarioideae, Cercidoideae, Duparquetioideae, and Dialioideae [9]. Reconstructing the phylogenetic relationships of the Leguminosae is crucial for comprehending the origin and diversification of the ecologically and economically significant plant family [10]. *Astragalus* is part of the Papilionoideae subfamily, which contains approximately 600 genera and more than 10,000 species [11]. The main medicinal tissues of *Astragalus* are the dried roots of *Astragalus membranaceus* (Fisch.) Bge. var. *mongholicus* (Bge.) Hsiao (*Astragalus mongholicus* (*A. mongholicus*)) and *Astragalus membranaceus* (Fisch.) Bge [12]. *A. mongholicus* is mainly distributed in Gansu, Heilongjiang, Shanxi, Hebei and Inner Mongolia, is characterized by its long taproot and abundant mealiness. It is currently the main source of medicinal materials and is superior in both quality and clinical efficacy compared to similar plants from other regions [13]. However, the limited genomic resources available for *A. mongholicus* have precluded an in-depth understanding of its genetics and phylogenetic relationships.

Glutathione (GSH) is commonly found in eukaryotic plants and exhibits antioxidant and comprehensive detoxification effects. It plays a crucial role in disease resistance, cell proliferation, root development, salt tolerance, protection against cold damage and alien organisms, as well as the detoxification of heavy metal metabolism [14]. Astragalus prefers sunny, cold, dry environments, probably due to the internal accumulation of GSH. In selenium-treated *Astragalus*, GSH is found to be the primary non-enzymatic compartment that plays a key role in an intricate network with phenolic compounds [15]. In addition, as a group of phenolic secondary metabolites in *Astragalus*, flavonoids not only contribute to plant growth and enhance plant defense capabilities [16], but also exhibit significant them valuable for clinical applications such as anticancer therapy, anti-inflammatory treatment, and nervous system regulation [17, 18]. The metabolic pathway of flavonoids encompasses phenylpropane, anthocyanin, flavonol, and isoflavone biosynthesis pathways, which constitute the pivotal routes in *Astragalus* secondary metabolite synthesis pathway [19]. However, the evolutionary mechanism of abundant flavonoids in the subsurface rhizome of *Astragalus* remains unclear, which is critical for understanding the biosynthesis of *Astragalus* flavonoids.

In this study, we assembled a high-quality genomic map of *A. mongholicus*. We performed phylogenetic and comparative analyses to determine the evolutionary relationships of *A. mongholicus* in Leguminosae. Moreover, the genes involved in GSH biosynthesis and flavonoid biosynthesis in *A. mongholicus* were analyzed. Overall, the objective of this study was to: (1) assemble a high-quality genome of *A. mongholicus* (2) investigate its phylogenetic position within Leguminosae; and (3) excavate the genes linked to GSH and flavonoid biosynthesis. The *A. mongholicus* genome assembly provides insights into Leguminosae phylogenomics and flavonoid biosynthesis, laying a solid foundation for molecular breeding *of Astragalus* medicinal materials.

Materials and methods

Plant material

This study collected a two-year-old *A. mongholicus* on April 10, 2023 at the experimental base of Shanxi Agricultural University. The voucher specimen of *A. mongholicus* used in this study has been stored in the Herbarium of Traditional Chinese Medicine Resources of the College of Life Sciences of Shanxi Agricultural University, with the collection number Fanshu Gong 20230410–01. It was identified by Researcher Y.L.Liu (Y.L.Liu det, 2023). Total genomic DNA was isolated from fresh leaves using the traditional cetyltrimethylammonium bromide method.

Genome sequencing

First, the Oxford Nanopore library was constructed. After sample quality inspection, large DNA fragments (≥ 15 kDa) were enriched and purified with magnetic beads. The ends of fragmented DNA were then repaired. The DNA library was quantified using QuBit. The purified library was loaded into R 9.4 Spot-On Flow Cells for sequencing using a PromethION P48 sequencer (Oxford Nanopore Technologies (ONT), Oxford, UK). Using the Oxford Nanopore GUPPY v4.0.2, the original image file from high-throughput ONT sequencing was processed with Base Calling and converted to fast5 raw data.

Approximately 100× Illumina short reads *+(insertion length = 150 bp) were generated on the HiSeq2000 platform (Illumina, San Diego, CA, USA) for polishing the assembled genome. FastQC sequencing projects/FastQC (www.bioinformatics.babraham.ac.uk/) was used for processing, detection of low-quality reads, and identifying expressed sequences in the original data. The Hi-C library was prepared according to standard procedures. In situ cross-linked DNA was extracted from approximately 700 ng of high-quality genomic DNA and digested with restriction enzymes. The sticky ends of digested DNA were labeled with biotin and randomly joined together to form chimeric junctions. These biotinylated DNA fragments were further enriched and cleaved to prepare libraries for Illumina sequencing.

Genome size estimation

The genome size of *A. mongholicus* was estimated using a K-mer (k = 19) analysis and Illumina short read sequences. K-mers in DNA samples were counted using

Jellyfish v2.1.4, and genome size was estimated using GenomeScope v1.0. Predicted genome size was approximately 1.24 Gb. Subsequently, 19-mer was used to analyze sequencing data. The frequency-depth distribution of K-mer species deviated from the Poisson distribution indicated the heterozygosity of *A. mongholicus* genome.

Genome assembly and quality assessment

The genome was assembled with a hybrid strategy. First, preliminary assembly was performed using NextDenovo (https://github.com/Nextomics/NextDenovo). Next, with third-generation ONT sequencing data, two rounds of third-generation error correction were performed through the preliminary assembly results of Racon v1.4.11. Two rounds of second-generation error correction were performed on preliminary assembly results after third-generation error correction with Pilon v1.23 [20]. For species with low heterozygosity, the corrected genome was the final assembly. For highly heterozygous species, Purge_haplotigs v1.0.4 was used for reduce heterozygosity after error correction, yielding the final assembly. To assess assembly quality, the BWA-MEM algorithm in BWA v0.7.17 was used to compare Illumina short fragments and assembled transcripts with the genome. The assembly was further evaluated by the Benchmarking Universal Single-Copy Orthologs (BUSCO) v4.1.2 database.

Gene prediction and functional annotation

Repeats were divided into scattered and tandem. Scattered repeats, also known as transposon elements, were grouped into four subcategories: Long Terminal Repeat (LTR) retrotransposons, Long Interspersed Nuclear Elements (LINEs), Short Interspersed Nuclear Elements (SINEs), and DNA transposons. Genomic sequences were subjected to analysis in RepeatMasker v4.0.9 [21], based on the RepBase library (http://www.girinst.org/repbase). RepeatModeler v1.0.11 [22] was then used to build the de novo prediction library based on its own sequence characteristics. Next, RepeatMasker v4.0.9 [21] was used to compare and predict repetitive sequences. Finally, predictions of all replicates were combined to remove redundancy, yielding the final set of genomic repeats. For the prediction of noncoding RNAs (ncRNAs), tRNA sequences in the genome were identified using RNAscan-SE v1.23 [23]; rRNA was obtained using the rRNA database; other ncRNA sequences (e.g., small nuclear RNAs and microRNAs) were searched from against the Rfam database using INFERNAL v1.1.2 [24].

Gene structure was predicted to obtain detailed gene distribution and structural information. Gene structure prediction included the prediction of gene loci, open reading framework, start and stop codons, intron and exon regions, promoters, variable splicing sites, and protein-coding sequences. Homology prediction in Exonerate v2.4.0 [24] selected *Cicer arietinum*, *Arabidopsis thaliana*, *Phaseolus vulgaris*, *Glycine max*, and *Medicago truncatula* as proximate species. De novo assembly was conducted using Augustus v3.3.2 [25], Genscan v1.0, and GlimmerHMM v3.0.4 [26]. Transcripts were obtained from RNA-seq data using Stringtie v2.1.4 [27], and coding frames were predicted utilizing TransDecoder v5.1.0. Finally, MAKER v2.31.10 [28] was used to integrate all predicted gene sets.

Functional annotation of genes was performed via searching the encoded protein sequences against existing protein databases Uniprot [29], the non-redundant (NR) protein, and KEGG [30] databases using diamond blastp v2.0.11.149 (parameter: –evalue 1e-5 v2.0.11.149) [31].

Synteny and WGD event analysis

Protein sequences of different species were compared using BLAST v2.6.0+ (parameters: -evalue 1e-5-outfmt 6) [32]. Genomic collinearity blocks were analyzed using MCScanX (https://github.com/wyp1125/MCScanx; parameters: a-e 1e-5-s 5) [33]. The synonymous mutation frequency (K_s) was calculated with the yn00 module of PAML v4.9 [34]. Finally, WGD events were determined using mapping density in ggplot2 v2.2.1.

Comparative genomics.

For phylogenetic analysis, a phylogenetic tree from the supergenes of 14 species was constructed using model PROTGAMMAWAG in RAxML v8.2.10 (Alexandros n.d.). First, Muscle v3.8.31 [35] was used for multiple sequence alignment of the protein sequences for each single-copy gene family. The results were then filtered using Trimal v1.4.rev22 (parameters: -gt 0.2) [36]. These filtered alignments were combined and connected as a super gene. *Vitis vinifera* (https://www.ncbi.nlm.nih.gov/genome/401?genome_assembly_id=214125) was selected to be the outgroup because its chromosome number was closest to that of dicotyledonous plants, making it an ideal reference genome for studying the genomic evolution of dicotyledons [37].

Species divergence time was estimated with the PAML package mcmctree subroutine v4.9 (parameters: nsample = 1,000,000; Burnin = 200,000; Seqtype = 0; model = 4). Fossil time tables were available from TIMETREE (http://www.timetree.org/).

For comparative genomic analyses, all amino acid sequences from the selected species were first subjected to gene family clustering in Orthofinder v2.3.12 (parameter: -M msa) [38]. Based on the clustering results from Orthofinder, genes with structural similarities were identified within a gene family, thereby identifying single-copy orthologous genes for subsequent evolutionary analysis. Gene Ontology (GO) [39] and KEGG [30] pathway enrichment analyses of *A. mongholicus* genes

were performed using R package clusterProfiler [40]. The number of ancestral gene-family members per branch was estimated by CAFÉ v5.0 [41] using the birth-mortality model, based on the phylogenetic tree and clustering results. The analysis determined gene-family contraction and expansion of a given species relative to its ancestor [42].

Positive selection analysis was performed using the CodeML program within the PAML v4.9 package. Protein sequences per gene family were aligned using MAFFT (parameter: -- localpair -- maxiterate 1000) [43], and then inverted to codon-aligned sequences using PAL2NAL v14 [44]. The CodeML subroutine of PAML v4.9 was used to analyze protein sequences based on the branch-site model. Additionally, the chi2 program in PAML was employed to analyze the two models: model A (assuming that foreground branches ω are all in positive selection, namely ω >1) and a null model (ω ≤ 1). Likelihood ratio tests (LRTs) were conducted to assess the significance of positive selection, with genes considered to be under significant positive selection if $p < 0.05$.

Gene family analysis

The structural analysis of a metabolic gene cluster required performing sequence alignment analysis with resemblant gene clusters from at least two different species to identify conserved and functional regions [45]. Here, the analysis included *Arabidopsis thaliana*, *Medicago truncatula* (NCBI database) in addition to *A. mongholicus*. Sequences from six gene families (OG0000021, OG0000659, OG0001993, OG0000636, OG0000113, and OG0001404) were aligned in MEGA 11.0. A neighbor-joining (NJ) tree was also built using MEGA 11.0 with 1000 bootstrap values and a 50% site coverage cutoff. FIGTREE (https://github.com/rambaut/figtree/releases) was used to set each parameter of the Neighbor-Joining (NJ) tree, and MEME was used to analyze conserved protein domains.

Table 1 Assembly and annotation statistics of the *A. mongholicus* genome

Parameter	Value
Genome size	1.60Gb
GC content	38.53%
Contig number	128
N50 length (contig)	20.73 Mb
N90 length (contig)	6.66 Mb
Longest contig	63,558,937 bp
Gene number	41,520
Average gene length	12,532,050.35 bp
Chromosomes	8
Exons per gene	5.62
Total anchored	1,331,772,021 bp
Repeat	71.60%
Anchored rate	92.73%

Results

Genome sequencing and assembly

The K-mer analysis showed that *A. mongholicus* genome size was approximately 1,510 Mb, with a repetition rate and heterozygosity of 71.60% and 0.91%, respectively (Figure S1). After preliminary assembly, third-generation ONT sequencing data were used for two rounds of third-generation error correction (149.89 Gb) (Table S1a). Next, second-generation sequencing data were used to conduct two rounds of second-generation error correction (179.40 Gb) (Table S1b). During Hi-C library construction, 351,777,194 raw paired-end reads were generated. Next, 22,871,476 valid interaction reads were obtained from the raw reads (Table S2). Sequence data (1.33 Gb) were assigned to eight chromosomes (Table S3). Intrachromosomal interaction signals were intense, indicating a high-quality Hi-C assembly that covered 92.73% of the genome (Figure S2 and Table S4). The final genome sequence was 1.60 Gb, with a contig N50 of 20.73 Mb (Table 1).

The assembled genome was evaluated for consistency and integrity. Consistency was evaluated via calculating alignment rate (99.92%) and coverage (94.10%) of second-generation data. Integrity was assessed using the BUSCO database, which contained 1,614 plant-specific orthologous genes. A total of 1,543 (95.6%) genes were identified, including 1,323 single, 220 complete-duplicate, and 16 fragmented genes. Only 55 patients (3.4%) had missing BUSCO genes (Table S4). These results demonstrated that the final genome assembly was highly complete and contiguous (Fig. 1).

Subsequently, by comparing the data of the *A. mongholicus* genome constructed by Chen et al. [8], we found that the data volume was larger in terms of genome prediction and genome sequencing data processing, but the sequences anchored by Hi-C chromosome assembly were smaller (Table S5). In order to verify the reliability of the *A. mongholicus* genome assembly data, we compared the continuity, consistency and integrity of the assembly. Finally, we found that the N50 value of the genome assembled was about 10,9403,21 bp higher than that assembled by Chen et al. [30], and the map rate reached 99.92%. These results illustrated the high quality of the *A. mongholicus* genome that we assembled by ONT sequencing (Table S6).

Genome annotation

Approximately 1050.39 Mb of repetitive sequences was identified in the *A. mongholicus* genome, accounting for 73.13% of all sequences. Approximately 653.51 Mb of repetitive sequences were LTRs (45.50%). Thus, LTRs are

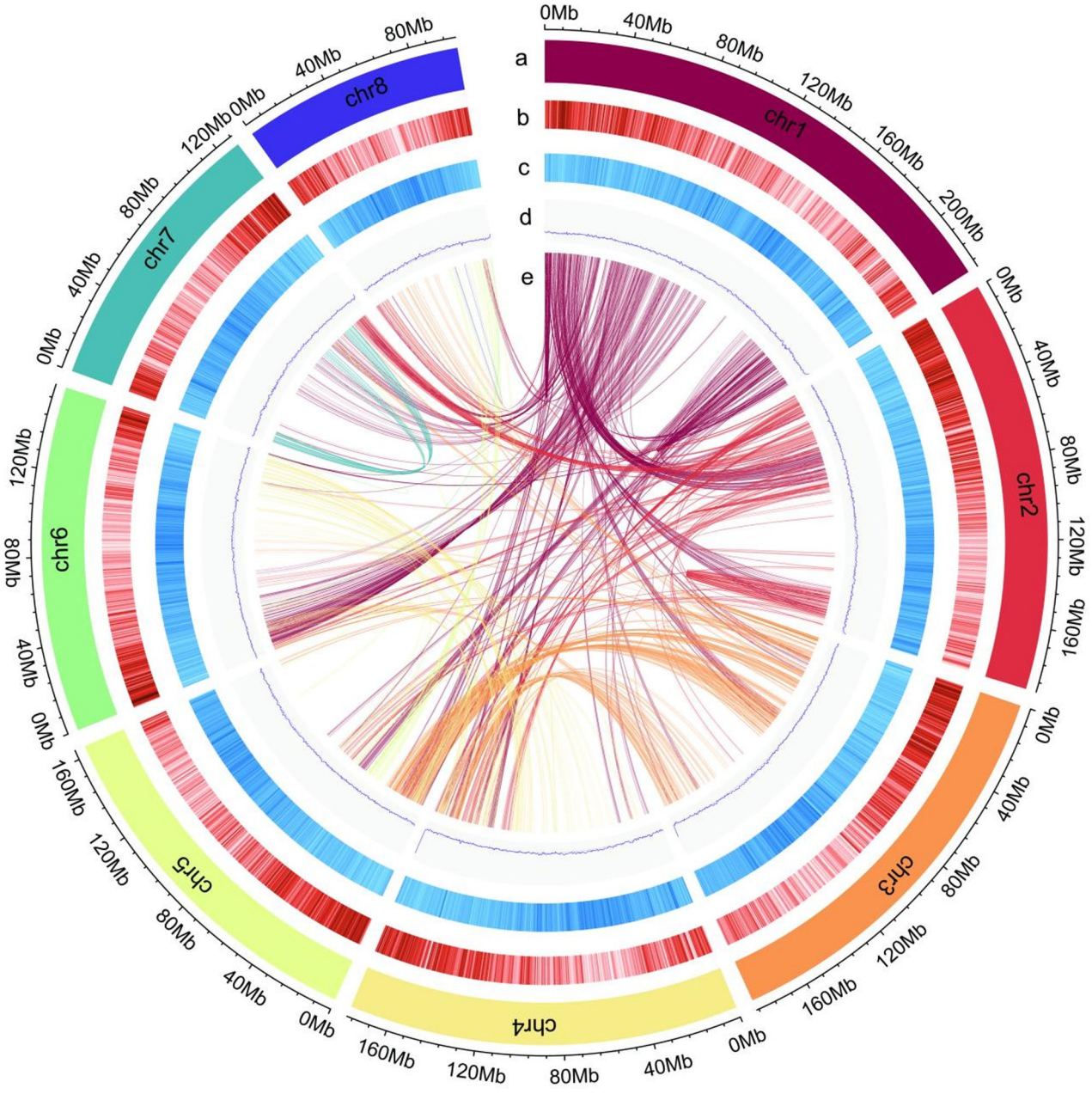

Fig. 1 Genome assembly and annotation map of *A. mongholicus*. (**a**) Chromosome-scale pseudomolecules (Chr1–Chr8). (**b**) Gene density. (**c**) GC content. (**d**) Repeat sequence density and genome syntenic blocks. (**e**) Syntenic blocks in different chromosomes are distinguished with colors

the dominant repetitive sequences in the *A. mongholicus* genome. Further analysis showed that *Copia* (115.40 Mb) and *Gypsy* (520.48 Mb) were the most common LTR subtypes, accounting for 8.03% and 36.24%, respectively (Table S7). By integrating de novo, homologous species, and transcriptome predictions, we identified 41,520 protein-coding genes in nonrepetitive regions, with an average length of 6,197.96 bp and an average exon length of 240.25 bp (Table S8 and Table S9). The ncRNA prediction identified 172 microRNAs, 1,562 tRNAs, and 1,328 rRNAs (Table S10). Functional analysis of the assembled genome annotated 96.59% of the predicted genes (Table S11).

Phylogenetic analysis

Leguminosae is the third largest family of angiosperms, comprising six subfamilies, 765 genera, and approximately 19,500 species [9]. To investigate the evolutionary position of *A. mongholicus* in Leguminosae, we included the species in a phylogenetic analysis with 13 other plants with whole genomes sequenced. We identified 471 strict single-copy homologous proteins of *A. mongholicus*

(Table S12). A phylogenetic tree was constructed on the basis of these homologous proteins. *V. vinifera* was used as the outgroup. Maximum likelihood analysis of a concatenated single-copy-gene supermatrix [46] tended to place *A. mongholicus* in a lineage adjacent to *Cicer arietinum* and *Medicago truncatula.* However, *A. mongholicus* was distant from other legumes (Figure S3). The divergence time of these plants was estimated with MCMCtree. Estimated divergence of *A. mongholicus*, *Cicer arietinum* and *Medicago truncatula* was ~ 34.8 Mya. The most recent common ancestor with *Abrus precatorius* diverged further at ~47.3 Mya (Fig. 2). Notably, plants from the subfamilies Yunshiinae and Sphenoideae may have diverged from the clade around 61.7 to 77.7 Mya, consistent with previous results [9].

WGD analysis

Gene age distribution and co-directional dots provided clear visual evidence of an ancient WGD event in the *A. mongholicus* genome (Figs. 1e and 3, Figure S4). The age distribution of duplicated genes was first used to identify significant peaks of gene replication consistent with WGD events. The mixed model exhibited a peak near K_s = 0.58 (Fig. 3). We compared this age distribution to duplicated gene age distributions of the *C. arietinum* genome, as well as the *M. turcatula* and *V. vinifera* transcriptomes (Figure S5, Figure S6 and Figure S7). Similar repeated peaks were also present in *C. arietinum* and *M. turcatula*, with median K_s of approximately 0.74 and 0.64, respectively (Figure S5 and Figure S7). Therefore, ancient WGD events had occurred in the genomes of *M. turcatula* and *C. arietinum* around the time corresponding to K_s 0.6–0.8. Median K_s (0.58) was earlier than the K_s of *A. mongholicus* and *M. turcatula* (K_s = 0.39), or of

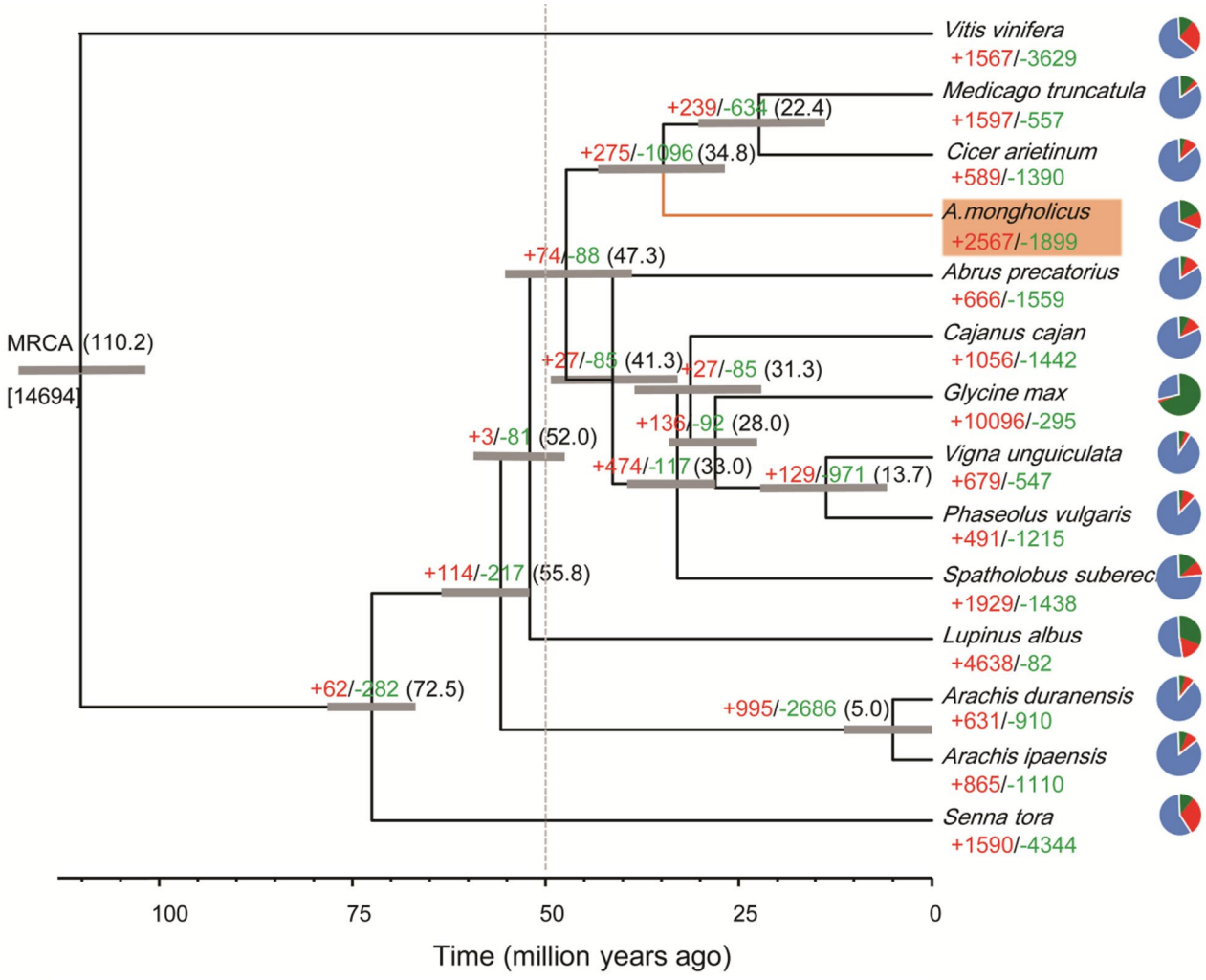

Fig. 2 Phylogenetic analysis on 13 legumes (including *A. mongholicus*) and an outgroup. Expansion and contraction of gene families are represented by numbers with plus and minus signs, respectively. Green numbers in parentheses represent estimated divergence times for each node, and green bars represent 95% confidence intervals for divergence times in millions of years. All nodes are 100% bootstrapped

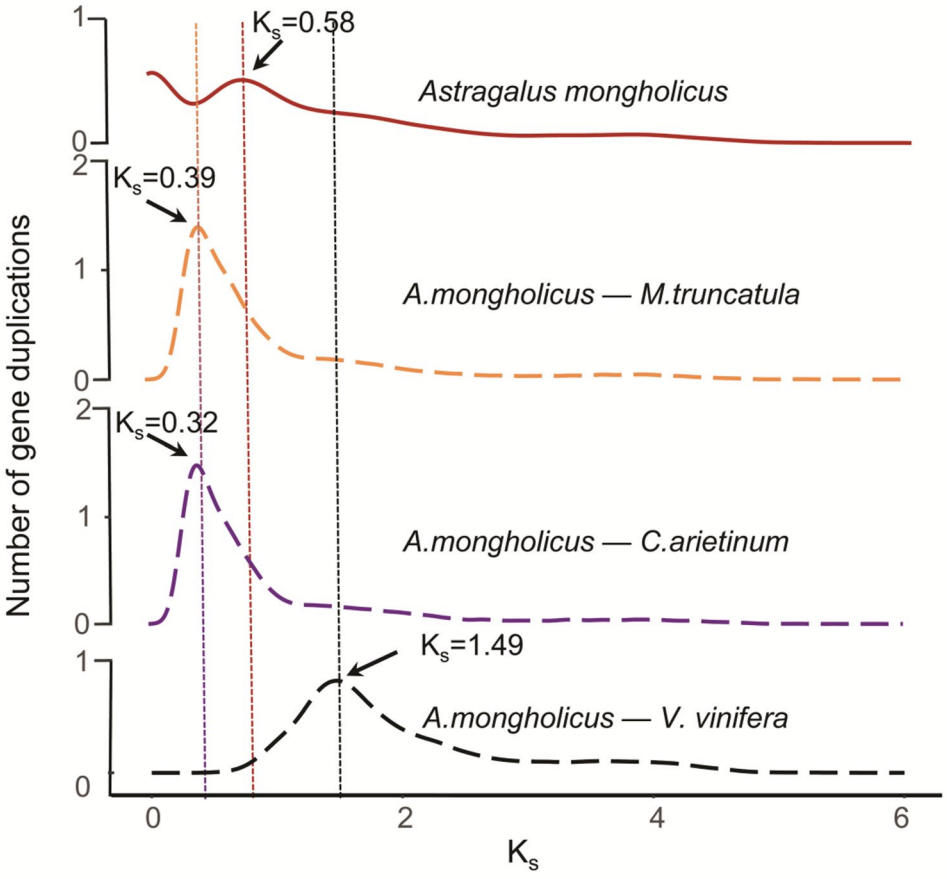

Fig. 3 Synonymy divergence (Ks) distributions of homologous repeats between *A. mongholicus* and *M. runcatula*, as well as between *C. arietinum* and *V. vinifera*

A. mongholicus and *C. arietinum* (K_s = 0.32) (Fig. 3). This pattern suggested that an ancient WGD event occurred in each of these three species, which is consistent with the WGD event of the differentiation of Papilionaceae. Thus, we used K_s analysis and orthogonal divergence between these genomes to assess whether WGD(s) of *A. mongholicus* were shared with *M. turcatula* and *C. arietinum.* In addition, we found that the divergence time between the WGD (K_s = 1.49) of *A. mongholicus* and *V. vinifera* (modern representative closest to ancestral eudicot karyotype [AEK]) was shorter than the divergence between *M. turcatula* and *C. arietinum* (K_s = 0.39, K_s = 0.32).

To investigate whether *A. mongholicus* had undergone ancient WGD, the syntenic depth ratio of the *A. mongholicus* genome was compared with the ratios for *M. turcatula* and *V. vinifera*. We observed an overall depth ratio of two-to-two between *A. mongholicus* and *M. turcatula* (Figure S8), meaning that the two *A. mongholicus* regions were aligned with the two *M. turcatula* regions. These results strongly suggested that *A. mongholicus* did not experience lineage-specific genome-wide replication events. We then compared the genome of *A. mongholicus* with that of *V. vinifera* (replicated three times during the hexaploid event in dicotyledons). The syntenic depth ratio between *A. mongholicus* and *V. vinifera* was two-to-four (Figure S9), consistent with two additional rounds of WGD events in *V. vinifera* compared with *A. mongholicus*. Similarly, the absence of lineage-specific genome-wide replication events in *A. mongholicus* was confirmed by the linear pattern collinearity plot (Fig. 4).

Comparative genomic analysis

To identify potential genes encoding enzymes involved flavonoid biosynthetic pathways, we compared *A. mongholicus* with 11 species of Papilionoideae (*G. max*, *C. arietinum*, *V. unguiculata*, *P. vulgaris*, *S. suberectus*, *M. truncatula*, *A. precatorius*, *C. cajan*, *L. albus*, *A. ipaensis*, *A. duranensis)*, as well as with two species of Caesalpinioideae (*S. tora* and *V. vinifera*). We identified 60,584 direct homologous gene families and 536,263 genes. Of the gene families, 9,517 were shared and 3,567 were unique to *A. mongholicus* (Table S13). Subsequently, we investigated patterns of gene family sharing among *A.*

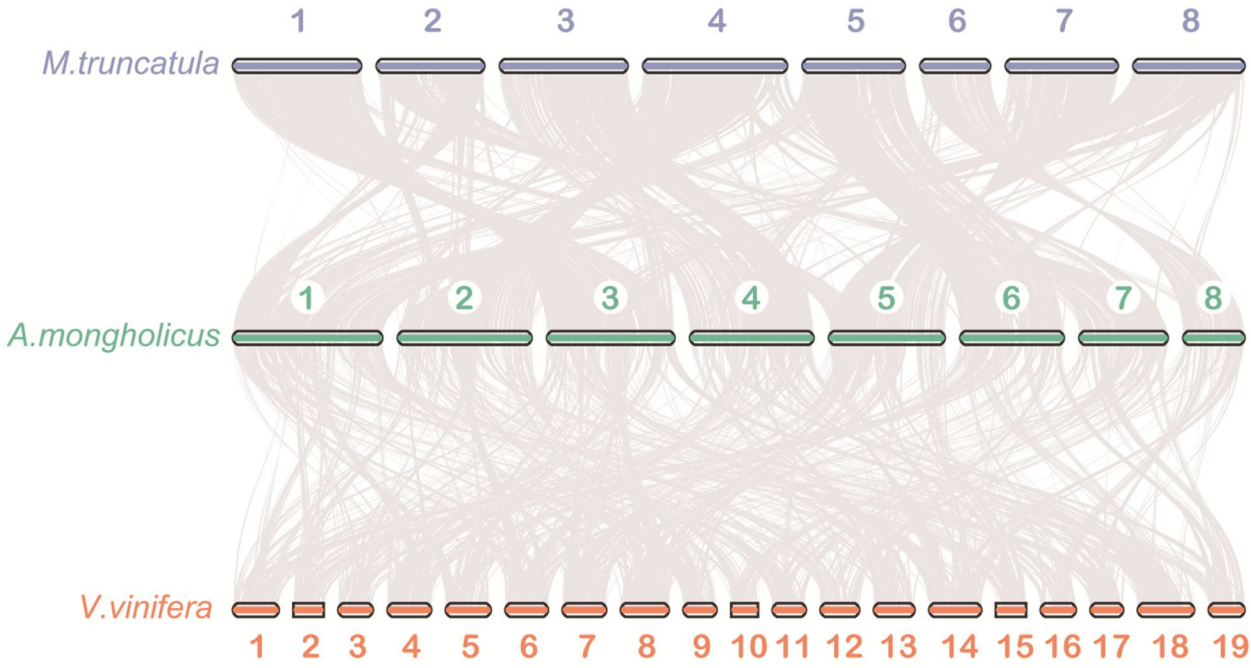

Fig. 4 Synteny analysis of *M. truncatula*, *A. mongholicus*, and *V. vinifera*

mongholicus, *C. arietinum*, *M. truncatula*, and *G. max* (Fig. 5). We identified 13,920 shared genes. Functional (GO and KEGG enrichment) analyses revealed that most of these genes were related to flavonoid biosynthesis, diterpene biosynthesis, zeatin biosynthesis, and circadian rhythms (Figure S10).

Positive selection analysis between *A. mongholicus* and 13 other species identified 78 *A. mongholicus* genes under significant positive selection. GO functional enrichment analysis showed that most of these genes were related to embryo development ending in seed dormancy, DNA (apurinic or apyrimidinic site) endonuclease activity, phosphodiesterase I activity, mitochondrion organization, and chloroplast inner membrane. Next, KEGG pathway analysis indicated that genes were enriched in base excision repair and the sulfur relay system (Figure S11).

Analysis of genes related to GSH synthesis

GSH is known as the "master antioxidant" or "super defender", which plays a necessary role in plant root development, plant disease resistance, low temperature defense and salt tolerance [14]. Through clustering analysis of the gene families of *A. mongholicus* genome expansion, 42 genes were related to GSH biosynthesis in *A. mongholicus*.

KEGG pathway analysis showed that 42 genes were mainly concentrated in the synthesis of glutathione S-transferase (GST) and ribonucleoside-diphosphate reductase (RRM1\RRM2) enzymes, which in turn affected GSH biosynthesis (Fig. 6a). These 42 genes were classified into three gene families (OG0000021; OG0000659; OG0001993) by GO functional classification (Table S14). In order to understand the specific regulation of these genes or the types of synthetic proteins, a NJ tree was built. The 42 proteins were compared with known protein sequences of *Arabidopsis thaliana* and *Medicago truncatula* by using the MEME Suite, 33 genes of OG0000021 and OG0000659 families were found to be involved in the synthesis of GST-related enzyme proteins, in which 24 *A. mongholicus* proteins of cluster I were inferred to be GSTU8/GST, amino terminal domain protein, The two proteins in cluster Ⅱ were respiratory burst oxidase homolog protein C, the two proteins in cluster Ⅲ were GSTU4, the three proteins in cluster Ⅳ were GSTU17, and the three proteins in cluster Ⅴ were GST23 (Fig. 6b, Figure S12). After sequence comparison, we found that seven genes in the *A. mongholicus* genome were related to RRM enzyme synthesis (Figure S13). Through MEME analysis, the inference that 42 genes had similar functions in GSH biosynthesis in *A. mongholicus* was finally confirmed (Figure S12, Figure S13).

Analysis of genes related to flavonoid synthesis

Flavonoid metabolism pathway is a widely existed and very important pathway in plants [47]. Branches of the *A. mongholicus* flavonoid metabolic pathway produced important flavonoids in the flavonoid, flavonol, and isoflavone biosynthesis pathways (Fig. 7a).

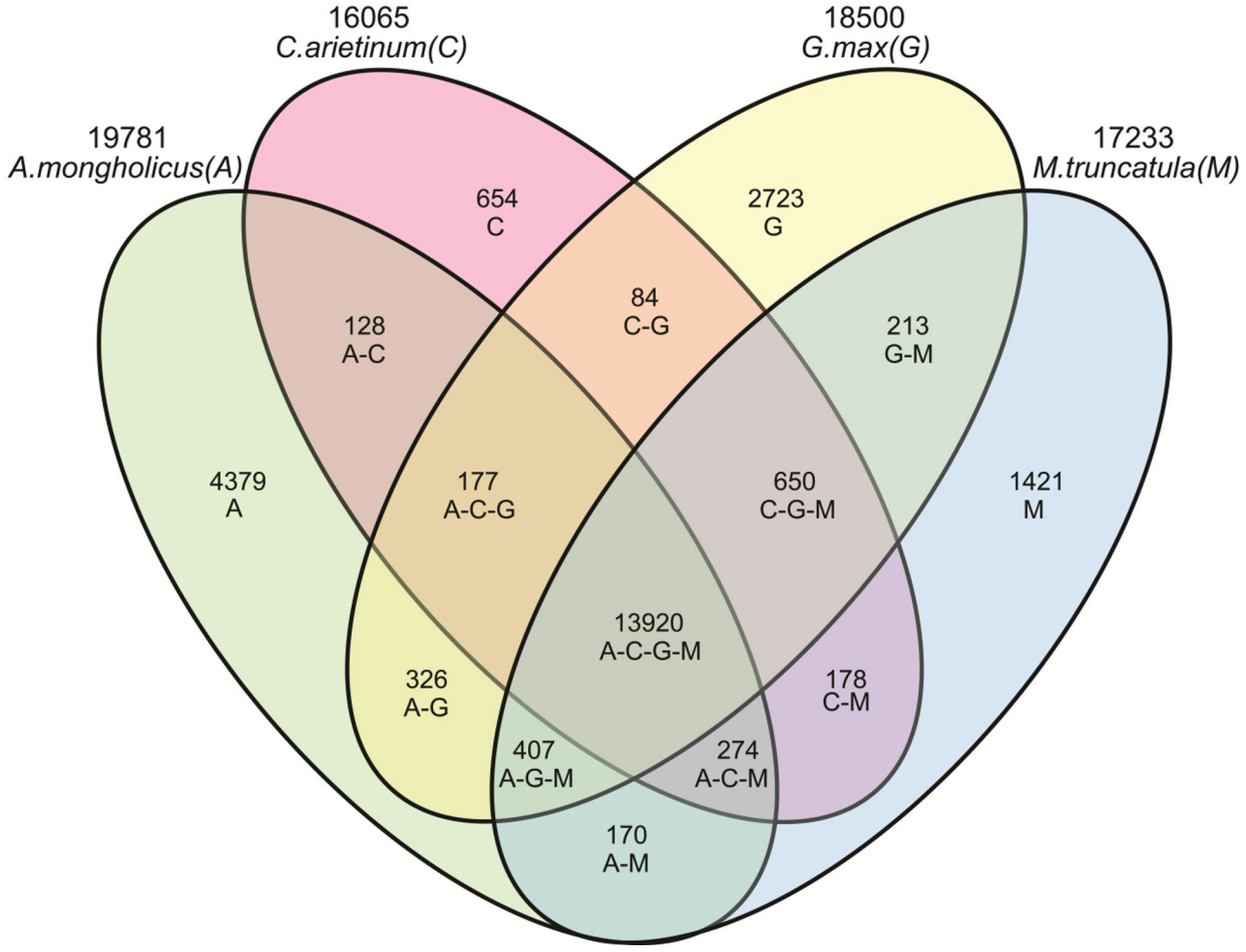

Fig. 5 Venn diagram of *A. mongholicus* and three related plant gene-family clusters

In the isoflavone metabolic pathway, the pterocarpan reductase (PTR) family was found to be significantly expressed in *A. mongholicus*. In the isoflavone biosynthesis of *A. mongholicus*, PTR participated in the process of (-)-medicarpin generation (-)-vestitol. In the biosynthesis of flavonoids and flavonols, flavonol-3-O-glucoside/galactoside glucosyltransferase (FG3) was significantly expressed in *A. mongholicus*. Sophoraflavonoloside, Trifolin -- >Kaempferol 3-O-beta-D-glucosylgalactoside, Isotrifoliin -- >Three biosynthetic pathways of Baimaside. In the biosynthesis of flavonoids, it mainly participated in Caffeoyl-CoA. The caffeoylCoA O-methyltransferase (CCoAOMT) gene family synthesized by Feruloyl CoA was significantly expressed. A total of 23 genes were found to participate in the biosynthesis of flavonoids in the genome of *A. mongholicus*, which were derived from OG0000636, OG0000113 and OG0001404 gene families, respectively (Table S15).

To investigate the specific types of proteins regulated or synthesized by these genes in *A. mongholicus*, the translated proteins of the 23 genes were compared with known protein sequences of *Arabidopsis thaliana* and *Medicago truncatula*, a model plant in the leguminous family. The results showed that 22 genes were mainly involved in the synthesis of UDP-glucosyltransferase family protein and cyanidin3-O-galactoside 2-O-xylosyltransferase FGGT1, caffeoyl-coa-o-methyltransferase, phenylcoumaran benzylic ether reductase POP1 (Fig. 7b). Subsequently, through MEME analysis, the gene of *A. mongholicus* was highly similar to the protein sequences of the two plants, which confirmed our inference of these genes (Figure S14).

Discussion

A. mongholicus is a well-known medicinal and food species renowned for its various pharmacological effects and health benefits. To explore the phylogenetic position of *A. mongholicus* and excavate the specific genes, we assembled a high-quality, chromosomal-scale genome of *A. mongholicus*, and found that *A. mongholicus* and other legumes participated in the WGD event during the evolution of Papilionaceae. Additionally, we identified genes

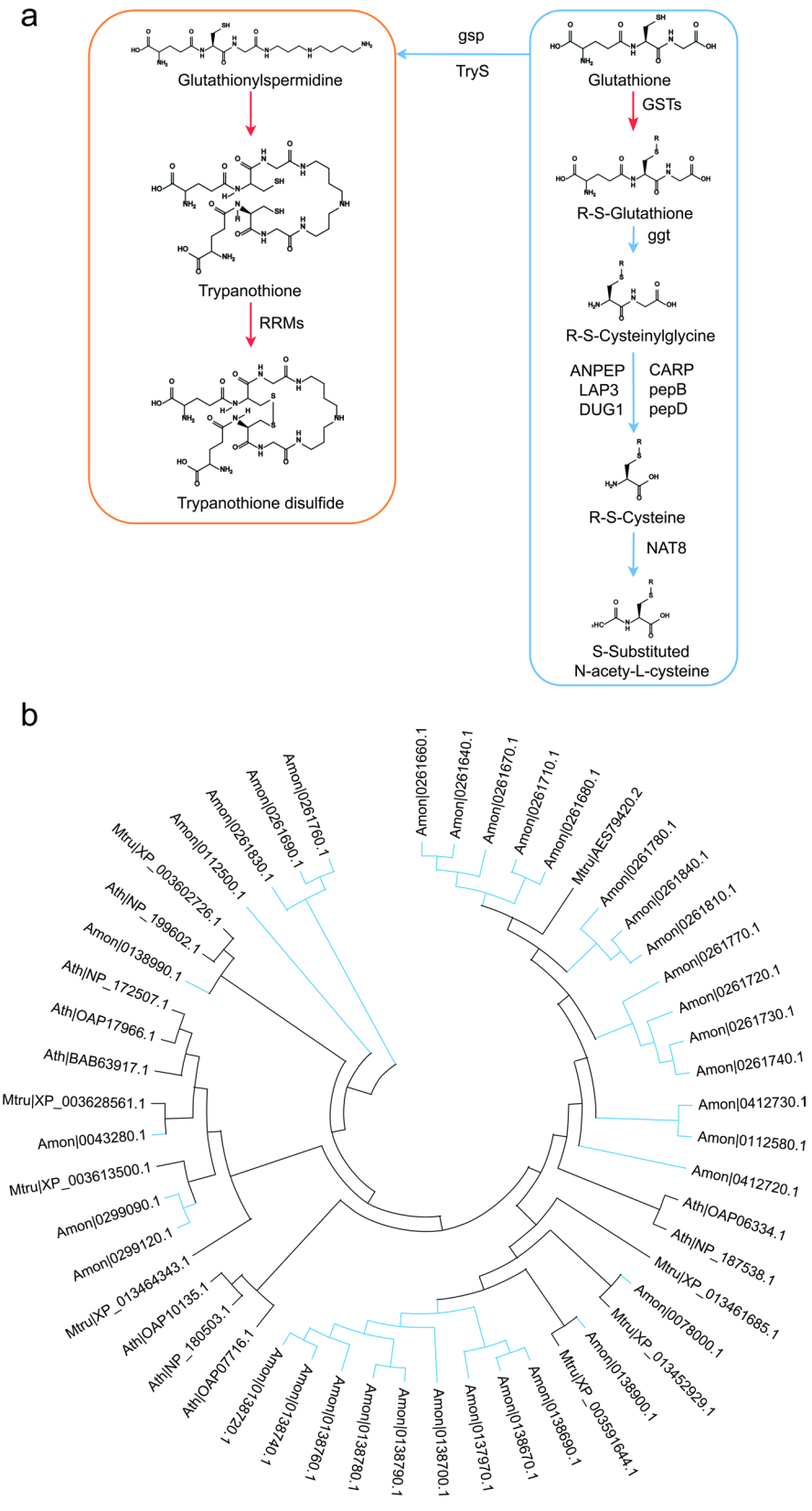

Fig. 6 Partial metabolic pathways and evolutionary analysis of glutathione in *A. mongholicus*. (**a**) glutathione ⟶ R-S-Glutathione pathway (blue), RX⟶ R-S-Glutathione pathway (yellow). (**b**) 33 genes were compared with GSTs protein sequences of *Arabidopsis thaliana* and *Medicago truncatula*

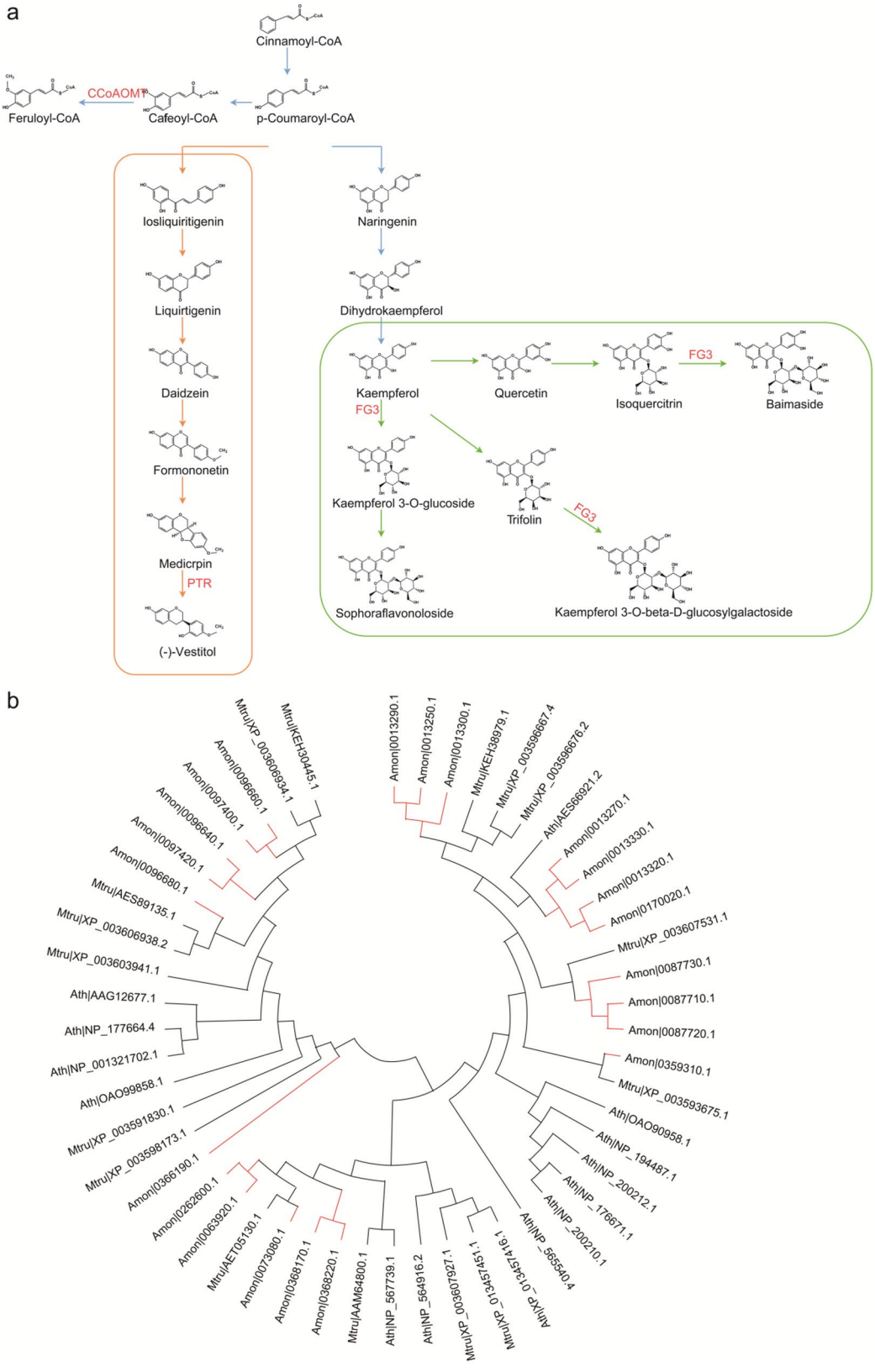

Fig. 7 Metabolic pathways and evolutionary analysis of flavonoids, flavonols, and isoflavones in *A. mongholicus*. (**a**) Flavonoid pathway (blue), isoflavone pathway (yellow), flavonoid and flavonol pathway (green). (**b**) 23 genes were compared with flavone biosynthesis of *Arabidopsis thaliana* and *Medicago truncatula*

associated with the glutathione and flavonoid synthesis pathways in *A. mongholicus*, which might play a pivotal role in flavonoid-related biosynthesis in this species.

This study assembled the *A. mongholicus* genome (coverage rate of 92.73%) using Nanopore and Hi-C technologies. The method we used is similar to techniques for whole genome sequencing in other medicinal plants. Examples include *Coptis chinensis* [48], *Aristolochia contorta* [49], *Bupleurum chinense* DC [50], *Lonicerae japonica* [51], and *Taxus* [52]. The assembled genome allowed us to provide a detailed genome layout of *A. mongholicus*, including its large number of repetitive elements (71.6% of total genome). Gypsy transposons were extremely common, accounting for 36.24% of the *A. mongholicus* genome. Our phylogenetic analysis revealed the position of *A. mongholicus* in legume evolution. As more genomes of flowering species are sequenced, we gain a greater understanding of legume evolutionary relationships. Based on the K_s plot, homology, synchrony, and phylogenetic analysis, we conclude that *A. mongholicus* has not experienced a recent WGD event. However, it has experienced an ancient WGD event with *M. turcatula* and *C. arietinum*, consistent with a recent overview of legume genomes [53]. In the evolutionary analysis of *A. mongholicus*, Chen et al. showed that *A. mongholicus* experienced two rounds of WGD events, the most recent of which was 5.329 billion years ago [8]. *M. turcatula and C. arietinum* experienced a WGD event common to the species after their differentiation from *V. vinifera*, which was consistent with our prediction. Therefore, we expect that additional legume genomes will enable the construction of highly reliable ancestral genomes for angiosperms. This will facilitate the analysis of gene-number changes in the *A. mongholicus* genome after WGD, in comparison with the ancestral angiosperm genome.

Astragalus has been systematically genetically differentiated from other legumes, allowing it to not only retain shared homologous genes but also distinguish itself with its specific genes. This is very important for the unique medicinal ingredients of *Astragalus*. WGD events are revealed as a key driving force in the expansion of gene families and their functional diversification [54, 55]. This process can contribute to the biosynthesis of secondary metabolites in plants [56], which often relies on a large number of duplicated genes that undergo functional divergence after the genome duplication. In this study, the functional enrichment analysis revealed that the majority of these genes are related to flavonoid biosynthesis, diterpene biosynthesis, zeatin biosynthesis, and circadian rhythms. Meanwhile, the expanded gene family of *A. mongholicus* was mainly concentrated in flavonoid biosynthesis. Therefore, our analysis suggests that the ancient WGD events in the evolutionary analysis of *A. mongholicus* may lead to the expansion of key gene families involved in secondary metabolism, such as those responsible for flavonoid biosynthesis like FG3 and GST. The expansion of these gene families, likely due to WGD, may contribute to the increased complexity of its secondary metabolism, particularly the flavonoid pathways that are central to its medicinal properties.

Flavonoids are crucial for various biological processes and help plants respond to environmental stimuli [57]. Flavonoids exhibit a range of pharmacological activities, including antibacterial, antioxidant, anti-inflammatory, antiviral, and anti-allergic properties, thus are responsible for various therapeutic action in humans [58, 59]. Flavonoid, isoflavones, and flavonols are the predominant subclasses of plant flavonoids. Due to the minimal adverse effects and extensive pharmacological activities, flavonols possess functions in the prevention and treatment of various diseases, such as cardiovascular disease [60] and cancers [61]. Notably, quercetin serves as a significant compound within the class of flavonols due to its potent inhibition of thrombine-induced platelet agglutination [62]. Flavonols are synthesized in the flavonoid pathway by flavanone 3-hydroxylase (F3H) and flavonol synthase (FLS), and subsequently undergo modifications by hydroxylase, methyltransferase, and glucosyltransferase enzymes. These pathways regulate the conversion of Astragalin ⟶ Sophoraflavonoloside, Trifolin ⟶ Panasenoside, and Isoquercitrin ⟶ Baimaside. A key enzyme in these processes is FG3. It catalyzes the glycosylation of substrates such as Astragalin, Trifolin, and Isoquercitrin, thereby synthesizing various other flavonoid glycosides through alterations in location, type, and quantity. In this study, FG3 is significantly expressed in the biosynthesis of flavonoids and flavonols in *A. mongholicus*. In the report on soybean (Glycine max) flavonols, researchers found that soybean leaves contain a variety of flavonol glycosides (FGs), which are derivatives of quercetin and kaempferol [63]. Additionally, isoflavone compounds, unique to legumes, are synthesized by isoflavone synthetase catalyzed by flavonoid substances from the upstream flavonoid pathway and subsequently modified by various groups to form diverse isoflavone monomer compounds. In the isoflavone metabolic pathway, we found the PTR family is significantly expressed in *A. mongholicus*. As early as 2006, the PTR family of *Lotus japonicus* was identified and named by Akashi T et al. [64]. Isoflavones and pterocarpols are the main biosynthetic plant antitoxins in legumes, and PTR can split dihydrofuran from pterocarpons into isoflavones (–)-Vestitol. In the isoflavone biosynthesis of *A. mongholicus*, PTR participated in the process of (-)-medicarpin generation (-)-vestitol, both of which have anti-stress and anti-cancer effects [65].

GSH is a small tripeptide molecule composed of glutamic acid, cysteine, and glycine [66]. The synthesis of GSH enables to adapt to various adversities including

detoxification, stress resistance, defense and signal transduction in *A. mongholicus* [67]. Moreover, GST enzymes, an important group of enzymes in the GSH biosynthesis pathway, are related to the synthesis and accumulation of flavonoids [68]. The GSTs represent a ubiquitous class of dimeric proteases within the superfamily, exhibiting diverse biological functions primarily involved in mediating the biosynthesis pathway of Glutathion ⟶ R-S-Glutathione in the metabolic cascade of GSH. Concurrently, GSTs also possess detoxification capabilities by catalyzing the conjugation of GSH with xenobiotic substances or toxic secondary metabolites within an organism, thereby facilitating their metabolism, compartmentalization, sequestration or elimination [62]. Furthermore, GSTs play a crucial role in hormone regulation in plants [69], cellular signaling regulation [70], and apoptosis [71]. In the stressful environmental, GSTs can finely regulate plant adaptability and tolerance, thus impact detoxification, stress resistance, defense, signal transduction and other functions [16]. Meanwhile, studies have shown that GSTs participate in the accumulation of flavonoid by ABC transporters and MATE transporters [72]. In the synthesis of GSH, there were 33 genes focused on the synthesis of GST enzymes in this study. Therefore, we hypothesize that the identified 33 GST genes may regulate the accumulation of flavonoids in *A. mongholicus* ABC transporters and MATE transporters, the mechanisms observed in other legumes.

However, one limitation of this study is the lack of gene expression data, such as RNA sequencing (RNA-seq), to confirm the activity of key genes like FG3 and PTR. Future studies should focus on RNA-seq to validate the expression of these genes, providing deeper insight into the molecular mechanisms behind flavonoid accumulation in *A. mongholicus*.

Conclusions

In summary, the assembled *A. mongholicus* genome provides an excellent resource for the genomics and genetics of this species. It complements and confirms the genomic resources of legumes, and through systematic analysis of the regulatory mechanisms of *A. mongholicus* flavones and related genes, it provides important insights into the metabolic process of this important medicinal ingredient, which will promote future molecular breeding to improve the medicinal ingredients of *A. mongholicus* and promote high yield and rapid growth.

Abbreviations

Abbr.	Meaning(restriction)
A. mongholicus	Astragalus mongholicus
GSH	Glutathione
WGD	Whole Genome Duplication
GST	Glutathione-S-transferase
RRM	ribonucleoside-diphosphate reductase
MYB	v-myb avian myeloblastosis viral oncogene homolog
PTR	pterocarpan reductase
FG3	flavonol-3-O-glucoside/galactoside glucosyltransferase
CCoAOMT	caffeoyl-CoA O-methyltransferase
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
LRT	Likelihood Ratio Test
LTR	Long Terminal Repeat
ONT	Oxford Nanopore Technologies
BUSCO	Benchmarking Universal Single-Copy Orthologs
M. truncatula	Medicago Truncatula
V. vinifera	Vitis Vinifera

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12353-8.

Supplementary Material 1.

Supplementary Material 2.

Authors' contributions

MY and PZ designed the project. MY and FG analyzed the data and wrote the manuscript. YL and HG revised the manuscript. MY contributed to the data analysis. MY and PZ helped with the preparation of samples. All authors read and agreed to the final article.

Funding

This research was funded by the Biotechnology Project of Shanxi Agricultural University (YZGC136), the earmarked fund for Modern Agro-industry Technology Research System (ZYC20), and the National Natural Science Foundation of China (32070378).

Data availability

All data generated or analyzed during this study are included in this published article and its Supplementary information files. The raw sequence reads and genome assembly have been deposited in NCBI under the BioProject accession number PRJNA1358049.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 24 July 2024 / Accepted: 17 November 2025
Published online: 24 November 2025

References

1. Fu J, Wang Z, Huang L, Zheng S, Wang D, Chen S, et al. Review of the botanical characteristics, phytochemistry, and pharmacology of astragalus membranaceus (Huangqi). Phytother Res. 2014;28(9):1275–83.
2. Li S, Sun Y, Huang J, Wang B, Gong Y, Fang Y, Liu Y, Wang S, Guo Y, Wang H, et al. Anti-tumor effects and mechanisms of astragalus Membranaceus (AM) and its specific immunopotentiation: status and prospect. J Ethnopharmacol. 2020;258:112797.
3. Li B, Zhang Q, Chen Y, Su Y, Sun S, Chen G. Different crop rotation systems change the rhizosphere bacterial community structure of astragalus Membranaceus (Fisch) Bge. var. Mongholicus (Bge.) Hsiao. Appl Soil Ecol. 2021;166:104003.
4. Wang Y, Liu X, Hu T, Li X, Chen Y, Xiao G, Huang J, Chang Y, Zhu Y, Zhang H, et al. Astragalus saponins improves stroke by promoting the proliferation

of neural stem cells through phosphorylation of Akt. J Ethnopharmacol. 2021;277:114224.
5. Zheng Y, Ren W, Zhang L, Zhang Y, Liu D, Liu Y. A review of the pharmacological action of astragalus polysaccharide. Front Pharmacol. 2020;11:349.
6. Abd EA, Elkader HT, Essawy AE, Al-Shami AS. Astragalus species: phytochemistry, biological actions and molecular mechanisms underlying their potential neuroprotective effects on neurological diseases. Phytochemistry. 2022;202:113293.
7. Agyemang K, Han L, Liu E, Zhang Y, Wang T, Gao X. Recent advances in astragalus Membranaceus Anti-Diabetic research: Pharmacological effects of its phytochemical constituents. Evidence-based Complement Altern Medicine: eCAM. 2013;2013:654643.
8. Chen J, Ma D, Bao J, Zhang Y, Deng G. Roots of *Astragalus propinquus* Schischkin regulate transmembrane iron transport and ferroptosis to improve cerebral ischemia-reperfusion injury. Evid Based Complement Alternat Med. 2022;2022:7410865.
9. Zhao Y, Zhang R, Jiang KW, Qi J, Hu Y, Guo J, Zhu R, Zhang T, Egan AN, Yi TS, et al. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in fabaceae. Mol Plant. 2021;14(5):748–73.
10. Wojciechowski MF, Lavin M, Sanderson MJ. A phylogeny of legumes (Leguminosae) based on analysis of the plastid MatK gene resolves many well-supported subclades within the family. Am J Bot. 2004;91(11):1846–62.
11. Castillon EE, Quintanilla JAV, Delgado-Salinas A, Rebman JP. The genus *Astragalus* (Leguminosae: Papilionoideae: Galegeae) in Mexico. Phytotaxa. 2023;586(1):1–162.
12. Liu J, Zhang X, Sheng J. Integrative analysis of the transcriptome and metabolome reveals the mechanism of saline–alkali stress tolerance in astragalus Membranaceus (Fisch) Bge. var. Mongholicus (Bge.) Hsiao. J Food Qual Saf. 2022;6:fyac001.
13. Li L, Zheng S, Yang Q, Chen S, Huang L. Distinguishing *Astragalus mongholicus* and its planting soil samples from different regions by ICP-AES. Molecules. 2016;21(4):482.
14. Aslam B, Khurshid M, Arshad MI, Muzammil S, Rasool M, Yasmeen N, Shah T, Chaudhry TH, Rasool MH, Shahid A, et al. Antibiotic resistance: one health one world outlook. Front Cell Infect Microbiol. 2021;11:771510.
15. Maassoumi N, Ghanati F, Zare-Maivan H, Gavlighi HA. Metabolic changes network in selenium-treated *Astragalus* cells derived by glutathione as a core component. Plant Cell Tissue Organ Cult (PCTOC). 2022;149(1):455–65.
16. Nabavi SM, Šamec D, Tomczyk M, Milella L, Russo D, Habtemariam S, Suntar I, Rastrelli L, Daglia M, Xiao J, et al. Flavonoid biosynthetic pathways in plants: versatile targets for metabolic engineering. Biotechnol Adv. 2020;38:107316.
17. Tungmunnithum D, Thongboonyou A, Pholboon A, Yangsabai A. Flavonoids and other phenolic compounds from medicinal plants for pharmaceutical and medical aspects: an overview. Medicines (Basel). 2018. https://doi.org/10.3390/medicines5030093.
18. Zhang L, Ravipati AS, Koyyalamudi SR, Jeong SC, Reddy N, Smith PT, Bartlett J, Shanmugam K, Münch G, Wu MJ. Antioxidant and anti-inflammatory activities of selected medicinal plants containing phenolic and flavonoid compounds. J Agric Food Chem. 2011;59(23):12361–7.
19. Bratkov VM, Shkondrov AM, Zdraveva PK, Krasteva IN. Flavonoids from the genus astragalus: phytochemistry and biological activity. Pharmacogn Rev. 2016;10(19):11–32.
20. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963.
21. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics. 2009, Chap. 4:4.10.11–14.10.14.
22. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117(17):9451–7.
23. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.
24. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31.
25. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinf (Oxford England). 2008;24(5):637–44.
26. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinf (Oxford England). 2007;23(6):673–9.
27. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20(1):278.
28. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.
29. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2004;32(Database issue):D115–119.
30. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
31. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366–8.
32. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
33. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
34. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
35. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
36. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinf (Oxford England). 2009;25(15):1972–3.
37. Yang N. Construction of a comparative genomic map of Cotton's multigenomes and functional evolution analysis. North China University of Science and Technology; 2017.
38. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238.
39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25(1):25–9.
40. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov (Cambridge (Mass). 2021;2(3):100141.
41. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinf (Oxford England). 2021;36(22–23):5516–8.
42. Su X, Yang L, Wang D, Shu Z, Yang Y, Chen S, Song C. 1 K medicinal plant genome database: an integrated database combining genomes and metabolites of medicinal plants. Hortic Res. 2022;9:uhac075.
43. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;47(W1):W5–10.
44. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;34(Web Server issue):W609–612.
45. Nützmann HW, Osbourn A. Gene clustering in plant specialized metabolism. Curr Opin Biotechnol. 2014;26:91–9.
46. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinf (Oxford England). 2014;30(9):1312–3.
47. Liu W, Feng Y, Yu S, Fan Z, Li X, Li J, et al. The flavonoid biosynthesis network in plants. Int J Mol Sci. 2021. https://doi.org/10.3390/ijms222312824.
48. Liu Y, Wang B, Shu S, Li Z, Song C, Liu D, et al. Analysis of the *Coptis chinensis* genome reveals the diversification of protoberberine-type alkaloids. Nat Commun. 2021;12(1):3276.
49. Cui X, Meng F, Pan X, Qiu X, Zhang S, Li C, et al. Chromosome-level genome assembly of *Aristolochia contorta* provides insights into the biosynthesis of benzylisoquinoline alkaloids and aristolochic acids. Hortic Res. 2022. https://doi.org/10.1093/hr/uhac005.
50. Zhang Q, Li M, Chen X, Liu G, Zhang Z, Tan Q, Hu Y, Fan Y, Liu Y, Zhu T, et al. Chromosome-Level genome assembly of bupleurum Chinense DC provides insights into the Saikosaponin biosynthesis. Front Genet. 2022;13:878431.
51. Pu X, Li Z, Tian Y, Gao R, Hao L, Hu Y, He C, Sun W, Xu M, Peters RJ, et al. The honeysuckle genome provides insight into the molecular mechanism of carotenoid metabolism underlying dynamic flower coloration. New Phytol. 2020;227(3):930–43.
52. Xiong X, Gou J, Liao Q, Li Y, Zhou Q, Bi G, et al. The *taxus* genome provides insights into Paclitaxel biosynthesis. Nat Plants. 2021;7(8):1026–36.
53. Egan AN, Vatanparast M. Advances in legume research in the genomics era. Aust Syst Bot. 2019;32(6):459–83.

54. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. Plant Physiol. 2016;171(4):2294–316.
55. Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. Mol Plant. 2018;11(3):414–28.
56. Wang JP, Yu JG, Li J, Sun PC, Wang L, Yuan JQ, et al. Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the Actinidiaceae family. iScience. 2018;7:230–40.
57. Shen N, Wang T, Gan Q, Liu S, Wang L, Jin B. Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. Food Chem. 2022;383:132531.
58. Mutha RE, Tatiya AU, Surana SJ. Flavonoids as natural phenolic compounds and their role in therapeutics: an overview. Future J Pharm Sci. 2021;7(1):25.
59. Ullah A, Munir S, Badshah SL, Khan N, Ghani L, Poulson BG, et al. Important flavonoids and their role as a therapeutic agent. Molecules (Basel). 2020. https://doi.org/10.3390/molecules25225243.
60. Kozłowska A, Szostak-Węgierek D. Targeting cardiovascular diseases by flavonols: an update. Nutrients. 2022. https://doi.org/10.3390/nu14071439.
61. Gangopadhyay A, Chakraborty S, Jash SK, Gorai D. Cytotoxicity of natural flavones and flavonols against different cancer cells. J Iran Chem Soc. 2022;19(5):1547–73.
62. Jiang P, Sheng Y, Ji L. The age-related change of glutathione antioxidant system in mice liver. Toxicol Mech Methods. 2013;23(6):396–401.
63. Di S, Yan F, Rodas FR, Rodriguez TO, Murai Y, Iwashina T, Sugawara S, Mori T, Nakabayashi R, Yonekura-Sakakibara K, et al. Linkage mapping, molecular cloning and functional analysis of soybean gene Fg3 encoding flavonol 3-O-glucoside/galactoside (1 → 2) glucosyltransferase. BMC Plant Biol. 2015;15(1):126.
64. Akashi T, Koshimizu S, Aoki T, Ayabe S. Identification of cDNAs encoding Pterocarpan reductase involved in Isoflavan phytoalexin biosynthesis in *Lotus japonicus* by EST mining. FEBS Lett. 2006;580(24):5666–70.
65. F Y. Effects of promoting osteogenic differentiation of rBMSCs and robs using five kinds of flavonoids from hedysari radix. Chin Traditional Herb Drugs. 2019;19(24):632–8.
66. WJ M, Z. The role of glutathione in plant stress resistance. Guangxi Bot Sci. 2005;6:570–5.
67. Xie Y, He X, Wang D, Wang M, Li W, Chen W, et al. Characterization of dark septate endophytes under drought and rehydration and their compensatory mechanisms in *Astragalus membranaceus*. Microorganisms. 2024;12(11):2254.
68. Nianiou-Obeidat I, Madesis P, Kissoudis C, Voulgari G, Chronopoulou E, Tsaftaris A, Labrou NE. Plant glutathione transferase-mediated stress tolerance: functions and biotechnological applications. Plant Cell Rep. 2017;36(6):791–805.
69. Bilang J, Sturm A. Cloning and characterization of a glutathione S-transferase that can be photolabeled with 5-azido-indole-3-acetic acid. Plant Physiol. 1995;109(1):253–60.
70. Loyall L, Uchida K, Braun S, Furuya M, Frohnmeyer H. Glutathione and a UV light-induced glutathione S-transferase are involved in signaling to chalcone synthase in cell cultures. Plant Cell. 2000;12(10):1939–50.
71. Kampranis SC, Damianova R, Atallah M, Toby G, Kondi G, Tsichlis PN, Makris AM. A novel plant glutathione S-transferase/peroxidase suppresses Bax lethality in yeast. J Biol Chem. 2000;275(38):29207–16.
72. Zhang Y, Zhang F, Wang L, Wei K, Cheng HJPPJ. Plant glutathione S-transferases: roles in flavonoid accumulation. 2015; 51(11):1815–20.

Publisher's note