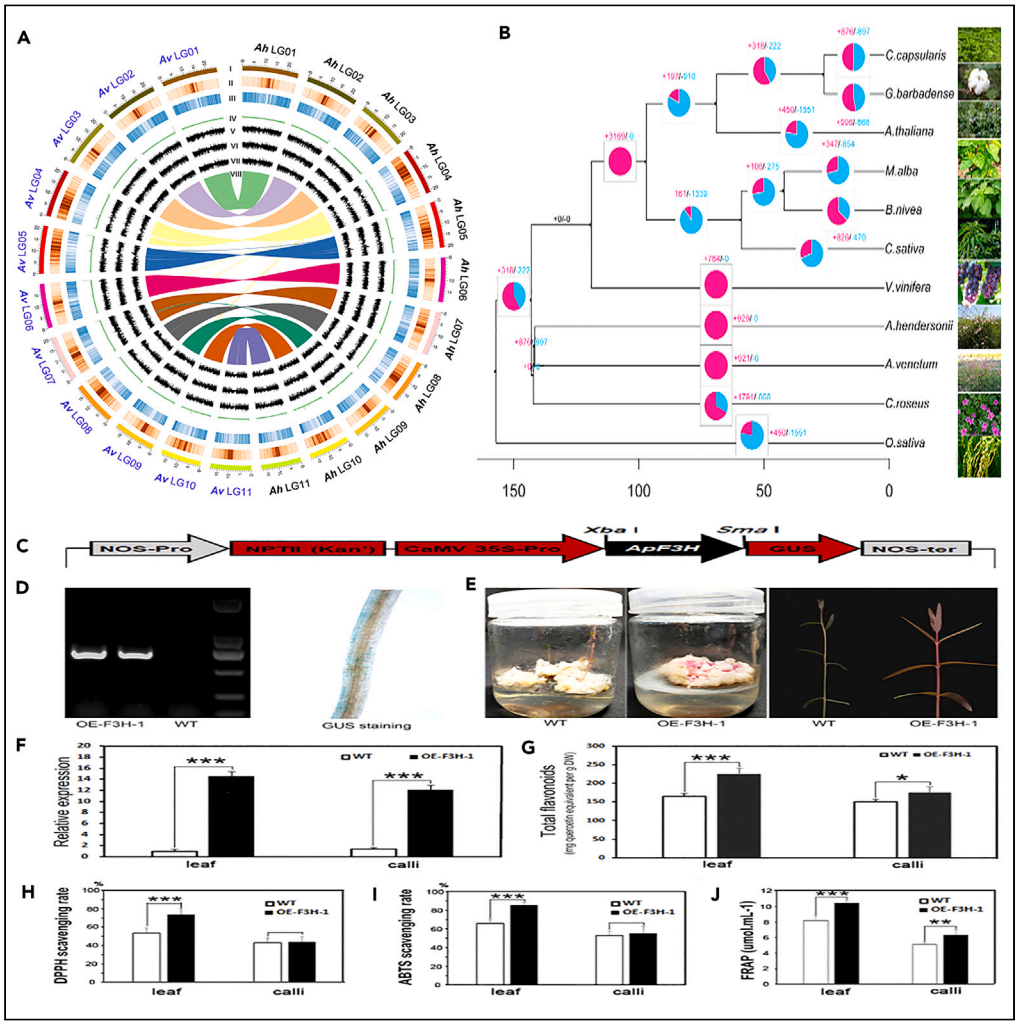
iScience

CellPress
OPEN ACCESS

Article

Comparative genome and metabolome analyses uncover the evolution and flavonoid biosynthesis between *Apocynum venetum* and *Apocynum hendersonii*



Gang Gao, Aminu Shehu Abubakar, Jikang Chen, ..., Yue Wang, Yu Chen, Aiguo Zhu

zhuaiguo@caas.cn

Highlights

The draft genomes of two *Apocynum* species were reported

ApF3H and ApUFGT genes were found critical for flavonoid biosynthesis in Apocynum

Overexpression of ApF3H-1 enhanced total flavonoid content and antioxidant capacity

ApUFGT5 and ApUGT6 explained the diversification of flavonoids in Apocynum

Gao et al., iScience *26*, 106772
May 19, 2023
https://doi.org/10.1016/j.isci.2023.106772

iScience

CellPress
OPEN ACCESS

Article

Comparative genome and metabolome analyses uncover the evolution and flavonoid biosynthesis between *Apocynum venetum* and *Apocynum hendersonii*

Gang Gao,[1,4] Aminu Shehu Abubakar,[1,2,4] Jikang Chen,[1] Ping Chen,[1] Kunmei Chen,[1] Chunming Yu,[1] Xiaofei Wang,[1] Xiaojun Qiu,[1] Xiaoyu Huang,[1] Deyi Shao,[1] Yue Wang,[1] Yu Chen,[1] and Aiguo Zhu[1,3,5,*]

SUMMARY

***Apocynum* species have great application prospects in textile and phytoremediation of saline soil, are rich in flavonoids, and possess medicinal significance. Here, we report the draft genomes of *Apocynum venetum* and *Apocynum hendersonii*, and elucidate their evolutionary relationship. The high synteny and collinearity between the two suggested that they have experienced the same WGD event. Comparative analysis revealed that flavone 3-hydroxylase (*ApF3H*) and differentially evolved flavonoid 3-*O*-glucosyl transferase (*ApUFGT*) genes are critical for determining natural variation in flavonoid biosynthesis between the species. Overexpression of *ApF3H-1* enhanced the total flavonoid content and promoted the antioxidant capacity of transformed plants compared to the wild-type. *ApUFGT5* and *6* explained the diversification of flavonoids or their derivatives. These data provide biochemical insight and knowledge on the genetic regulation of flavonoid biosynthesis, supporting the adoption of these genes in breeding programs aimed at the multipurpose utilization of the plants.**

INTRODUCTION

Apocynum species are perennial plants with rhizomes extending 5–6 m underground. They are adapted to survive under extreme environments, including low rainfall, saline environments, [1]alkaline deserts, riverbanks, alluvial plains, and the Gobi desert.[1] These species are distributed worldwide and primarily grow in Asia, North America, and Eastern Europe. Two notable species (*Apocynum venetum* and *Apocynum hendersonii*), well known in China and Asia, are commonly used in herbal medicine and tea, in addition to their bast fiber.[2] Their medicinal significance is linked to their rich flavonoid constituents, which confer a wide range of physiological functions.[3,4] The two species have a distinguishing morphological difference, with *A. venetum* having characteristically more red flowers and stems, while *A. hendersonii* has a conspicuous white flower, so their common name of red hemp and white hemp, respectively.[2,5] This variation in the flower and stem colors may be due to differences in the functional expression of flavonoid biosynthesis genes.[6,7]

Flavonoids are one of the largest groups of plant secondary metabolites. They play essential roles in a wide range of physiological processes and are becoming increasingly crucial for yield stability and ecological adaptability in the face of environmental deterioration. In addition, their beneficial effects on human health as dietary sources of antioxidants have attracted increasing attention.[8] Because of their similar geographical distribution, *A. venetum* and *A. hendersonii* are usually confused; however, there are distinct differences between the two species, with the former having a red-colored stem and flower with relatively broader leaves.[5] These stem and flower colors may be due to differences in the functional expression of anthocyanins, as recent findings have shown that *A. venetum* has the highest flavonoid content.[7]

The natural variation in flavonoids between *A. venetum* and *A. hendersonii*[6,7] implies that a comparative analysis of the two species could provide insight into the mechanistic themes. Therefore, elucidating the mechanism underlying flavonoid biosynthesis is critical for *Apocynum* breeding. However, the paucity of genomic information has restricted progress on the molecular modification of species.[9] Flavonoids are biosynthesized through a chain of processes, beginning with malonyl-CoA, and 4-coumaroyl-CoA catalyzes by a multi-enzyme complex,[10–12] including chalcone synthase, chalcone isomerase, flavanone 3-hydroxylase,

[1]Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410221, People's Republic of China

[2]Department of Agronomy, Bayero University Kano, PMB 3011, Kano, Nigeria

[3]Key Laboratory of Biological and Processing for Bast Fiber Crops, MARA, Changsha 410221, People's Republic of China

[4]These authors contributed equally

[5]Lead contact

*Correspondence: zhuaiguo@caas.cn

https://doi.org/10.1016/j.isci.2023.106772

flavonoid 3′-hydroxylase, flavonoid 3′,5′-hydroxylase, dihydroflavonol reductase, and UDP-glucose flavonoid3-*O*-glucosyl transferase. Flavonols, anthocyanins, and proanthocyanidins are the major classes of flavonoids and play significant roles in synthesizing other classes, physiology, and morphology in plants based on their structural diversity.[13,14] Flavonoids are mostly glycosylated via the action of flavonoid O-glucosyltransferase (UFGTs) at either the C-3 or C-7 position with glucose or rhamnose sugar. They exhibit diverse physiological functions, are primarily involved in stress response, and serve as co-pigments.[15] Flavonoids produced in response to developmental or environmental needs[16] are often found in nature as derivatives of flavylium ions or 2-phenylbenzopyrilium, in the form of polyhydroxylated or methoxylated heterosides.[17–19]

Recently, combined technologies of bulked segregate analysis (BSA) coupled with randomly amplified polymorphic DNA have been developed to distinguish between the two species.[20] While few studies have assessed the genetics of *A. venetum* and *A. hendersonii*, the inadequacy of gene sequence resources has limited their in-depth study. Chromosome assembly and higher-quality genomes facilitate comparative studies of the evolution of significant plant traits.[21] Therefore, genome data of *Apocynum* species is imperative in providing valuable resources for elucidating flavonoid biosynthesis and related genes, which have always been linked to higher stress tolerance and may aid in the improvement of the species and economic values. Here, we report a chromosome-scale assembly to facilitate in-depth research on *A. venetum* and *A. hendersonii*. Multiomic techniques have provided insight into complex biological processes. Using the draft genomes of *A. venetum* and *A. hendersonii* coupled with transcriptomic, metabolomic, and experimental data, we investigated the evolutionary relationship and natural variation in the accumulation of total flavonoids in the two species. The findings will increase our understanding of the functions of structural genes and facilitate molecular breeding for genetic improvements.

RESULTS

Genome assembly and karyotype analysis

The genomes of *A. venetum* and *A. hendersonii* were estimated by k-mer statistics based on a genome survey, followed by *de novo* assembly of long reads generated by the combined Oxford Nanopore and Hi-C technologies (Figures 1A and 1B, related to STAR Methods). A total of 32.48 Gb clean reads (~ 139.53-fold coverage) for *A. venetum* and 60.48 Gb clean reads (~ 269.97-fold coverage) for *A. hendersonii* were obtained. After correction based on error correction data, we obtained 232.77 Mb contig with an average contig N50 length of 9.68 Mp for *A. venetum* and 240.24 Mb contig with an average contig N50 length of 9.60 Mp for *A. hendersonii* (Tables 1 and S1. Statistics on assembly results and S2. Genome statistics after Hi-C assembly, related to STAR Methods). The chromosome numbers and ploidies of the two species were determined by karyotype analysis. The *A. venetum* and *A. hendersonii* genomes were composed of 22 chromosomes, and the number of chromosomes was confirmed by fluorescence *in situ* hybridization (FISH) and telomere repeat probes (Data S1. TS, Telomere-specific-repeat probe, related to STAR Methods). The ploidy of the two species was determined using FISH with 18S rDNA repeat sequence probes (Data S1. 18S rDNA repeat sequence probe, related to STAR Methods), and the 18S rDNA hybridization signal showed that the cells in the two samples contained two pairs of chromosomes, indicating that both *A. venetum* and *A. hendersonii* were diploid, with 2n = 22 (Figure 1C). The LACHESIS programs employed to investigate the Hi-C scaffolding reliability oriented and anchored the 232.77 Mb contig (representing 99.2% total length) and 240.24 Mb contig (representing 97.4% total length) into 11 pseudochromosomes for *A. venetum* and *A. hendersonii*, respectively (Table 1 and Figure 1A). The 11 chromosomes were clearly distinguished within each group, with the intensity of the interaction at the diagonal position higher than that at the non-diagonal position in both species (Figure 1B). This is consistent with the principle of HiC-assisted genome assembly, supporting the genome assembly effect.

Genome annotation

Genome annotation revealed a total of 756,249,732 (96.49%) and 782,208,273 (99.49%) reads for *A. venetum* and *A. hendersonii*, respectively, mapping to the ref.[22] (Table S3. Statistical information on second-generation data comparison). Gene repertoire completeness revealed 93.61% complete BUSCOs (91.88 single copies and 1.74% duplicated copies) for *A. venetum*. There were 93.40% complete BUSCOs for *A. hendersonii* (91.60% single copies and 1.81% duplicated copies) (Tables 1 and S4. BUSCO evaluation results), indicating higher quality of the annotation and assembly. The assembly covered 98.03 and 97.60% of the 458 core eukaryotic genes (CEG) for *A. venetum* and *A. hendersonii*,

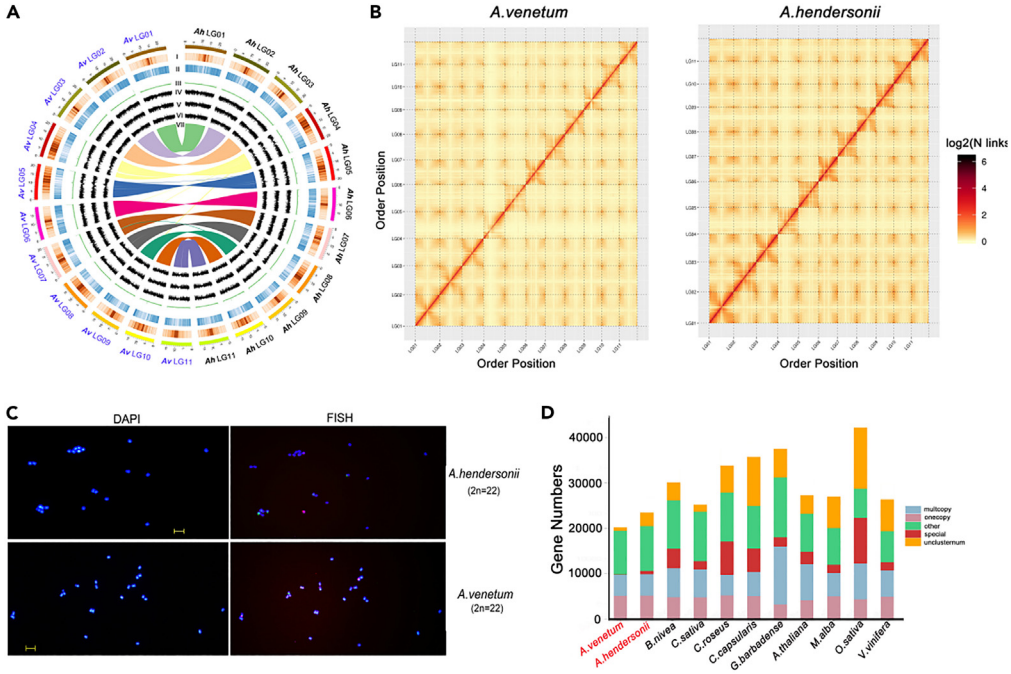


Figure 1. Chromosome karyotype analysis and genomic features of *Apocynum venetum* and *Apocynum hendersonii*

(A) Overview of genome assembly, repeat density (I), gene density (II), GC content (III), gene expression in the root, stem, and leaves (IV, V, VII), and gene synteny (VIII) between *A. venetum* (left half-circle) and *A. hendersonii* (right half-circle) genomes.

(B) Hi-C assembly chromosome interaction heatmap for *A. venetum* and *A. hendersonii*, LG01-LG11 represents lachesis group 01–11, the abscissa and ordinate coordinates represent the order of each bin on the corresponding chromosome group.

(C) DAPI and FISH identification of chromosome karyotype between *A. venetum* and *A. hendersonii*.

(D) Statistics of unique gene families among *A. venetum*, *A. hendersonii*, and nine other plant species.

respectively (Table 1), reflecting completeness.[23,24] These data indicated that the genome assemblies of *A. venetum* and *A. hendersonii* presented high contiguity and sequence quality.

There were 20,292 protein-coding genes (PCGs), including 5,173 one-copy genes and 4,632 multi-copy genes clustered into 15,298 gene families in *A. venetum*. Whereas in *A. hendersonii*, 23,456 PCGs

Table 1. Summary statistics of *A. venetum* and *A. hendersonii* genome assemblies

Assembly parameter	*A. venetum*	*A. hendersonii*
Clean data (Gb)	32.48	60.48
Contig number	62	64
Contig length (bp)	232,771,289	240,238,302
Contig N50 (Mp)	9.68	9.60
Number of 458 CEG* present in the assembly	449 (98.03%)	447 (97.60%)
Number of 248 highly conserved CEGs	227 (91.53%)	230 (92.74%)
Complete BUSCOs	1348 (93.61%)	1345 (93.40%)
Genome size (Mb)	232.77	240.24
GC content	32.38%	32.93%
Repeat density	35.21%	34.80%
Number of protein-coding genes	19,790	22,303
Average length of protein-coding genes	3240	2853

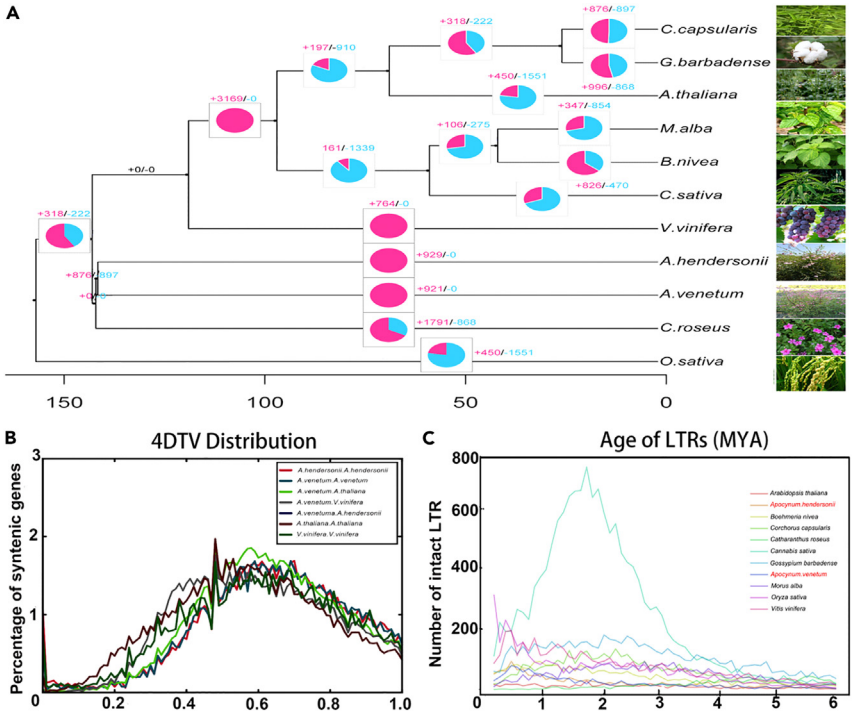


Figure 2. Evolutionary and synteny analysis of *A. venetum*, *A. hendersonii*, and other species

(A) Expansion and contraction of the gene family. The pie diagram on each tree branch represents the proportion of genes undergoing gain (pink) or loss (blue) events.

(B and C) 4DTV distribution and C. Insertion time of LTRs in Apocynum species and other representative species.

comprising 5,151 one-copy genes and 4,717 multi-copy genes clustered into 15,592 gene families were identified (Table S5. Statistics of gene prediction results). Among these, 15,274 orthologous gene families were common to all the 11 selected plant species (Figure 1D, and Table S6. Gene family statistics of *Apocynum species* and 9 others plant species). A total of 24 families containing 68 genes and 245 families containing 733 genes were unique to *A. venetum* and *A. hendersonii*, respectively (Figure S1. Venn diagram of unique gene families among Apocynum species and 9 other plant species), which may be related to environmental adaptation and metabolic pathways within the *Apocynum* lineage. These results indicated that the *A. hendersonii* genome contained more unclustered and unique genes than the *A. venetum* genome. Repetitive sequences and transposable elements (TEs) in *A.* venetum and *A. hendersonii* accounted for 35.21 and 34.80% of the total genome, respectively (Table S7. Statistics for repeat sequences in the *Apocynum* species), and the sequence distribution was more abundant in the LG04 and LG05 chromosomes in both species (Figure 1). Class I TEs formed the dominant repeats in *A. venetum* and *A. hendersonii* at 30.38 and 29.41%, respectively. Copia and Gypsy long terminal repeats (LTRs), as well as large retrotransposon derivatives (LARDs), were the most abundant TEs at 11.80, 10.52, and 6.65%, respectively, in *A. venetum* and 11.70, 10.60, and 5.82% in *A. hendersonii*.

Whole-genome duplication and evolutionary analyses

Gene families that underwent discernible changes and were divergently involved along different branches, with emphasis on those involved in *Apocynum* traits, were analyzed (Figure 2A). The results showed that 921 and 929 gene families exhibited significant expansions ($p < 0.001$) in the *A. venetum* and *A. hendersonii* lineages, respectively, compared with their ancestors. No contraction of this gene family was observed in either species, similar to *Vitis vinifera*. Gene family annotation demonstrated that the primary enriched expansion genes belonged to secondary metabolites, such as those involved in flavonoid biosynthesis. WGD events were investigated by estimating 4-fold degenerate synonymous site (4DTv) mutations in collinear genes (Figure 2B). The peak around 0.53 4DTv indicated shared WGD among *A. venetum*, *A. hendersonii*, *V. vinifera*, and *Arabidopsis thaliana* before divergence. At 0.20–0.80,

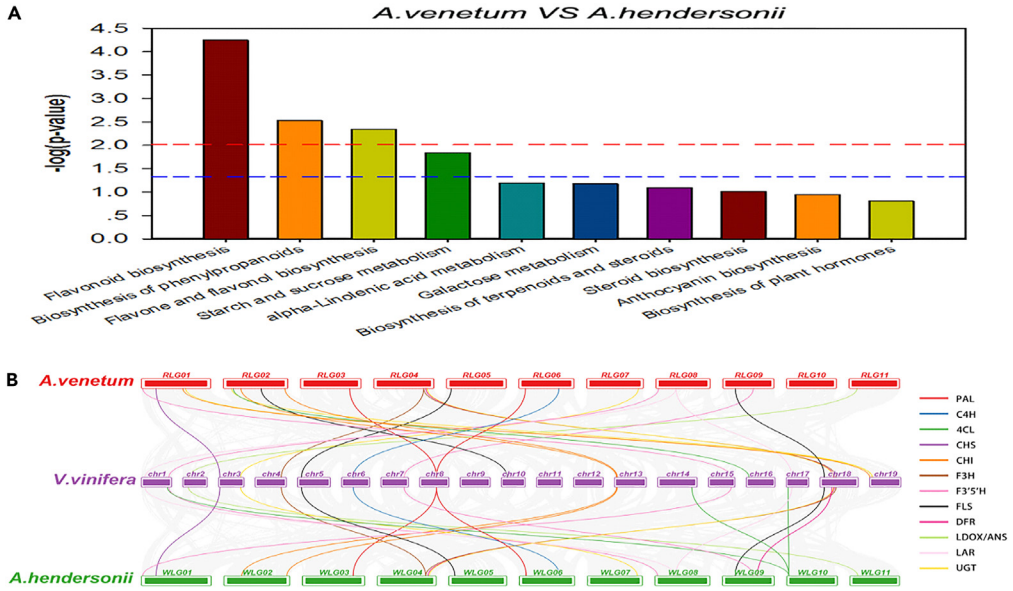


Figure 3. Pathway enrichment and flavonoids biosythesis related genes analysis

KEGG pathway enrichment analysis of differentially accumulated metabolites (A) and genomic arrangement and syntenic analysis of flavonoid biosynthesis-related genes (B) in *A. venetum* and *A. hendersonii*.

A. venetum and *A. hendersonii* had a shared WGD event consistent with the phylogenetic data of the two species (Figures 2B and 2C). The higher synteny and shared collinearity between *A. venetum* and *A. hendersonii* suggested that the two species experienced the same WGD event, indicating that divergence might be more recent, which is in line with the results of the Ks analysis (Data S1. LTRs in the species, related to STAR Methods). The evolutionary analysis confirmed the placement of *A. venetum* and *A. hendersonii* in Apocynaceae, sharing a monophyletic clade with *Catharanthus roseus* (a representative member of the Apocynaceae).[22,25] Both *Apocynum* species diverged successively and showed the closest evolutionary relationships among the 11 species investigated (Figure S2. Evolutionary analysis of *A. venetum*, *A. hendersonii* and other plant species, related to STAR Methods). These species shared a common ancestor approximately 140.8 million years ago (MYA).

Flavonoid profiling and identification of biosynthesis genes

The flavonoid biosynthesis pathway was predominantly enriched in *A. venetum* and *A. hendersonii* (Figure 3A). A total of 223 flavonoids or their derivatives, comprising flavones (74), flavonols (85), flavanones (23), isoflavones (13), and anthocyanins/proanthocyanins (28), were identified in the two species (Data S1. Identified flavonoid components in Apocynum, related to STAR Methods). Among these, 51 flavonoids were differentially accumulated between the two species, with 31 being predominant in *A. venetum* and 20 in *A. hendersonii* (Data S1. Differentially accumulated flavonoids, related to STAR Methods). Quantitative analysis by targeted LC/MS (Table 2) confirmed that in addition to the total flavonoid content and the candidate discrimination marker quercetin 3-D-galactoside,[6] colored compounds, such as gallocatechin, proanthocyanidin, and delphinidin 3-glucoside were also significantly accumulated in *A. venetum* compared to *A. hendersonii*.

Genome assembly allowed us to locate all functionally characterized flavonoid metabolism genes and their closely related homologs. Unlike specialized metabolites, such as momilactones and morphinan, which are produced in high plants by conserved biosynthetic gene clusters,[26,27] the potential genes involved in *A. venetum* and *A. hendersonii* flavonoid biosynthesis were widely distributed on all 11 chromosomes (Figure S3. Chromosomal gene arrangement of flavonoid biosynthesis candidate genes in *A. venetum* and *A. hendersonii*, related to STAR Methods). Gene synteny analysis of candidate flavonoid biosynthesis genes in *A. venetum* and *A. hendersonii* with those in *V. vinifera* (which showed close homologies and richly accumulated high flavonoids) revealed the presence of essential genes encoding enzymes (Figure 3B). Although some isoforms involved in flavonoid biosynthesis are missing in one Apocynum genome and

Table 2. Differentially accumulated flavonoids in *A. venetum* and *A. hendersonii* identified by target LC/MS

	Average content (μg/g)		
Component	*A. venetum*	*A. hendersonii*	Log2(FC)
Quercetin 3-D-galactoside	296.55 ± 32.92	48.35 ± 10.03	2.616
(±)-Gallocatechin	318.34 ± 10.03	89.16 ± 6.78	1.836
Proanthocyanidin A2	0.55 ± 0.05	0.15 ± 0.05	1.874
Procyanidin B1	36.14 ± 4.62	4.11 ± 1.22	3.136
Procyanidin B3	2.31 ± 0.48	0.43 ± 0.03	2.425
Procyanidin B2	673.06 ± 83.9	193.38 ± 35.63	1.799
Delphinidin 3-glucoside	188.17 ± 15.07	32.31 ± 9.44	2.541
Fraxin	5.61 ± 1.28	2.38 ± 0.35	1.237
Phlorizin	35.93 ± 5.44	2.48 ± 0.39	3.856
Rutin	240.06 ± 12.52	236.94 ± 16.34	0.018
Quercetin	270.53 ± 14.12	223.52 ± 13.26	0.275
Epicatechin	553.62 ± 47.52	397.86 ± 53.12	0.476
Catechin	27.21 ± 7.15	16.76 ± 6.12	0.699
(-)-Epigallocatechin	223.23 ± 10.84	220.43 ± 10.52	0.018
Astragalin	121.05 ± 10.58	113.03 ± 7.89	0.098
Aesculin	5.26 ± 0.58	3.39 ± 0.28	0.633
Hexahydroxyflavanone	5.28 ± 1.12	4.41 ± 0.77	0.259
Nicotiflorin	30.11 ± 2.12	20.13 ± 2.03	0.58
Quercetin3-*O*-glucuronide	80.47 ± 10.52	161.08 ± 44.42	−1.001
Myricetin 3-galactoside	27.78 ± 7.55	33.15 ± 7.98	−0.254
Trans-Piceid	3.38 ± 0.39	4.29 ± 0.79	−0.343

vice versa, flavonoid metabolic pathways are well conserved in both. The expression patterns of potential synthetases involved in flavonoid biosynthesis (including PAL, C4H, 4CL, CHS, CHI, F3H, F3′5′H, FLS, DFR, LDOX/ANS, LAR, and UFGT) were analyzed. Among these, the ApF3H and ApUFGT gene families were significantly up/downregulated.

Overexpression of ApF3H genes and determination of total flavonoids

Three pairs of highly conserved F3H genes (with 100% sequence similarity between each pair) were identified in the *A. venetum* and *A. hendersonii* genome: *ApF3H-1* (Av, EVM0009869; Ah, EVM0007724), *ApF3H-2* (Av, EVM0017534; Ah, EVM0010080), and *ApF3H-3*(Av, EVM0019829; Ah, EVM0016058). *ApF3H-2* and *ApF3H*-3, located on chromosome 04, were expressed at the same levels in different tissues of *A. venetum* and *A. hendersonii*. In contrast, *ApF3H-1,* located on chromosome 09, was significantly upregulated in *A. venetum* leaves compared to *A. hendersonii* leaves (Data S1. Fold changes of the F3H genes, related to STAR Methods). Combining metabolic and gene expression data obtained for *A. venetum* and *A. hendersonii*, we identified *ApF3H-1* as a critical enzyme required for total flavonoid accumulation. This variation may significantly affect flavonoid biosynthesis in the duos. Thus, the ORF of *ApF3H-1* under the control of the CaMV35S promoter (Figure 4A) was transformed into *A. hendersonii* and the transgenic lines were screened using kanamycin (100 mg/L) and identified by PCR analysis using a gene-specific primer (Data S1. ApF3H1OE, related to STAR Methods) and GUS staining (Figure 4B). The expression levels of *ApF3H-1* in the calli and leaves were significantly higher in the transgenic lines than in the WT plants, resulting in a significant increase in total flavonoid content and enhanced antioxidant capacity of the plant (Figures 4C–4H).

ApUFGT enzymatic assay and polymorphisms underlying flavonoid variation

After excluding the too-long and partial amino acid-encoding sequences, eight pairs of glycosylated family homologous genes (*AvUFGT1-8* and *AhUFGT1-8*) from *A. venetum* and *A. hendersonii* were characterized. The phylogenetic analysis and unrooted tree obtained using MEGA 10 (neighbor-joining algorithm)

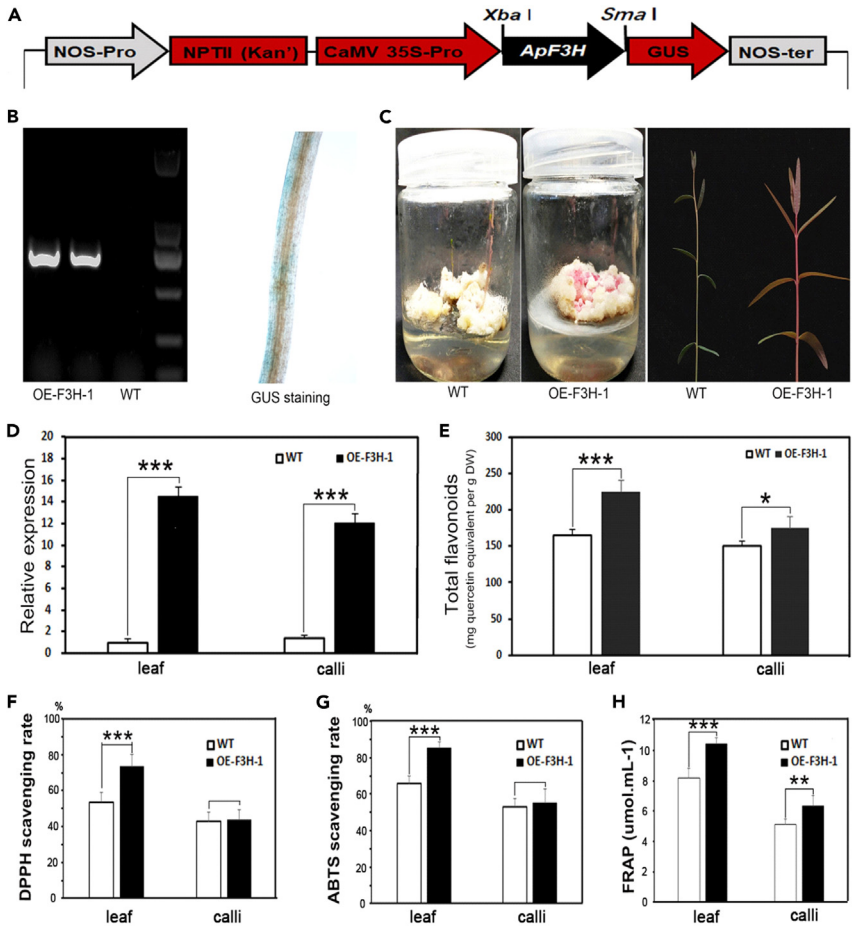


Figure 4. Total flavonoid content in *A. hendersonii* overexpressing ApF3H-1 and its antioxidant capacity

(A) Schematic diagram of the pBI121-ApF3H-1 construct showing the ORF of ApF3H-1 inserted behind the CaMV35S promoter, followed by the GUS and NOS terminator.

(B) Phenotypic comparison of the WT and ApF3H-1-overexpressing *A. hendersonii*.

(C) GUS staining and PCR identification of the ApF3H-1-overexpressing plant.

(D) Real-time PCR analysis of ApF3H-1-overexpressing plant. The relative expression level of ApF3H was normalized to that of AhGAPDH.

(E) The content of total flavonoids in the WT and ApF3H-1-overexpressing plant.

(F–H) DPPH, ABTS radical scavenging activity and reducing antioxidant capacity (FRAP) between the WT and ApF3H-1-overexpressing *A. hendersonii*. Error bars indicate standard deviations. ***$p < 0.001$.

categorized the *ApUFGTs* into unique clusters based on their region-specificity,[28,29] comprising 3 GT (*AvUFGT1*, *AhUFGT1*, *AvUFGT6*, *AhUFGT6*), 5 GT (*AvUFGT3*, *AhUFGT3*), GGT (*AvUFGT2*, *AhUFGT2*, *AvUFGT4*, *AvUFGT4*, *AvUFGT7*, *AhUFGT7*), and CGT (*AvUFGT5*, *AhUFGT5*, *AhUFGT8*, *AhUFGT8*) subfamilies (Figure 5). Eight pairs were highly correlated or differentially evolved *ApUFGTs* suggesting novel specificities. Due to the significance of UFGT glycosylation in the stability, solubility, and bioavailability of flavonoids,[30] *ApUFGTs* were prokaryotic expressed, and recombinant proteins were used for the enzymatic assay. Six ApUFGTs were amplified and successfully expressed in *Escherichia coli* using the pET-30a prokaryotic vector (Data S1. Primers for prokaryotic expression, related to STAR Methods). Enzyme-substrate specificity was determined using UDP-glucose as a sugar donor and three typical flavonoid aglycones (flavonols—kaempferol and quercetin; flavones—apigenin, and flavanones naringenin) as substrates (Table 3).

The ApUFGTs exhibited specific activity against kaempferol, quercetin, and apigenin aglycones but showed no activity against naringenin. *AvUFGT3* and *AhUFGT3* showed the highest catalytic efficiency, with *kcat* values of 0.198 $\pm$ 0.002 and 0.201 $\pm$ 0.002 s^{-1} for kaempferol and 0.097 $\pm$ 0.001 and 0.101 $\pm$ 0.002 s^{-1}

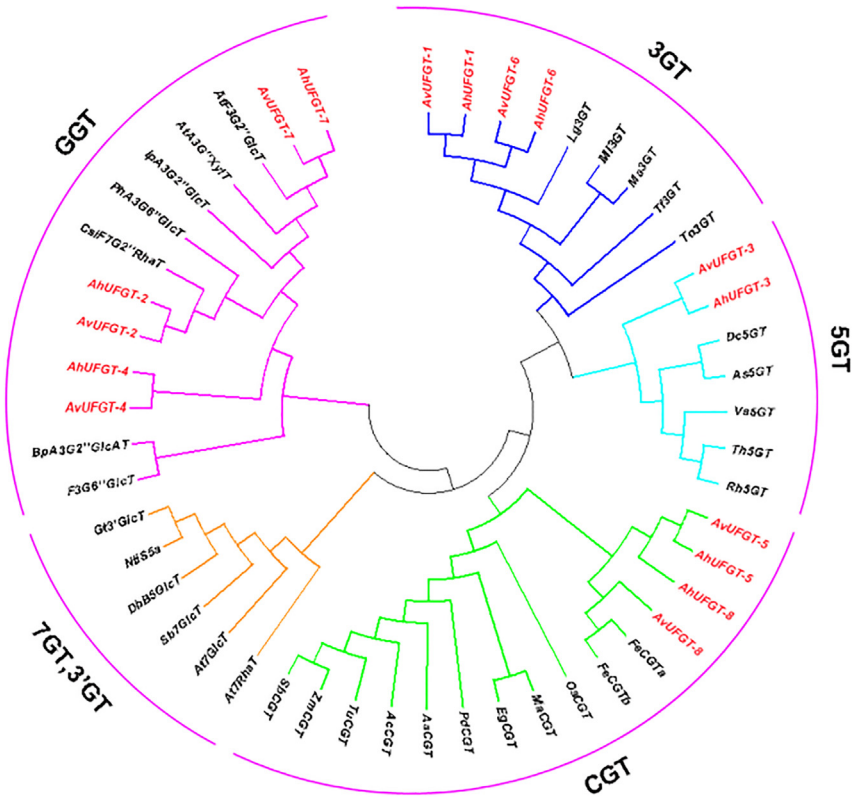


Figure 5. Identification and phylogenetic analysis of flavonoid UFGTs in *A. venetum* and *A. hendersonii*

for quercetin, respectively. In addition, *UFGT3* and *UFGT1* of both *A. venetum* and *A. hendersonii* showed similar enzymatic properties on the selected substrates. All ApUFGTs displayed higher activities with kaempferol than with quercetin, except for *AvUFGT5* and *AhUFGT6*. *AvUFGT5* from *A. venetum* exhibited higher catalytic activity for quercetin and lower activity for kaempferol compared to its *AhUFGT5* ortholog from *A. hendersonii*. *AhUFGT6* exhibited the highest catalytic activity, with a significantly higher *kcat* value of 0.224 $\pm$ 0.005 s^{-1}. Regarding UFGT6, higher activities of flavonols (both kaempferol and quercetin) and considerably lower activities of flavones (apigenin) were observed in *AvUFGT6* compared to *AhUFGT6* (Figure 6).

DISCUSSION

Apocynum species (*A. venetum* and *A. hendersonii*) are perennial herbaceous plants with medicinal significance and excellent ecological adaptability.[3,7,31] Quality genomes and chromosome assembly are crucial for the comparative analysis of significant traits in plants.[21] High-quality genome and chromosome assembly data for the two species revealed substantial differences in repeat sequences, with a higher number of DNA transposons in *A. hendersonii* than *A. venetum*, which might have been attributed to the species' morphology. In morning glory, many mutations that lead to morphological changes, including diversity of flower color, result from TEs.[32,33] Due to the pleiotropic role of the flavonoid biosynthesis pathway in plants, any single mutation may lead to phenotypic differences.[32] Integrated multi-omics analysis of the two species revealed the accumulation of total flavonoids. Overexpression of Ap*F3H-1* in *A. hendersonii* demonstrated potential for improving total flavonoid accumulation, suggesting its exploitation for molecular breeding targeted at enhancing flavonoid content. *ApF3H-1*, which catalyzes the early biosynthesis pathway, determines the differential accumulation of total flavonoids. F3H acts on flavonoids to yield dihydrokaempferol (colourless dihydroflavonol), which is subsequently hydroxylated by flavonoid 3′ hydroxylase (F3′H) or flavonoid 3′,5′ hydroxylase (F3′5′H) to produce red and blue anthocyanins precursors (cyanidin and delphinidin), respectively.[34] The *ApF3H-1* gene was expressed at high levels in *A. venetum* leaves compared to *A. hendersonii*, despite no differences in amino acids between these paralogous

Table 3. Kinetic parameters of ApUFGTs obtained using flavonol (kaempferol and quercetin), flavones (apigenin), and flavanones (naringenin) as substrates

	Flavonol				Flavones		Flavanones	
	Kaempferol		Quercetin		Apigenin		Naringenin	
	Kcat (S^{-1})	*Km* (mM)	*Kcat* (S^{-1})	*Km* (mM)	*Kcat* (S^{-1})	*Km* (mM)	*Kcat*(S^{-1})	*Km* (mM)
AvUFGT1	0.076 ± 0.005	0.194 ± 0.005	0.058 ± 0.003	0.105 ± 0.001	0.073 ± 0.002	0.131 ± 0.002	0.000	0.000
AhUFGT1	0.071 ± 0.003	0.189 ± 0.001	0.061 ± 0.002	0.102 ± 0.001	0.069 ± 0.001	0.126 ± 0.002	0.000	0.000
AvUFGT2	–	–	–	–	–	–	–	–
AhUFGT2	–	–	–	–	–	–	–	–
AvUFGT3	0.198 ± 0.002	0.168 ± 0.001	0.097 ± 0.001	0.248 ± 0.001	–	–	–	–
AhUFGT3	0.201 ± 0.002	0.171 ± 0.001	0.101 ± 0.002	0.256 ± 0.007	–	–	–	–
AvUFGT4	–	–	–	–	–	–	–	–
AhUFGT4	–	–	–	–	–	–	–	–
AvUFGT5	0.019 ± 0.001	0.027 ± 0.001	0.050 ± 0.001	0.083 ± 0.002	–	–	0.001 ± 0.001	0.001 ± 0.001
AhUFGT5	0.021 ± 0.003	0.028 ± 0.002	0.011 ± 0.003	0.019 ± 0.002	–	–	0.000	0.000
AvUFGT6	0.099 ± 0.002	0.719 ± 0.006	0.081 ± 0.003	0.459 ± 0.006	0.033 ± 0.002	0.136 ± 0.001	0.000	0.000
AhUFGT6	0.037 ± 0.003	0.135 ± 0.001	0.045 ± 0.01	0.152 ± 0.0003	0.224 ± 0.005	0.761 ± 0.004	0.000	0.000

copies. This may be because variations in chromosome location or regulatory genes affected gene expression. Consistent with previous studies,[35] the F3H genes in *Apocynum* were highly conserved. Flavonoids endow *Apocynum* with crucial medicinal value, improving the species' resistance and adaptability to adverse conditions. Elucidating the genetic basis underlying the natural variation of plant metabolites provides a better understanding of gene function within ecological and evolutionary contexts. It represents an effective way of determining gene function. This makes it useful for gain-of-function breeding approaches.[36,37]

Given the multi-step nature of flavonoid biosynthesis,[13,38] variation in flavonoids is also determined by tailoring enzymes. Natural variation and diversity of flavonoid constituents between *A. venetum* and *A. hendersonii* were determined by the differentially evolved UFGT, *ApUFGT5,* and *ApUFGT6*. These demonstrated variations in preference for the flavone apigenin and exhibited various kinetic parameters. Decorating enzymes are responsible for the final steps of metabolite synthesis, resulting in a more significant contribution to the natural variation in metabolite abundance. The glycosylation of flavonoids confers structural complexity, molecular stability, subcellular transportability, and biological activity.[39–41] However, knowledge of flavonoid UGTs derives almost exclusively from research on flavonols, which accumulate in *Arabidopsis* and other plants.[15,42,43] In this study, diverse catalytic activity in *ApUFGT5* (CGT subfamily) and *ApUFGT6* (3 GT subfamily) was observed between *Apocynum* species. The mRNA and amino acid analyses revealed that the homologous *AvUFGT1-4* and *AhUFGT1-4* exhibited almost 100% sequence similarity, with some nonsense amino acid mutations. There are nine single nucleotide polymorphisms for *AvUFGT5* and *AhUFGT5*, whereas *AhUFGT6* has a 48 aa deletion mutant at the N-terminus compared to *AvUFGT6*. These results suggested that the two UFGTs are major enzymes that control natural variation in downstream metabolites of the flavonoid biosynthesis pathway between *A. venetum* and *A. hendersonii*. Subsequent investigation of UFTG5 and UFGT6 activities on flavones *in planta* and flavonoid-targeted profiling between the WT and overexpression lines will further contribute to flavonoid biosynthesis in *Apocynum* species.

Apocynum species are rich sources of flavonoids that have medicinal significance, including antimicrobial and antioxidant properties. This may explain their tolerance to adverse environments, such as saline-alkaline soils, where they naturally exist.[2,3] As natural antioxidants, flavonoids protect plants from unfavorable environmental conditions and enhance tolerance to abiotic and biotic stresses. Thus, we correlated flavonoid accumulation with Apocynum's free radical scavenging activity, transgenic lines overexpressing the F3H flavonoid biosynthesis gene, and the concentration-dependent activity of ABST+, DPPH, and Fe^{3+}-TPTZ. Our data showed that *ApF3H-1* promoted the accumulation of total flavonoids, which conferred

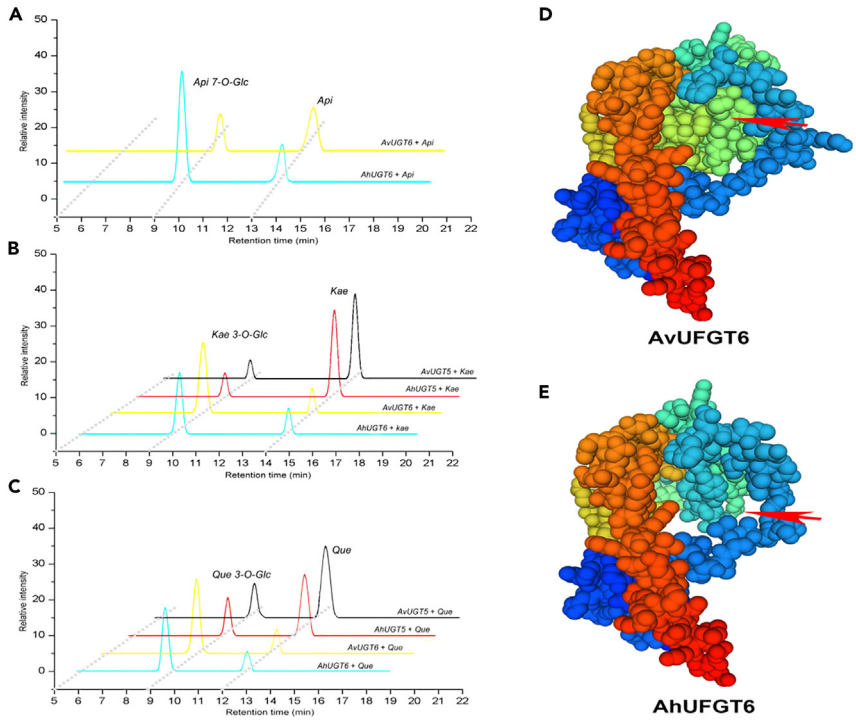


Figure 6. The enzymatic properties and protein structure anlysis of ApUFGTs

Enzymatic assay of ApUFGT5 and 6 proteins by HPLC analysis (A, B, and C) and the tertiary structure of AvUFGT6 (D) and AhUFGT6 (E). Substrates of the enzymatic reaction were flavonols (kaempferol and quercetin) and flavones (apigenin). The red arrows denote the deletion mutant at the N-terminus of the ApUFGT6 protein.

enhanced tolerance to reactive oxygen species to *Apocynum*. Recent studies on Arabidopsis have demonstrated that glycosylated flavonoids can protect against UVB.[15] Flavonoids improve drought tolerance,[44,45] suggesting that increasing flavonoid contents may enhance the species' ecological adaptability and stress resistance. Thus, overexpression of F3H may be exploited for molecular breeding to improve flavonoid accumulation in *Apocynum*. This study has provided important information for future studies of the two species. And further studies are recommended to unravel the roles of repeat sequences in the flavonoid variation and other significant characteristics of *A. venetum* and *A. hendersonii*.

Limitations of the study

Although the genome data and genetic transformation report are of great significance for further study of Apocynum plants, investigating decorating enzymes ApUFTG5 and ApUFGT6 *in planta* and flavonoid-targeted profiling between the WT and overexpression lines will further give a clearer understanding of flavonoid biosynthesis in *Apocynum* species.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Genomic DNA extraction and sequencing
  - Genome assembly and karyotype analysis

- Gene annotation and identification of repeat sequences
- Evolutionary and synteny analysis
- Transcriptome and metabolite profiling
- Identification and analysis of genes involved in flavonoids biosynthesis
- Overexpression of *ApF3H-1* and determination of total flavonoids
- Prokaryotic expression and enzymatic assay of ApUFGTs

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis of the total flavonoid contents, ABTS and DPPH radical scavenging capacities
  - Statistical analysis of the multi-omics data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106772.

ACKNOWLEDGMENTS

This work was funded by the Excellent Young Innovators of Changsha (kq2107017), the Youth Talent Program of Hunan Province (2022RC1013), and the Agricultural Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-IBFC04).

AUTHOR CONTRIBUTIONS

G.G., A.A.S., and Z.A. conceived the study. G.G. and A.A.S. designed and conducted the experiments. C.J., C.P., C.K., and W.X. analyzed the data. Y.C., Q.X., and H.X. selected the plant material and analyzed the phenotypic data. S.D., W.Y., and C.Y. participated in metabolite profiling and RNA-seq. G.G., A.A.S., wrote the manuscript and Z.A. contributed to the final version of the manuscript and supervised the work. G.G., Y.C., and Z.A. acquired the funding. All authors approved the manuscript.

DECLARATION OF INTERESTS

The authors have no conflict of interest to declare.

Received: January 1, 2023
Revised: February 21, 2023
Accepted: April 24, 2023
Published: April 28, 2023

REFERENCES

1. Thevs, N., Zerbe, S., Kyosev, Y., Rozi, A., Tang, B., Abdusalih, N., and Novitskiy, Z. (2012). Apocynum venetum L. and Apocynum pictum Schrenk (Apocynaceae) as multi-functional and multi-service plant species in Central Asia: a review on biology, ecology, and utilization. J. Appl. Bot. Food Qual. 159–167.

2. Gao, G., Chen, P., Chen, J., Chen, K., Wang, X., Abubakar, A.S., Liu, N., Yu, C., and Zhu, A. (2019). Genomic survey, transcriptome, and Metabolome analysis of Apocynum venetum and Apocynum hendersonii to reveal major flavonoid biosynthesis pathways. Metabolites *9*, 296. https://doi.org/10.3390/metabo9120296.

3. Abubakar, A.S., Gao, G., and Zhu, A. (2021). Apocynum venetum, A bast fiber plant with medicinal significances and potentials for drought tolerance and phytoremediation studies – a review. null, 1–13. https://doi.org/10.1080/15440478.2021.1889436.

4. Zhang, Y., Liu, C., Zhang, Z., Wang, J., Wu, G., and Li, S. (2010). Comprehensive separation and identification of chemical constituents from Apocynum venetum leaves by high-performance counter-current chromatography and high performance liquid chromatography coupled with mass spectrometry. J. Chromatogr. B *878*, 3149–3155. https://doi.org/10.1016/j.jchromb.2010.09.027.

5. Abubakar, A.S., Feng, X., Gao, G., Yu, C., Chen, J., Chen, K., Wang, X., Mou, P., Shao, D., Chen, P., and Zhu, A. (2022). Genome wide characterization of R2R3 MYB transcription factor from Apocynum venetum revealed potential stress tolerance and flavonoid biosynthesis genes. Genomics *114*, 110275. https://doi.org/10.1016/j.ygeno.2022.110275.

6. Chan, C.-O., Lau, C.-C., Ng, Y.-F., Xu, L.-J., Chen, S.-B., Chan, S.-W., and Mok, D.K.W. (2015). Discrimination between leave of Apocynum venetum and its adulterant, A. Pictum based on antioxidant assay and chemical profiles combined with multivariate statistical analysis. Antioxidants *4*, 359–372. https://doi.org/10.3390/antiox4020359.

7. Gao, G., Liu, N., Yu, C., Chen, P., Chen, J., Chen, K., Wang, X., Liu, B., and Zhu, A. (2021). UPLC-ESI-MS/MS based characterization of active flavonoids from Apocynum spp. and anti-bacteria assay. Antioxidants *10*, 1901. https://doi.org/10.3390/antiox10121901.

8. Nakabayashi, R., and Saito, K. (2015). Integrated metabolomics for abiotic stress responses in plants. Curr. Opin. Plant Biol. *24*, 10–16. https://doi.org/10.1016/j.pbi.2015.01.003.

9. Chang, L., Yu, X., Wang, W., and Tian, X. (2020). The complete chloroplast genome of *Apocynum venetum* (Apocynaceae). Mitochondrial DNA Part B *5*, 2601–2602. https://doi.org/10.1080/23802359.2020.1781567.

10. Panche, A.N., Diwan, A.D., and Chandra, S.R. (2016). Flavonoids: an overview. J. Nutr. Sci. *5*, e47. https://doi.org/10.1017/jns.2016.41.

11. Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S., and Vianello, A. (2013). Plant flavonoids—biosynthesis, transport and involvement in stress responses. IJMS *14*, 14950–14973. https://doi.org/10.3390/ijms140714950.

12. Treutter, D. (2005). Significance of flavonoids in plant resistance and enhancement of their biosynthesis. Plant Biol. *7*, 581–591. https://doi.org/10.1055/s-2005-873009.

13. Falcone Ferreyra, M.L., Rius, S.P., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front. Plant Sci. *3*, 222. https://doi.org/10.3389/fpls.2012.00222.

14. Xu, W., Dubos, C., and Lepiniec, L. (2015). Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. Trends Plant Sci. *20*, 176–185. https://doi.org/10.1016/j.tplants.2014.12.001.

15. Peng, M., Shahzad, R., Gul, A., Subthain, H., Shen, S., Lei, L., Zheng, Z., Zhou, J., Lu, D., Wang, S., et al. (2017). Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. Nat. Commun. *8*, 1975. https://doi.org/10.1038/s41467-017-02168-x.

16. Quattrocchio, F., Wing, J.F., Leppen, H., Mol, J., and Koes, R.E. (1993). Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell *5*, 1497–1512. https://doi.org/10.1105/tpc.5.11.1497.

17. Di Paola-Naranjo, R.D., Sánchez-Sánchez, J., González-Paramás, A.M., and Rivas-Gonzalo, J.C. (2004). Liquid chromatographic–mass spectrometric analysis of anthocyanin composition of dark blue bee pollen from Echium plantagineum. J. Chromatogr. A *1054*, 205–210. https://doi.org/10.1016/j.chroma.2004.05.023.

18. Hou, Z., Qin, P., Zhang, Y., Cui, S., and Ren, G. (2013). Identification of anthocyanins isolated from black rice (Oryza sativa L.) and their degradation kinetics. Food Res. Int. *50*, 691–697. https://doi.org/10.1016/j.foodres.2011.07.037.

19. Lohachoompol, V., Mulholland, M., Srzednicki, G., and Craske, J. (2008). Determination of anthocyanins in various cultivars of highbush and rabbiteye blueberries. Food Chem. *111*, 249–254. https://doi.org/10.1016/j.foodchem.2008.03.067.

20. Lu, C., Zhang, W., Peng, X., Gu, G., Chen, M., and Tang, Z. (2010). Development of randomly amplified polymorphic DNA-sequence characterized amplified region marker for identification of Apocynum venetum LINN. From A. Biol. Pharm. Bull. *33*, 522–526. https://doi.org/10.1248/bpb.33.522.

21. Chen, D.X., Pan, Y., Wang, Y., Cui, Y.-Z., Zhang, Y.-J., Mo, R.Y., Wu, X.L., Tan, J., Zhang, J., Guo, L.A., et al. (2021). The chromosome-level reference genome of Coptis chinensis provides insights into genomic evolution and berberine biosynthesis. Hortic. Res. *8*, 121. https://doi.org/10.1038/s41438-021-00559-2.

22. Guimarães, G., Cardoso, L., Oliveira, H., Santos, C., Duarte, P., and Sottomayor, M. (2012). Cytogenetic characterization and genome size of the medicinal plant Catharanthus roseus (L.) G. Don. AoB PLANTS *2012*. https://doi.org/10.1093/aobpla/pls002.

23. Pei, L., Wang, B., Ye, J., Hu, X., Fu, L., Li, K., Ni, Z., Wang, Z., Wei, Y., Shi, L., et al. (2021). Genome and transcriptome of Papaver somniferum Chinese landrace CHM indicates that massive genome expansion contributes to high benzylisoquinoline alkaloid biosynthesis. Hortic. Res. *8*, 5. https://doi.org/10.1038/s41438-020-00435-5.

24. Ryan, D.E., Pepper, A.E., and Campbell, L. (2014). De novo assembly and characterization of the transcriptome of the toxic dinoflagellate Karenia brevis. BMC Genom. *15*, 888. https://doi.org/10.1186/1471-2164-15-888.

25. Kellner, F., Kim, J., Clavijo, B.J., Hamilton, J.P., Childs, K.L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., et al. (2015). Genome-guided investigation of plant natural product biosynthesis. Plant J. *82*, 680–692. https://doi.org/10.1111/tpj.12827.

26. Guo, L., Winzer, T., Yang, X., Li, Y., Ning, Z., He, Z., Teodor, R., Lu, Y., Bowser, T.A., Graham, I.A., and Ye, K. (2018). The opium poppy genome and morphinan production. Science *362*, 343–347. https://doi.org/10.1126/science.aat4096.

27. Mao, L., Kawaide, H., Higuchi, T., Chen, M., Miyamoto, K., Hirata, Y., Kimura, H., Miyazaki, S., Teruya, M., Fujiwara, K., et al. (2020). Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. Proc. Natl. Acad. Sci. USA *117*, 12472–12480. https://doi.org/10.1073/pnas.1914373117.

28. Miners, J.O., McKinnon, R.A., and Mackenzie, P.I. (2002). Genetic polymorphisms of UDP-glucuronosyltransferases and their functional significance. Toxicology *181–182*, 453–456. https://doi.org/10.1016/S0300-483X(02)00449-3.

29. Xiao, J., Muzashvili, T.S., and Georgiev, M.I. (2014). Advances in the biotechnological glycosylation of valuable flavonoids. Biotechnol. Adv. *32*, 1145–1156. https://doi.org/10.1016/j.biotechadv.2014.04.006.

30. Li, Y., Baldauf, S., Lim, E.-K., and Bowles, D.J. (2001). Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. J. Biol. Chem. *276*, 4338–4343. https://doi.org/10.1074/jbc.M007447200.

31. Abubakar, A.S., Huang, X., Birhanie, Z.M., Gao, G., Feng, X., Yu, C., Chen, P., Chen, J., Chen, K., Wang, X., and Zhu, A. (2022). Phytochemical composition, antioxidant, antibacterial, and enzyme inhibitory activities of various organic extracts from Apocynum hendersonii. Plants *11*, 1964. https://doi.org/10.3390/plants11151964.

32. Clegg, M.T., and Durbin, M.L. (2000). Flower color variation: a model for the experimental study of evolution. Proc. Natl. Acad. Sci. USA *97*, 7016–7023. https://doi.org/10.1073/pnas.97.13.7016.

33. Wei, L., and Cao, X. (2016). The effect of transposable elements on phenotypic variation: insights from plants to humans. Sci. China Life Sci. *59*, 24–37. https://doi.org/10.1007/s11427-015-4993-2.

34. Falginella, L., Castellarin, S.D., Testolin, R., Gambetta, G.A., Morgante, M., and Di Gaspero, G. (2010). Expansion and subfunctionalisation of flavonoid 3',5'-hydroxylases in the grapevine lineage. BMC Genom. *11*, 562. https://doi.org/10.1186/1471-2164-11-562.

35. Shoeva, O.Y., and Khlestkina, E.K. (2013). F3h gene expression in various organs of wheat. Mol. Biol. *47*, 901–903. https://doi.org/10.1134/S0026893313060137.

36. Mitchell-Olds, T., and Schmitt, J. (2006). Genetic mechanisms and evolutionary significance of natural variation in Arabidopsis. Nature *441*, 947–952. https://doi.org/10.1038/nature04878.

37. Quadrana, L., Almeida, J., Asís, R., Duffy, T., Dominguez, P.G., Bermúdez, L., Conti, G., Corrêa da Silva, J.V., Peralta, I.E., Colot, V., et al. (2014). Natural occurring epialleles determine vitamin E accumulation in tomato fruits. Nat. Commun. *5*, 4027. https://doi.org/10.1038/ncomms5027.

38. Ishihara, H., Tohge, T., Viehöver, P., Fernie, A.R., Weisshaar, B., and Stracke, R. (2016). Natural variation in flavonol accumulation in Arabidopsis is determined by the flavonol glucosyltransferase BGLU6. EXBOTJ *67*, 1505–1517. https://doi.org/10.1093/jxb/erv546.

39. Sun, W., Liang, L., Meng, X., Li, Y., Gao, F., Liu, X., Wang, S., Gao, X., and Wang, L. (2016). Biochemical and molecular characterization of a flavonoid 3-O-glycosyltransferase responsible for anthocyanins and flavonols biosynthesis in freesia hybrida. Front. Plant Sci. *7*, 410. https://doi.org/10.3389/fpls.2016.00410.

40. Yoshihara, N., Imayama, T., Fukuchi-Mizutani, M., Okuhara, H., Tanaka, Y., Ino, I., and Yabuya, T. (2005). cDNA cloning and characterization of UDP-glucose: anthocyanidin 3-O-glucosyltransferase in Iris hollandica. Plant Sci. *169*, 496–501. https://doi.org/10.1016/j.plantsci.2005.04.007.

41. Zhao, Z.C., Hu, G.B., Hu, F.C., Wang, H.C., Yang, Z.Y., and Lai, B. (2012). The UDP glucose: flavonoid-3-O-glucosyltransferase (UFGT) gene regulates anthocyanin biosynthesis in litchi (Litchi chinesis Sonn.) during fruit coloration. Mol. Biol. Rep. *39*, 6409–6415. https://doi.org/10.1007/s11033-011-1303-3.

42. Jones, P., Messner, B., Nakajima, J.-I., Schäffner, A.R., and Saito, K. (2003). UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. J. Biol. Chem. *278*, 43910–43918. https://doi.org/10.1074/jbc.M303523200.

43. Yonekura-Sakakibara, K., Tohge, T., Niida, R., and Saito, K. (2007). Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis

by transcriptome coexpression analysis and reverse genetics. J. Biol. Chem. *282*, 14932–14941. https://doi.org/10.1074/jbc.M611498200.

44. Nakabayashi, R., Yonekura-Sakakibara, K., Urano, K., Suzuki, M., Yamada, Y., Nishizawa, T., Matsuda, F., Kojima, M., Sakakibara, H., Shinozaki, K., et al. (2014). Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. Plant J. *77*, 367–379. https://doi.org/10.1111/tpj.12388.

45. Tohge, T., Wendenburg, R., Ishihara, H., Nakabayashi, R., Watanabe, M., Sulpice, R., Hoefgen, R., Takayama, H., Saito, K., Stitt, M., and Fernie, A.R. (2016). Characterization of a recently evolved flavonol-phenylacyltransferase gene provides signatures of natural light selection in Brassicaceae. Nat. Commun. *7*, 12399. https://doi.org/10.1038/ncomms12399.

46. 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. https://doi.org/10.1093/bioinformatics/btv351.

47. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. Genome Res. *27*, 722–736. https://doi.org/10.1101/gr.215087.116.

48. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

49. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067. https://doi.org/10.1093/bioinformatics/btm071.

50. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

51. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259. https://doi.org/10.1186/s13059-015-0831-x.

52. Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. *31*, 1119–1125. https://doi.org/10.1038/nbt.2727.

53. Bayani, J., and Squire, J.A. (2004). Fluorescence in situ hybridization (FISH). Curr. Protoc. Cell Biol. *22*, 22.4. https://doi.org/10.1002/0471143030.cb2204s23.

54. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7. https://doi.org/10.1186/gb-2008-9-1-r7.

55. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinf. *19*, 189. https://doi.org/10.1186/s12859-018-2203-5.

56. Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. *268*, 78–94. https://doi.org/10.1006/jmbi.1997.0951.

57. Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19*, ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080.

58. Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics *20*, 2878–2879. https://doi.org/10.1093/bioinformatics/bth315.

59. Blanco, E., Parra, G., and Guigó, R. (2007). Using geneid to identify genes. In Current Protocols in Bioinformatics, A.D. Baxevanis, D.B. Davison, R.D.M. Page, G.A. Petsko, L.D. Stein, and G.D. Stormo, eds. (John Wiley & Sons, Inc.), p. bi0403s18. https://doi.org/10.1002/0471250953.bi0403s18.

60. Korf, I. (2004). Gene finding in novel genomes. BMC Bioinf. *5*, 59. https://doi.org/10.1186/1471-2105-5-59.

61. Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. *44*, e89. https://doi.org/10.1093/nar/gkw092.

62. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360. https://doi.org/10.1038/nmeth.3317.

63. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

64. Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. *43*, e78. https://doi.org/10.1093/nar/gkv227.

65. Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., and Buell, C.R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genom. *7*, 327. https://doi.org/10.1186/1471-2164-7-327.

66. Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. *33*, D121–D124. https://doi.org/10.1093/nar/gki081.

67. Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964. https://doi.org/10.1093/nar/25.5.955.

68. Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. *39*, D225–D229. https://doi.org/10.1093/nar/gkq1189.

69. Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol. *5*, R7. https://doi.org/10.1186/gb-2004-5-2-r7.

70. Dimmer, E.C., Huntley, R.P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M.J., Bely, B., Browne, P., Mun Chan, W., Eberhardt, R., et al. (2012). The UniProt-GO annotation database in 2011. Nucleic Acids Res. *40*, D565–D570. https://doi.org/10.1093/nar/gkr1048.

71. Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *27*, 29–34. https://doi.org/10.1093/nar/27.1.29.

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

73. Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21*, i351–i358. https://doi.org/10.1093/bioinformatics/bti1018.

74. Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., and Quesneville, H. (2014). PASTEC: an automatic transposable element classification tool. PLoS One *9*, e91929. https://doi.org/10.1371/journal.pone.0091929.

75. Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA *6*, 11. https://doi.org/10.1186/s13100-015-0041-9.

76. Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics *4*, 4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

77. Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., and Stoeckert, C.J. (2011). Using OrthoMCL to assign proteins to

OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. In Current Protocols in Bioinformatics, D.S. Goodsell, ed. (John Wiley & Sons, Inc.), p. bi0612s35. https://doi.org/10.1002/0471250953.bi0612s35.

78. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

79. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

80. dos Reis, M., and Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. Mol. Biol. Evol. *28*, 2161–2172. https://doi.org/10.1093/molbev/msr045.

81. De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271. https://doi.org/10.1093/bioinformatics/btl097.

82. Han, M.V., Thomas, G.W.C., Lugo-Martinez, 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. https://doi.org/10.1093/molbev/mst100.

83. Zhang, W., Zhang, Y., Qiu, H., Guo, Y., Wan, H., Zhang, X., Scossa, F., Alseekh, S., Zhang, Q., Wang, P., et al. (2020). Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. Nat. Commun. *11*, 3719. https://doi.org/10.1038/s41467-020-17498-6.

84. Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T. -h., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49. https://doi.org/10.1093/nar/gkr1293.

85. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

86. Abu Bakar, M.F., Mohamed, M., Rahmat, A., and Fry, J. (2009). Phytochemicals and antioxidant activity of different parts of bambangan (Mangifera pajang) and tarap (Artocarpus odoratissimus). Food Chem. *113*, 479–483. https://doi.org/10.1016/j.foodchem.2008.07.081.

87. Hassasroudsari, M., Chang, P., Pegg, R., and Tyler, R. (2009). Antioxidant capacity of bioactives extracted from canola meal by subcritical water, ethanolic and hot water extraction. Food Chem. *114*, 717–726. https://doi.org/10.1016/j.foodchem.2008.09.097.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Seeds of *Apocynum* species	Institute of Bast Fiber Crops Chinese Academy of Agricultural Sciences	N/A
Deposited data		
A. venetum sequence data	This study	NCBI:PRJNA787043
A. hendersonii sequence data	This study	NCBI:PRJNA787056
A. venetum root RNA-seq data	This study	SRA:SAMN23766539
A. venetum stem RNA-seq data	This study	SRA:SAMN23766540
A. venetum leaf RNA-seq data	This study	SRA:SAMN23766541
A. hendersonii root RNA-seq data	This study	SRA:SAMN23768475
A. hendersonii stem RNA-seq data	This study	SRA:SAMN23768476
A. hendersonii leaf RNA-seq data	This study	SRA:SAMN23768477
Experimental models: Organisms/strains		
Apocynum venetum	Xinjiang Yili Kazak autonomous prefecture, China	N/A
Apocynum hendersonii	Xinjiang Yili Kazak autonomous prefecture, China	N/A
Oligonucleotides		
Overexpression primer ApF3H1OE, see Data S1	This study	N/A
Primers for prokaryotic expression analysis, see Data S1	This study	N/A
Primers ApF3H1QR for qPCR quantification, see Data S1	This study	N/A
Telomere-specific-repeat probe (TS), see Data S1	Sangon Biotech Co., Shanghai, China	N608360-0096
18s rDNA repeat sequence probe, see Data S1	Sangon Biotech, Shanghai, China	N608380-0096
Software and algorithm		
BUSCO v2.0	Simão et al.[46]	https://busco.ezlab.org/
CEGMA v2.5	Parra et al.[49]	http://korflab.ucdavis.edu/Datasets
HiC-Pro	Servant et al.[51]	http://github.com/nservant/HiC-Pro
Progqenesis QI	Waters Corporation, USA	https://www.nonlinear.com/progenesis/qi/
Other		
PromethION	Oxford Nanopore Technologies	https://nanoporetech.com/products/promethion
ACQUITY UHPLC system	Waters Corpo MA, Milford, USA	https://www.waters.com/waters/en_US/UPLC-UHPLC-systems-and-detectors-for-sub-2-micron-separations/nav.htm?cid=10125009
Triple TOF 5600 System	AB SCIEX, Framingham, MA, USA	https://emphordlas.com/product/tripletof-5600-system/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Zhu Aiguo (zhuaiguo@caas.cn)

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Sequencing data generated have been deposited at NCBI and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Seeds of both *A. venetum* and *A. hendersonii* were initially germinated on MS media. The germinated seedlings and plantlets were used for the transformation. Shoots and leaves were excised, cut to appropriate sizes about 4 - 6 mm, and inoculated with fresh *Agrobacterium tumefaciens* strain, harbouring the pBI121-F3H construct (OD 600 = 0.6 - 0.8) for 30 minutes with continuous shaking. The explants were rinsed two times with double distilled water, and excess moisture removed using sterile filter paper. Subsequently, the explants were placed on MS supplemented with 1 mg/L naphthalene acetic acid (NAA) and incubated in dark at 25°C for two days. The explants were later transferred to MS media containing 1 mg/L NAA, 50 mg/L kanamycin and 100 mg/L Cefotaxime and incubated under 16/8 light/dark photoperiod for callus induction. Shoot and root were induced in MS supplemented with 0.04 mg/l NAA and 0.4 mg/l thidiazuron (TDZ). Hardening was done was done in a controlled chamber by transferring the plant to a potted organic soil mixture and watered using Hoagland solution (1.63 mg/L w/v).

METHOD DETAILS

Genomic DNA extraction and sequencing

DNA samples were extracted from fresh *A. venetum* and *A. hendersonii* leaves obtained from the Xinjiang Research Centre of the Chinese Academy of Agricultural Sciences using the CTAB method. The extracted DNA was fragmented, enriched, and purified, and the fragments were end-repaired to obtain the final library following the Nanopore library construction protocol. The DNA library was quantified using Qubit and Sequenced on the PromethION sequencer platform at Biomarker Biotechnology Co., Ltd. Beijing. Nanopore sequencing data were evaluated and assessed by filtering low-quality reads, removing the joint to obtain reads, and correcting the base error to get high-accuracy data for genome assembly, post-assembly evaluation, and other analyses.

Genome assembly and karyotype analysis

The third generations of Nanopores data obtained were error-corrected using Canu.[47] The integrity of the assembled genome was evaluated based on the second-generation sequencing read comparison rate, core genetic integrity, and BUSCO assessment. Second-generation read comparison refers to the proportion of clean reads in relation to the reference genome, and was performed using BWA.[48] CEGMA v2.5[49] and BUSCO v2.0[46] were used to assess the integrity of the final genome assembly. Hi-C fragment libraries with 300–700 bp insert sizes were constructed and sequenced using the Illumina platform according, as previously described.[50] HiC-Pro software[51] was used for quality control, and RACKES[52] was used to evaluate the assembly results. Accurate karyotyping was confirmed using fluorescence *in situ* hybridization.[53] A telomere-specific repeat probe was used to verify the number of intact chromosomes, and an 18s rDNA repeat sequence probe was used to identify multiple copies of chromosomes. Chromosomes were stained using 4′,6-diamidino-2-phenylindole (DAPI), and the dispersed metaphase chromosome cells were counted under a fluorescence microscope.

Gene annotation and identification of repeat sequences

The genetic structures of *A. venetum* and *A. hendersonii* were annotated based on *de novo* prediction, homologous species prediction, and unigene prediction.[54,55] Genscan,[56] August[57] v2.4, GlimmerHMM[58] v3.0.4, GeneID[59] v1.4S and NAP[60] (version 2006-07-28) set at default were used for the *de novo* prediction. GeMoMa[55,61] v1.3.1 with default parameters was used for homologous species based predictions with protein sequences of *Arabidopsis thaliana*, *Catharanthus roseus*, *Coffea canephora* and *Solanum lycopersicum* obtained from NCBI. Hisat[62] v2.0.4 (set at: –max-intronlen 20000,–min-intronlen 20) and Stringtie[63] v1.2.3 with default parameters were employed for the genome assembly based on reference transcripts. TransDecoder (http://transdecoder.github.io) v2.0 and GeneMarkS-T[64] v5.1 both set to default for gene prediction while PASA[65] v2.0.2 (parameters: -align_ tools gmap, -maxIntronLen 20000) was used in

validating the predicted unigene sequences and finally EVM[54] v1.1.1 was used to integrate the predicted results.

Non-coding RNA (microRNAs, rRNA, and tRNA) were also predicted. Rfam[66] with Blistn for genome-wide comparison was employed to identify microRNAs and rRNAs, and tRNAscan-SE[67] in the identification of tRNA. At the same time, the genome of both species was annotated for gene function with Nr,[68] KOG,[69] GO[70] and KEGG[71] metabolic pathway.

Functional annotations were performed on the gene set using BLASTP with default settings against the NCBI, NR, SwissProt, and KOG databases. LTR_FINDER[72] and RepeatScout[73] were used with default parameters to build a database of repeat genome sequences based on structural and *de novo* prediction principles. PASTEClassifier[74] was used to classify the database, and together with Repbase[75] formed the final repeating sequence database. Using RepeatMasker,[76] we obtained repeat predictions for the genome based on a constructed repeat sequence database.

Evolutionary and synteny analysis

Orthologous gene families were constructed between *A. venetum*, *A. hendersonii,* and nine other species (*Oryza sativa* GCA_001648735.1, *Arabidopsis thaliana* GCA_000222325.1, *Vitis vinifera* GCA_002922885.1, *Morus alba* GCA_012066045.3, *Cannabis sativa* GCA_001865755.1, *Gossypium barbadense* GCA_001856525.1, *Catharanthus roseus* GCA_028651295.1, *Boehmeria nivea* GCA_002806895.1, and *Corchorus capsularis* GCA_001974805.1) using OrthoMCL.[77] MUSCLE[78] was used for multiple sequence alignment of single-copy orthologs for all 11 species. A maximum likelihood tree was drawn using RAxML.[79] Subsequently, a phylogenetic tree was constructed using the GTRGAMMA model with *V. vinifera* as an outgroup. The protein sequences were transformed into CDS, and MCMCTREE[80] was used to estimate the divergence time of *Apocynum* species from other species. Expansions and contractions of gene families were computed using CAFÉ,[81] and 4DTv[82] obtained based on the CDS alignment was used to depict whole genome duplications (WGD). The 4DTv substitution per synonymous site (Ks) was computed for each homologous gene pairs within and between species. The age of LTR divergence time was calculated using DistMat software following Kimura model with a 7.3 × 10−9 molecular clock.[83] Synteny analysis was conducted between *Apocynum species* and *V. vinifera* using MCscanX.[84]

Transcriptome and metabolite profiling

RNA was extracted from the stems and leaves of *A. venetum* and *A. hendersonii* using the AG21019 SteadyPure Plant RNA Extraction Kit (Accurate Biotechnology Co., Ltd., Hunan, China), according to the manufacturer's instructions. These RNA-seq libraries were then sequenced through the Illumina HiSeq 2000 platform and aligned to the already assembled and annotated *Apocynum* genomes. Gene expression levels were computed as the number of reads per kilobase of gene length per million mapped reads (FPKM) using RSEM software.[85] Metabolites were profiled using targeted/untargeted metabolomics methods at Oebiotech Biotechnology Co., Ltd. (Shanghai, China) (http://www.oebiotech.com/), and six independent samples for each group were analyzed.

A liquid chromatography-electrospray ionization–tandem mass spectrometry (LC–ESI–MS/MS) system was used for relative metabolite quantification. The ACQUITY UHPLC system (Waters Corpo MA, Milford, USA) and Triple TOF 5600 System (AB SCIEX, Framingham, MA, USA) equipped with an UPLC BEH C18 column (1.7 μm, 2.1 × 100 mm) in both positive and negative ion modes was used for the analysis according to previous studies.[2,31] The binary gradient elution system were 0.1 % formic acid in water, v/v (A) and 0.1 % formic acid in acetonitrile, v/v (B) and separation done under the gradient: 0 min, 5% B; 2min, 20% B; 4min, 25% B; 9min, 60% B; 14min, 100% B; 18min, 100% B; 18.1 min, 5% B and 19.5min, 5%B. The flow rate was 0.4 mL/min and column temperature of 45 °C. The samples temperature was maintained at 4°C with an injection volume of 50 μL and a full scan mode (m/z ranges from 70 to 1000) was used for data acquisition.

Identification and analysis of genes involved in flavonoids biosynthesis

The sequences of genes associated with flavonoid biosynthesis were downloaded from UniProt (https://www.uniprot.org) and identified using BLASTP against the database with an E-value cut-off of 1e-.[10] These included genes related to enzymes potentially involved in flavonoid biosynthesis pathways, such as PAL, C4H, 4CL, CHS, CHI, F3H, F3'5'H, FLS, DFR, LDOX/ANS, LAR, and UFGT. Integrated analyses of

comparative transcriptomic, genomic, and phytochemical data based on gene annotation were performed to elucidate the molecular mechanisms of flavonoid biosynthesis in *A. venetum* and *A. hendersonii*.

Overexpression of *ApF3H-1* and determination of total flavonoids

The ORF sequence of Ap*F3H* was amplified using cDNA from *A. hendersonii* and cloned into the pBI121 overexpression vector. The recombinant vector was introduced into Agrobacterium strain LBA4404 and then transferred into *A. hendersonii* using tender stems as explants. The positively transformed plants were selected on kanamycin (100 mg/L)-supplemented MS media and identified by PCR using specific primers and GUS staining. Gene expression was quantified by qRT-PCR, and relative expression was calculated using the $2^{-\Delta\Delta CT}$ method. The primers used for qRT-PCR are listed in the supplementary file (Data S1. ApF3H1QR, ApF3H2QR, and ApF3H3QR).

Prokaryotic expression and enzymatic assay of ApUFGTs

The open reading frame (ORF) of ApUFGTs was amplified and inserted into the pET-30a vector. Recombinant pET-30a-ApUFGT plasmid DNA was then transformed into *Escherichia coli* BL21 cells for prokaryotic expression. Recombinant ApUFGT enzymes were purified using a Biologic DuoFlowTM chromatography system (Bio-Rad, USA), following the manufacturer's protocol, and then used for enzymatic assays.[15] *In vitro* enzyme reactions were performed in a total volume of 100 μL containing 200 μM flavonoid substrates, flavonols (kaempferol and quercetin), flavones (apigenin), and flavanones (naringenin, 1.5 mM UDP glucose, 5 mM $MgCl_2$, and 500 ng purified protein in Tris-HCl buffer (100 mM, pH 7.4). The mixture was incubated at 37°C for 20 min, and the reaction was stopped by adding 300 μL of ice-cold methanol. The reaction mixture was then passed through a 0.2 μm filter (Millipore) before use for LC-MS analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of the total flavonoid contents, ABTS and DPPH radical scavenging capacities

Total flavonoids from wild and transformed plants were extracted and measured as previously described.[86] Free radical scavenging capacity was measured in three test systems using the stable radical $ABTS^+$ [2,2-azino-bis-(3-ehylbenzothiazoline-6-sulfonic acid)], DPPH (2,2′-diphenyl-1-picrylhydrazyl), and FRAP assays. The percentage inhibition of ABTS+ (I_{ABTS+} %), DPPH (I_{DPPH} %), and FRAP (μmol·mL^{-1}) was calculated as previously reported.[87]Analysis of variation was performed with GenStat 17th edition and means separated using the Student's Newman Keul method.

Statistical analysis of the multi-omics data

The raw metabolic analysis data were analyzed by the progqenesis QI software (Waters Corporation, USA). Standard internal detection parameters were deselected for peak RT alignment and isotopic peaks excluded for analysis. Noise elimination level and minimum intensity were set at 10.00 and 15 % of base peak intensity respectively. The matrix obtained was further reduced by eliminating any peaks with missing value (ion intensity = 0) realized in more than 60 % samples. The internal standard was used for data QC (reproducibility). 0.3 mg/ml 2-Chloro-L-phenylalanine (CAS: 103616-89-3) in Methanol solution was use as reference substance.

The kinetic constants of ApUFGTs for flavonoid acceptance were determined using 0–400 μM of the different flavonoids at a fixed concentration of 1.5 mM UDP-glucose. Flavonoids and UDP sugars were purchased from Sigma–(Ald)rich USA. All kinetic parameters were calculated using the Michaelis-Menten model (Sigma Plot, version 12.5), and all reactions were performed in triplicate.