Plant Biotechnology Journal

WILEY

RESEARCH ARTICLE OPEN ACCESS

Multi-Omics Insights Into Anthraquinone Biosynthesis in *Rheum tanguticum*

Shuo Zhao[1,2] | Yuan Zong[3] | Tao Wang[2,4] | Jianan Li[2,4] | Zengrong Ye[2,4] | Famei Zhang[2,4] | Guoying Zhou[1,2]

[1]Sichuan Zoige Alpine Wetland Ecosystem National Observation and Research Station, Southwest Minzu University, Chengdu, China | [2]Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, Qinghai, China | [3]Hong Kong Baptist University, Hong Kong, China | [4]University of Chinese Academy of Sciences, Beijing, China

Correspondence: Guoying Zhou (zhougy@nwipb.cas.cn)

Received: 25 July 2025 | **Revised:** 3 December 2025 | **Accepted:** 29 December 2025

ABSTRACT

Rheum tanguticum is renowned for its medicinal properties, including purgative, anti-inflammatory and hepatoprotective effects, primarily attributed to anthraquinones (AQs). However, the molecular mechanisms of AQs biosynthesis have largely been hindered by insufficient genomic resources and functional genomics investigations. Here, we employed multi-omics approaches to address this knowledge gap. The high-quality T2T-level genome was constructed with a size of 2.68 Gb and a contig N50 of 233.65 Mb. Functional annotation revealed that the specific and expanded gene families of *R. tanguticum* are involved in efficient energy metabolism and secondary metabolite biosynthesis, providing a molecular basis for its stress adaptation and medicinal value. Integrated widely targeted metabolomics, spatial metabolomics and targeted quantification of AQs, we successfully elucidate the AQs spatiotemporal accumulation patterns. Seeds and leaves are key sites for the synthesis and transport of free AQs, while the root core serves as the primary location for the synthesis and accumulation of conjugated AQs. Through integrated genomic, transcriptomic, metabolomic and functional validation analyses, we preliminarily characterised the positive regulatory role of *RtPKSIII-8* in the synthesis of aloe-emodin and emodin, as well as the potential function of *RtUGT85AD11* in the biosynthesis of AQs and flavonoids. These findings provide essential genomic and functional data for deciphering AQs biosynthetic pathways and lay a theoretical foundation for the medicinal development and genetic improvement of *R. tanguticum*.

1 | Introduction

Rhubarb, the perennial herbaceous plants belonging to the genus Rheum in the family Polygonaceae, is recognised as one of the 'Four Diamonds' in traditional Chinese medicine (TCM) and has a long history of medicinal use (Zhuang et al. 2020). The dried roots and rhizomes are the primary medicinal parts, widely utilised in TCM clinical practice. It is one of the most frequently used herbs in both compound and single-ingredient formulations (Yang et al. 2024). Among the various species of rhubarb, *Rheum tanguticum* (*R. tanguticum*) stands out as a representative species due to its unique chemical composition and remarkable pharmacological effects, holding a significant position among medicinal rhubarbs (Dai et al. 2021; Li, Zhao, et al. 2024; Liang et al. 2022). Modern pharmacological studies have demonstrated that *R. tanguticum* exhibits a wide range of biological activities, including purgative, hepatoprotective, anti-inflammatory and antitumor effects (Chen, Chen, Tao, et al. 2020; Gui et al. 2007; Kon et al. 2014; Lande et al. 2023). These pharmacological properties are primarily attributed to its rich content of anthraquinones (AQs) (Wang et al. 2021). With the advancement of modern research, key AQs, such as emodin and rhein, have been demonstrated to possess significant antiviral properties, which are increasingly utilised in the treatment and prevention of viral epidemics, like COVID-19 (Chen et al. 2021; Xu et al. 2023). Simultaneously, AQs exhibit

Shuo Zhao and Yuan Zong have contributed equally to this work.

a range of biological activities, including antioxidant and anti-obesity effects, and can serve as natural food colourants (Shen et al. 2023; Tsagkaris et al. 2023; Wang et al. 2023). These properties have increasingly drawn attention to *R. tanguticum* in the development of dietary supplements and functional health products. Moreover, AQs are a significant source of yellow natural dyes and are widely utilised in industrial applications (Gürses et al. 2016).

AQs have been identified in 135 species of higher plants, with a predominant distribution in plant groups such as Polygonaceae, Fabaceae, Rhamnaceae and Rubiaceae (Duval et al. 2016). These AQs primarily exist in two forms: free AQs with lower polarity and bound AQs with increased polarity due to glycosylation (Wang et al. 2024). Despite the significance and vast application potential of AQs, their biosynthetic pathways remain largely unexplored. The biosynthetic pathways of AQs in plants are primarily categorised into two routes: the shikimic acid/o-succinylbenzoic acid pathway and the polyketide pathway (Mund and Čellárová 2023). Current research indicates that AQs are more likely synthesised via the polyketide pathway, which involves the participation of polyketide synthases (PKS) (Karppinen et al. 2008; Mizuuchi et al. 2009; Shkryl et al. 2011). Within the polyketide pathway, PKS III effectively catalyses the seven consecutive decarboxylative condensation reactions of malonyl-CoA to form an octaketide chain (Liu et al. 2020). This linear polyketide chain then undergoes cyclization and decarboxylation reactions to generate the core unit of polyketide compounds, thereby forming the backbone of AQs (Karppinen et al. 2008; Mizuuchi et al. 2009). Only the CHS-L gene of PKS III has been experimentally validated in the synthesis of emodin in *Senna tora* (L.) Roxb., providing definitive evidence for the catalytic role of PKS III in polyketide compound synthesis (Kang et al. 2020). Free AQs, limited by their poor solubility, frequently exhibit reduced medicinal efficacy in practical applications, which significantly impedes their widespread adoption and utilisation (Malik and Muller 2016). Studies have demonstrated that glycosylation modification can markedly enhance the solubility of free AQs, thereby amplifying their medicinal potency (Ghimire et al. 2015). In this process, UDP-glycosyltransferase (UGT) genes play a pivotal role, and multiple UGT genes have been identified to catalyse the conversion of emodin into emodin glycosides (Xie et al. 2014; Yin et al. 2020; Zhao et al. 2023). However, the genetic mechanisms underlying the high abundance of AQs in *R. tanguticum* remain elusive, with few key anthraquinone synthesis genes functionally validated.

The elucidation of biosynthetic pathways of key compounds hinges on the in-depth exploration of functional genes. The integration of high-quality genome sequencing with transcriptomics and metabolomics provides comprehensive insights into gene regulation, protein function, and metabolic composition, offering a robust foundation and an efficient research platform for unravelling the complexities of metabolic networks. For instance, the biosynthetic pathways of natural products such as leonurine in *Leonurus japonicus* (Li, Yan, et al. 2024), artemisinin in *Artemisia annua* (Liao et al. 2022) and triterpenoid saponins in *Panax notoginseng* (Yang et al. 2021) have been progressively deciphered through multi-omics integrative analysis. These studies not only demonstrated the powerful capability of multi-omics approaches in dissecting the intricate metabolic networks of medicinal plants but also provided critical theoretical support for the investigation of biosynthetic mechanisms of natural products and the exploitation of medicinal plant resources.

This study employed and integrated multi-omics strategies, combining T2T genome assembly, metabolomics profiling and transcriptome analysis to systematically elucidate AQ biosynthesis pathways in *R. tanguticum*. Our comprehensive analysis revealed distinct spatiotemporal accumulation patterns, identifying leaves and seeds as previously unrecognised key biosynthetic sites for AQ production. Spatial distribution analysis demonstrated compartmentalization of pharmacologically active compounds: free AQs predominantly localised in periderm and cortical tissues, while AQ glycosides were concentrated in vascular cylinder and pith regions. Through weighted gene co-expression network analysis, we identified critical metabolic modules and functionally characterised key biosynthetic genes. Notably, we established the enzymatic roles of four *RtPKSIII* (*RtPKS-8*, 12, 17, 20) and the *RtUGT85AD11* through transient or stable gene transformation systems. This research not only deepens our comprehension of the molecular pathways governing AQs metabolite biosynthesis but also validates key enzymes in *R. tanguticum*. It lays the groundwork for breeding high-quality rhubarb varieties and provides valuable genetic resources for AQs biosynthesis.

2 | Results

2.1 | Spatiotemporal AQs Accumulation Patterns of *R. tanguticum*

UPLC-MS/MS was employed for metabolite analysis, resulting in the identification of 1973 metabolites across four tissues, which were categorised into 13 classes (Figure 1a and Figure S1a). Flavonoids were the most abundant, followed by phenolic acids (Figure S1b). There was significant tissue-specific accumulation of metabolites in *R. tanguticum*. The leaves displayed significant enrichment of flavonoids; the seeds, in turn, were abundant in a diverse array of metabolites, encompassing flavonoids, quinones, phenolic acids and alkaloids (Figure 1b). AQs predominantly accumulate in the roots, a finding corroborated by the quantitative analysis of 11 AQs (Figure 1b,c). Furthermore, elevated levels of five free AQs were observed in both leaves and seeds, with the seeds exhibiting particularly high concentrations (Figure 1c). This suggested that seeds and leaves may also serve as crucial sites for the biosynthesis of free AQs.

The rhizome of *R. tanguticum* serves as its primary medicinal organ, exhibiting significant age-dependent accumulation patterns of secondary metabolites (Zhao, Xiong, et al. 2024). In this study, matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) was employed to systematically investigate the spatial distribution characteristics of secondary metabolites in roots of 1-year-old (1Y), 2-year-old (2Y) and 3-year-old (3Y) *R. tanguticum*. High-resolution mass spectrometry analysis identified 57 AQs, providing the first comprehensive elucidation of their spatial distribution patterns in root cross-sections. Results demonstrated distinct spatial heterogeneity in the distribution of different AQ structural types within

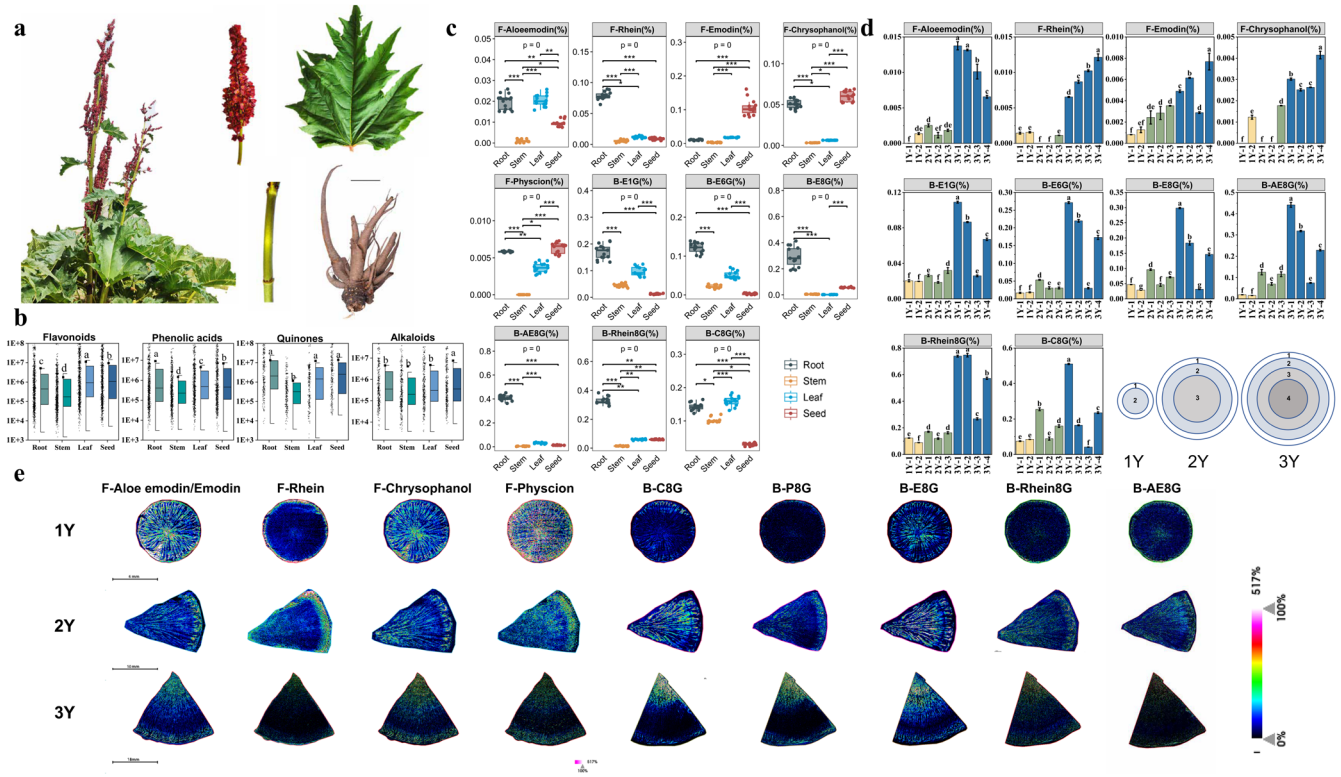

FIGURE 1 | Distribution and variation of metabolites in four tissues and aged roots of *R. tanguticum*. (a) Complete *R. tanguticum* plant; (b) Variability of the 4 metabolite classes in the *R. tanguticum* root, stem, leaf and seed; (c) Contents of 11 AQs of four tissues; (d) Spatial distribution of selected AQs in cross-sections of 1Y, 2Y and 3Y roots; (e) Quantitative analysis of 11 AQs in various layers of 1Y, 2Y and 3Y roots. E1G, emodin-1-O-β-D-glucoside; E6G, emodin-6-O-β-D-glucopyranoside; E8G, emodin-8-O-β-D-glucopyranoside; AE8G, aloe-emodin-8-O-β-D-glucopyranoside; rhein8G, rhein-8-O-glucopyranoside; C8G, chrysophanol-8-O-β-D-glucopyranoside. The number of asterisks indicates the *p*-value: $p < 0.05$, one asterisk (*); $p < 0.01$, two asterisks (**); $p < 0.001$, three asterisks (***).

root tissues (Figure 1d). Specifically, pharmacologically active free AQs (including rhein, emodin, chrysophanol and physcion) were predominantly localised in the periderm and cortical regions, while AQ glycosides were primarily distributed in the vascular cylinder and pith regions (Figure 1d). Stratified quantitative analysis demonstrated age-dependent accumulation of 11 key AQs with marked radial gradients (Figure 1e): free AQs decreased centripetally while bound AQs accumulated preferentially in the root core, consistent with spatial metabolomics findings. The spatiotemporal distribution patterns of AQs were closely associated with their biosynthetic pathways, transport mechanisms and physiological functions, providing crucial metabolomic evidence for understanding the accumulation dynamics of AQs.

2.2 | T2T-Levelled *R. tanguticum* Genome Construction

The karyotype analysis revealed that *R. tanguticum* possesses a diploid chromosome number of $2n = 22$ (Figure S2). Genomic characterisation through flow cytometry and k-mer analysis estimated a genome size of approximately 2.75 Gb (Table S1, Figures S3 and S4). To obtain a high quality *R. tanguticum* reference genome, we have performed the PacBio circular consensus

TABLE 1 | Statistics of *R. tanguticum* genome assembly information.

Item	Count
Assembly	
Total length (bp)	2 681 444 853
Scaffolds N50 (bp)	241 534 841
Number of contig	34
contig N50 (bp)	233 651 479
Gaps	23
Telomeres	22
Complete BUSCOs %	96.4%
Annotation	
Repcat sequenccs %	83.59
Number of predicted genes	70 967
Number of non-coding RNAs	24 999
Average gene length (bp)	1052.55
Average exon length (bp)	702.13
Average intron length (bp)	1002.41

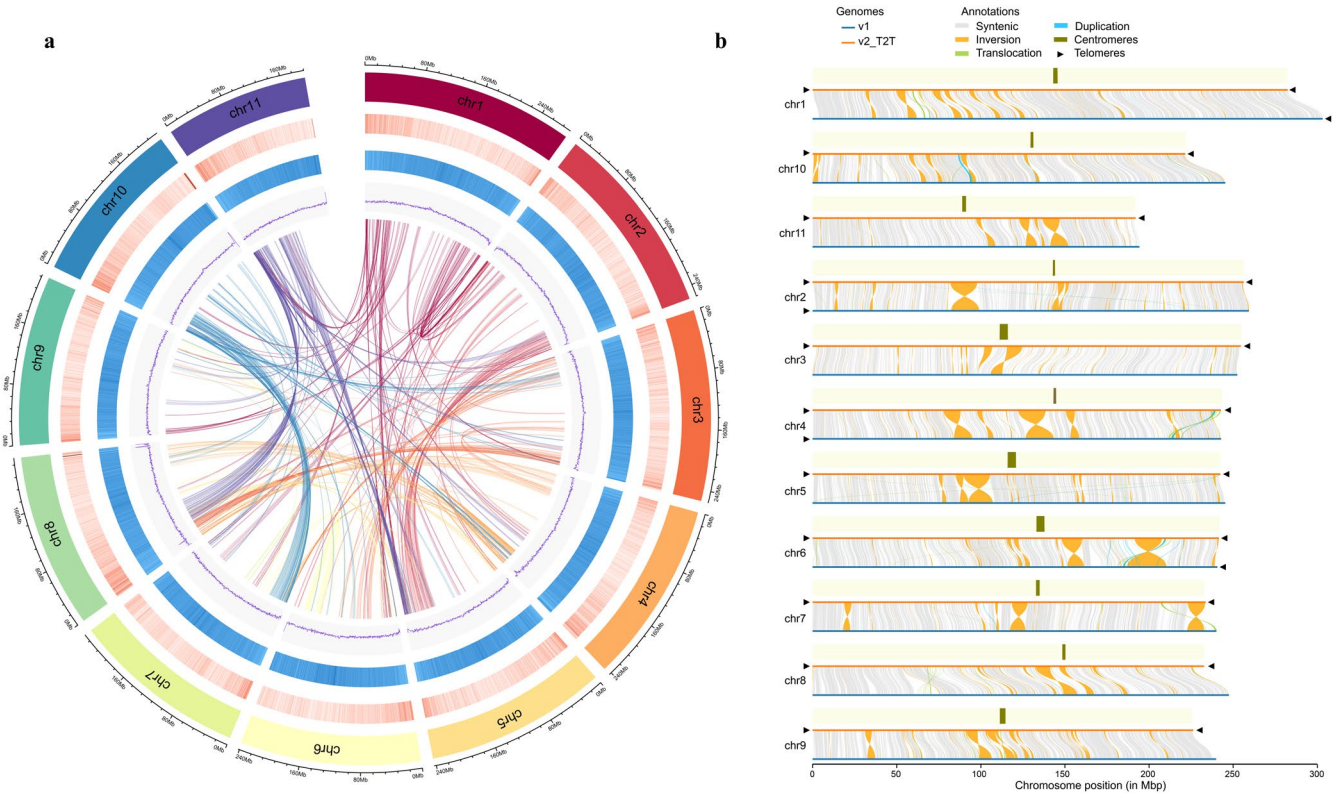

FIGURE 2 | High-quality genome assembly and genomic evolution analysis of *R. tanguticum*. (a) Distribution of *R. tanguticum* genomic features. From the outside in, there are circular representations of the 11 pseudomolecules, gene density (500 kb window), percentage of repeats (500 kb window), GC content (500 kb window), identified syntenic blocks; (b) Comparison of two chromosome-scale assemblies of *R. tanguticum*: The Nanopore-based V1 and our T2T-level V2, with V2 featuring full structural annotation of 11 centromeres and 22 telomeres.

sequencing (CCS) and generated 128.60 Gb CCS reads (46.76×) (Table S2). Then these reads were assembled into 985 contigs using hifiasm, resulting in a total length of 2.68 Gb with a contig N50 of 233.65 Mb (Table 1 and Table S4). Subsequently, a total of 400.67 Gb Hi-C data were used to successfully anchor 97.94% of contig sequences into 11 pseudochromosomes, with sizes ranging from 55.20 Mb to 282.51 Mb (Figure 2a and Figure S5, Tables S3 and S5). We further detected all the 11 centromeres and a complete set of 22 telomeres across all 11 chromosomes in our assembly, representing all chromosomes were telomere-to-telomere level (Figure 2b, Tables S6 and S7). 5 chromosomes fully met the T2T gap-free standard, whereas the remaining 6 chromosomes contained 23 gaps (Figure 2b, Table 1).

We firstly compared our assembly with the published one (Li, Wang, et al. 2023). Our assembly showed a great improvement of the contig N50 from 7.16 Mb to 233.65 Mb and the BUSCO score from 93.0% to 96.2% (Tables S8 and S9), which indicates a great improvement in continuity and completeness. Besides, the Illumina short reads were aligned to the assembled genome, resulting in a mapping rate of 99.98%, and about 85.05%–94.89% of RNA-seq reads from various tissues can be successfully aligned to the genome (Table S10). These assessments both indicate that the genome assembly of *R. tanguticum* had a high degree of accuracy and completeness.

Repeat elements were further comprehensively identified in our genome through a combination of ab initio and homology-based approaches. A total of 2241.53 Mb (83.59%) of repetitive sequences were successfully identified. Long terminal repeats (LTRs) were the predominant repeat elements, occupying 52.14% of the genome, among which the *Gypsy* and *Copia* superfamilies accounted for 43.88% and 7.42%, respectively (Table S11). Furthermore, a total of 70 967 protein-coding genes were predicted in the assembled genome (Table S12), with the average gene length of 6483 bp, coding DNA sequence (CDS) length of 6483 bp and exon number of 4.39 exons per gene (Table S13). The further functional annotation showed that 94.97% of genes were annotated in at least one database of Interpro, NR, Uniprot, GO and KEGG (Table S14). In addition, 24 999 noncoding RNAs in the *R. tanguticum* genome were identified, comprising 153 microRNAs (miRNA), 6707 transfer RNAs (tRNA), 16 777 ribosomal RNAs (rRNA) and 1362 small nuclear RNAs (snRNA) (Table S15).

2.3 | Genome Evolutionary Analysis of *R. tanguticum*

To unearth the evolutionary relics from polyploidization in *R. tanguticum*, we first analysed synonymous substitution rates (Ks) of intra-genomic collinear gene pairs within synteny blocks. Three similar Ks peaks were all observed in the five *Rheum* species, *R. tanguticum*, *R. palmatum*, *R. nobile*, *R. officinale* and *Rumex acetosa*, which suggests that two rounds of polyploidization events occurred after the γ event (whole-genome triplication, shared by all core eudicots) (Figure 3a). Notably, *R. officinale* contained an additional one Ks peak around 0.89,

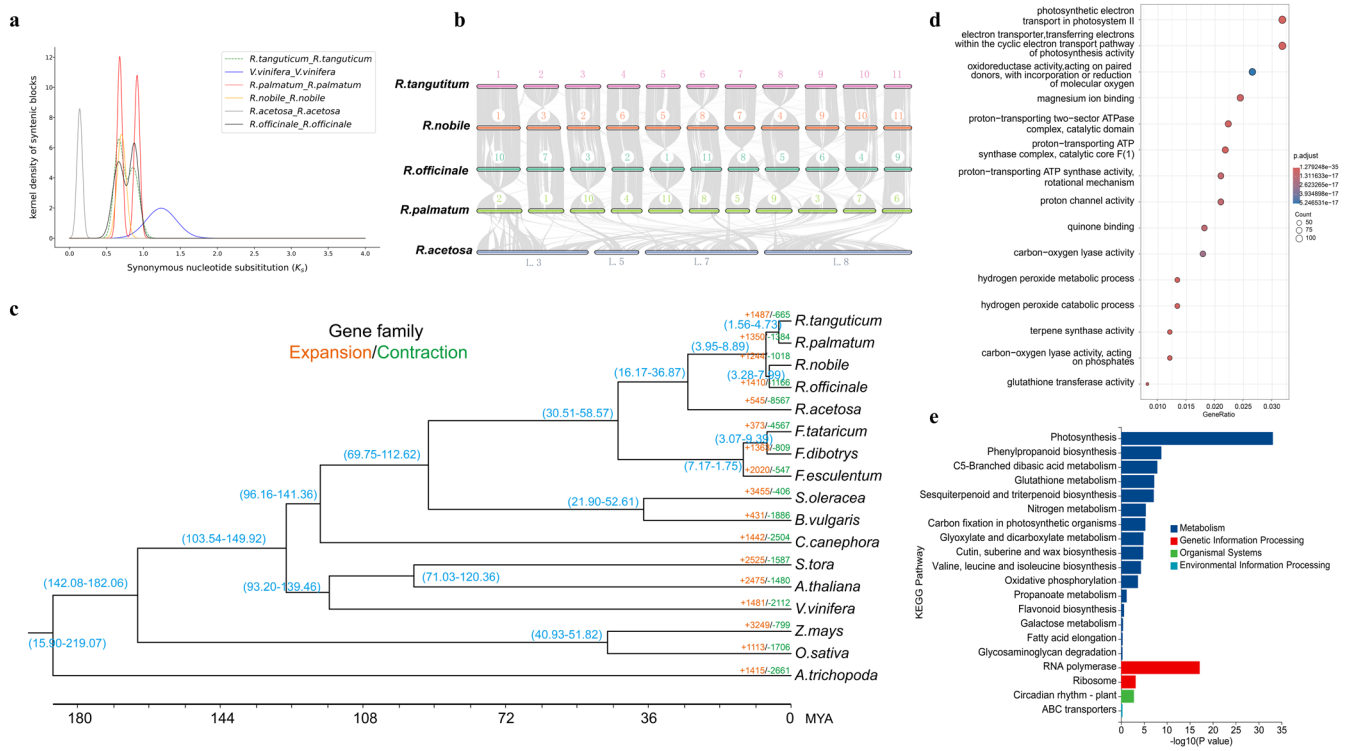

FIGURE 3 | Genomic evolution analysis of *R. tanguticum*. (a) Synonymous substitution rate (Ks) distributions of syntenic blocks for *R. tanguticum* with *V. vinifera*, *R. palmatum*, *R. officinale* (A), *R. nobile and R. acetosa*. (b) Syntenic analysis of *R. tanguticum*, *R. palmatum*, *R. officinale* (A), *R. nobile* and *R. acetosa*. (c) Divergence time estimating and clustering results of gene families of 17 species. The expansion and contraction of gene families are respectively indicated by red and green fonts next to the species names; (d) GO enrichment analysis of significantly expanded gene families in *R. tanguticum*; (e) KEGG enrichment analysis of significantly expanded gene families in *R. tanguticum*.

which was consistent with its a tetraploid species (Zhang, He, et al. 2024). The further inter-genomic synteny depth analysis showed a clear 4:1 ratio between *R. tanguticum* and *V. vinifera*, which proved that the recent polyploidization events in *R. tanguticum* were two rounds whole genome duplications (WGDs) (Figure S6). We also found these two rounds WGDs were shared by all other four Polygonaceae species, as the clear 1:1 chromosome-level intact collinear relationships were detected between homologous chromosomes with minimal interchromosomal rearrangements (Figure 3b).

To elucidate the phylogenetic relationship of the *R. tanguticum* genome, we conducted a comprehensive comparative genomic analysis involving 16 representative species (Table S16). A total of 231 265 gene families were constructed and 874 of them were single-copied genes that were used to reconstruct the phylogeny tree (Figure 3c). All four *Rheum* species clustered into a monophyletic clade, in which *R. tanguticum* was sister to *R. palmatum* with an approximate divergence time of 3.07 million years ago (MYA). Additionally, the divergence between the *Rheum* genus and the *Fagopyrum* genus within the same family occurred between 30.5 and 58.5 Mya.

A total of 9492 lineage specific gene families and 1487 expanded gene families were further identified in *R. tanguticum*. We found these unique genes were mainly involved in the photosynthetic system, such as proton-transporting ATP synthase complex, carbon fixation in photosynthetic organisms, oxidoreductase activity and so on (Figure S7), which may reflect the strong adaptive advantages in high altitude environments, such as tolerating high-light intensity environments and oxidative stress conditions. While for the expanded genes, they were also enriched in pathways related to photosynthesis and energy transfer related functions (Figure 3d). Additionally, these gene families are notably involved in various metabolic synthesis processes, such as quinone binding, terpene synthase activity, and glutathione transferase activity (Figure 3d). KEGG pathway enrichment analysis further confirmed that the expanded gene families primarily participate in the biosynthesis of multiple metabolites, including phenylpropanoid biosynthesis, C5-branched dibasic acid metabolism, glutathione metabolism, as well as sesquiterpenoid and triterpenoid biosynthesis (Figure 3e). These findings indicate that the specific and expanded gene families in *R. tanguticum* may provide a molecular-level explanation for the rich bioactive components and remarkable environmental adaptability.

2.4 | Transcriptomic Analysis and Integration With Metabolomic Profiling

To investigate the relationship between metabolite accumulation and gene expression in *R. tanguticum*, we selected 28 468 genes expressed in four tissues (with an average FPKM > 1) for weighted gene co-expression network analysis (WGCNA). By setting a soft threshold of 18 (Figure S8a), we identified 14 co-expression modules. Correlation analysis between the modules and 14 selected AQs revealed that the magenta and turquoise

modules exhibited correlation coefficients greater than 0.90 ($p < 0.01$) with the majority of AQs, suggesting a strong association with major AQs and indicating that these modules may contain key genes involved in AQs biosynthesis (Figure S8b). Furthermore, the MEpurple and MEbrown modules showed significant correlations with aloe-emodin and emodin, while the MEblue module was significantly correlated with aloe-emodin-8-O-glucoside and physcion-8-O-β-glucoside (Figure S8b). These findings further highlight the importance of these modules in the biosynthesis of AQs.

2.5 | Functional Characterisation of *RtPKSs* in AQs Biosynthesis

Through genome-wide homology comparison, we screened 12 enzyme genes that were potentially associated with the polyketide pathway and the shikimate/o-succinylbenzoate pathway in previous studies (Zhang, He, et al. 2024) (Figure 4a). The three official rhubarb species (*R. tanguticum*, *R. officinale* and *R. palmatum*) exhibited a significantly greater expansion in gene number compared to the non-official rhubarb species (*R. nobile*) across seven gene families (*SMK*, *EPSP*, *ICH*, *PHYLLO*, *MenE*, *MenB* and *PKC*) (Figure 4a). Similar to the metabolic and WGCNA results, most genes showed a high expression in leaves and seeds during AQ biosynthesis (Figure 4a).

Among the total pathway genes, one *CHS-L* gene in the *PKS III* family has been experimentally validated to participate in AQs polyketide synthesis in *S. tora* (Kang et al. 2020), which is the first step of the polyketide pathway (Figure 5a). So, we further carefully examined the *PKS III* family in *R. tanguticum* and identified a total of 24 PKS III genes (7 CHS and 17 *CHS-L*) (Figure 5a and Figure S9), which is greater than *R. nobile* but the same as *F. tataricum*. Among 17 *CHS-L*, *RtPKSIII-2*, *6*, *7*, *8*, *9*, *14*, *16* and *17* exhibited high expression levels in tissues other than stems (Figure 4a). Integrated with WGCNA analysis, *RtPKSIII-8* and *RtPKSIII-17* showed significant correlations with multiple AQs (Table S17). Although *RtPKSIII-12* belongs to the CHS class, it was grouped into the MEpurple module, which is significantly associated with aloe-emodin and emodin, thus being considered a potential functional candidate gene (Table S17). Additionally, *RtPKSIII-20* from the CHS family was selected as a control gene, hypothesized to be involved in flavonoid biosynthesis. qRT-PCR of 4 candidate genes showed *RtPKS III-8*, *17* and *20* had high expression in leaves and seeds, and *RtPKS III-12* in leaves (Figure S10a). When these 4 genes were transiently expressed in tobacco protoplasts, *RtPKS III-8* and *12* localised to the cytoplasm, whereas *RtPKS III-17* and *20* localised to both the cytoplasm and nucleus (Figure 5b). To further validate the functions of four candidate genes, we established an *Agrobacterium*-mediated transient gene expression system in *R. tanguticum*. Four candidate *RtPKSIII* genes were injected into young leaves, and the injection sites displayed patchy, intense fluorescence signals, indicating successful transient overexpression in the leaf tissues (Figure 5c and Figure S10b). HPLC analysis revealed that among the five free AQs, *RtPKSIII-8* and *RtPKSIII-12* significantly increased the content of aloe-emodin, with *RtPKSIII-8* showing particularly prominent effects. However, overexpression of *RtPKSIII-8* markedly reduced the content of rhein (Figure 5d). Furthermore, *RtPKSIII-8* and *RtPKSIII-17* increased the emodin content by 1.6-fold and 1.1-fold, respectively (Figure 5d).

2.6 | UGT Gene Family Identification and Functional Analysis

UGT-related genes play a crucial role in the glycosylation of free AQs, enhancing the solubility and bioavailability of glycosylated AQs. We identified 97 UGT genes (Figure S11) in the *R. tanguticum* genome and they could be clustered into 21 subfamilies, including 14 basic evolutionary groups (A-H, J-N and P) and 4 newly characterised groups (O, Q, R and OG) (Figure 6a). To further investigate the functions of UGT genes, we selected 20 *RtUGT* genes from groups D, G and L, which have been implicated in AQ glycosylation (Yin et al. 2020). Among these, 14 genes exhibited high expression levels in roots (group C2, C3, C7), particularly *RtUGT73BH7*, *RtUGT73BE21*, *RtUGT75R14*, *RtUGT75R15* and *RtUGT85AD11* (Figure 6b). Additionally, WGCNA analysis classified these genes into the MEpurple, MEbrown and MEblue modules, which are significantly associated with multiple bound AQs (Table S15). Notably, *RtUGT74CC3* was identified as a homologue of *FtUGT74L2* in *F. tataricum*, which is known to catalyse the glycosylation of emodin to produce emodin-8-O-glucoside (Yin et al. 2020), served as a control. To validate the functions of candidate genes, we transformed *RtUGT73BH7*, *RtUGT74CC3*, *RtUGT75R14*, *RtUGT75R15* and *RtUGT85AD11* into *G. inflata* hairy roots and successfully obtained positive transgenic lines (Figure 6c, Figure S14). HPLC analysis of six AQ glycosides revealed that, although the target compounds were not detected, new metabolite peaks were observed in all transgenic lines compared to the negative controls, particularly in the *RtUGT85AD11* transformants, indicating that the candidate *RtUGT* genes can recognise emodin as a substrate and generate novel metabolites (Figure 6d). To further explore metabolic changes, we performed widely targeted metabolomics using UPLC-MS/MS on GFP control and *RtUGT85AD11*-overexpressing hairy roots. The results showed that the contents of key AQs, including emodin, emodin-6-O-β-D-glucoside, and emodin-8-O-glucoside, were significantly reduced in *RtUGT85AD11*-overexpressing lines, while the rhein content increased by 7-fold compared to the GFP control (Figure 6e). Simultaneously, the levels of glycosylated derivatives of kaempferol, apigenin, naringenin, acacetin and chrysoeriol were significantly increased (Table S15), suggesting a potential regulatory role of *RtUGT85AD11* in the flavonoid metabolic pathway.

2.7 | MYB Transcription Factors Involved in AQs Regulation

Research in *R. palmatum* suggested that MYB transcription factors may be key regulatory components in the anthraquinone biosynthetic pathway (Zhao, Yan, et al. 2024). Cis-element analysis of the *RtPKSIII* gene promoter revealed a significant enrichment of MYB transcription factor binding sites. For instance, six MYB binding sites were identified in the *RtPKSIII-8* promoter, suggesting that MYB may regulate the expression of this structural gene and thereby promote anthraquinone biosynthesis (Figure S9b). Genome-wide identification systematically uncovered 242 *RtMYB* family

members, which were phylogenetically classified into R2R3-MYB, 3R-MYB and other subfamilies using *A. thaliana* MYB proteins as references. Notably, the R2R3-MYB subfamily exhibited the largest membership (118 members), implying its potential dominance in metabolic regulation (Figure S16). Tissue-specific transcriptomic screening identified 46 *RtMYB* candidate genes with elevated expression in roots, leaves and seeds (Figure S17). Co-expression network analysis demonstrated significant positive correlations between *RtMYB* genes and *RtPKSIII* genes, especially *RtMYB29*, *RtMYB111*, *RtMYB206* (Figure S18). Protein–protein interaction network constructed via STRING database revealed CHS and UGT

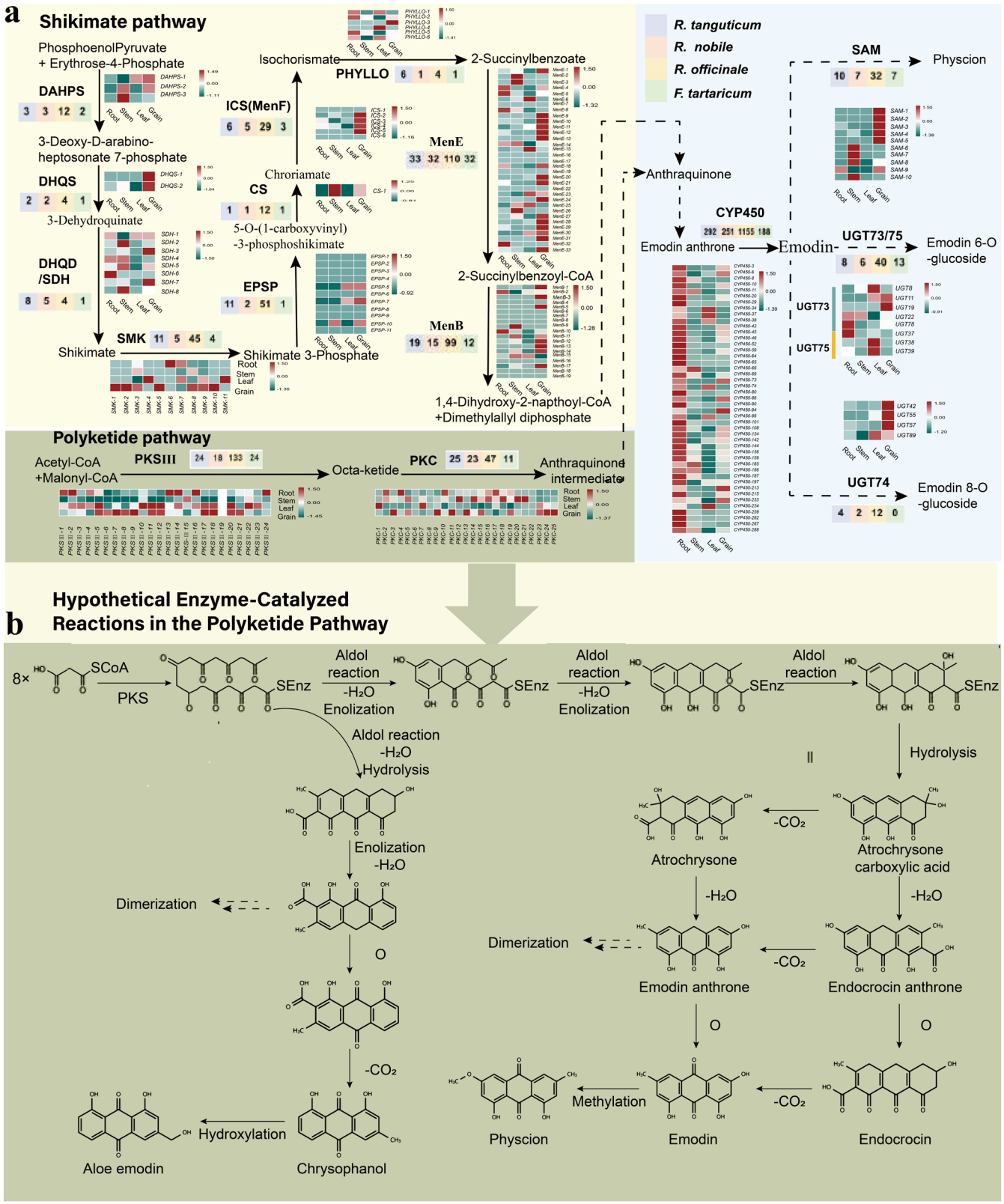

FIGURE 4 | Legend on next page.

FIGURE 4 | Genome-wide screening of key enzymes for AQs biosynthesis and functional validation of *PKSIII* involvement in *R. tanguticum*. (a) Genes involved in the shikimate and polyketide pathways, the dashed lines indicate multiple steps. The yellow and green backgrounds represent the shikimate pathway and the polyketide pathway, respectively. The purple, pink, yellow, and green frames in the figure display the number of genes in *R. tanguticum*, *R. nobile*, *R. officinale* and *F. tataricum*, respectively. The heatmap shows the expression levels of genes in different tissues; (b) Proposed anthraquinone biosynthesis via the polyketide pathway, involving polyketide synthases (PKS). PKS acts as an octaketide synthase responsible for the assembly and cyclization of the polyketide chain, producing the anthranoid scaffold precursors, such as atrochrysone carboxylic acid and its derivative endocrocin anthrone. PKS lacks the catalytic activity required to convert these precursors into final anthraquinones like chrysophanol, rhein and emodin. Further steps, including decarboxylation and oxidation, are necessary for the synthesis of chrysophanol, emodin and related anthraquinones.

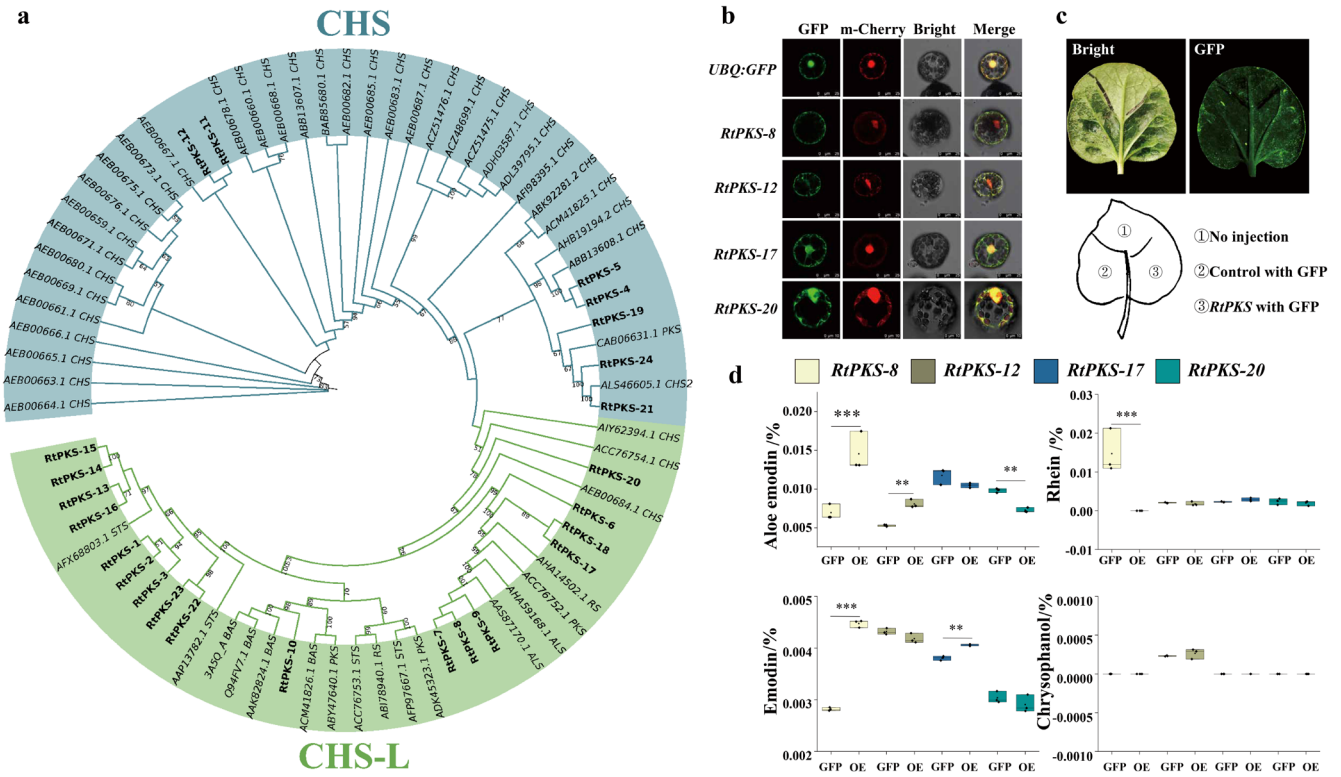

FIGURE 5 | Functional validation of *PKSIII* involvement in *R. tanguticum*. (a) Phylogenetic tree of *R. tanguticum* and 50 *PKSIII* protein family members from the Polygonaceae; (b) The subcellular localization of four *RtPKSIII* genes; (c) Detection of transient expression effects of candidate *RtPKSIII* genes; (d) Determination of the content of five free AQs in leaves with transient overexpression of four *RtPKSIII* genes.

as central hub nodes, with 9 *RtMYB* TFs directly interacting with both enzymatic genes. Remarkably, *RtanMYB111* exhibited simultaneous interactions with CHS and four UGT genes (Figure S19), indicating its potential role in multi-target metabolic coordination. *RtMYB040*, *RtMYB055* and *RtMYB183* displayed comparable regulatory potential. qRT-PCR validation of six core *RtMYB* genes confirmed high consistency between expression patterns and transcriptomic data across tissues (Figure S20). WGCNA analysis clustered three R2R3-MYB members (*RtMYB082*, *RtMYB114* and *RtMYB176*) into the MEturquoise module, which showed significant positive correlation with anthraquinone content. Notably, *RtMYB100* displayed high sequence homology with *RpMYB81*, a characterised AQs biosynthesis regulator in *R. palmatum*, which could be markedly induced by MeJA, accompanied by a significant increase in AQs content (Zhao, Yan, et al. 2024). These cumulative findings suggest that MYB genes are likely to regulate the expression of key enzyme genes, such as PKSIII, thereby coordinating AQs biosynthesis in *R. tanguticum*. However, definitive evidence requires further validation through molecular biology approaches.

3 | Discussion

The high quality T2T-level reference genome is fundamental for deciphering biosynthetic pathways (Cheng et al. 2025). Here we integrated high depth HiFi and Hi-C sequencing data to construct the T2T genome of *R. tanguticum*, providing a reliable genetic resource for functional genomics research. Our assembly showed a very high Contig N50 value of 233.65 Mb, representing a 32.6-fold improvement compared to the previously published Nanopore-based assembly (Li, Wang, et al. 2023) and also remarkable enhancements in continuity and completeness compared to those in three recently published *Rheum* species genomes (Li, Niu, et al. 2023; Zhang, He, et al. 2024; Zhang,

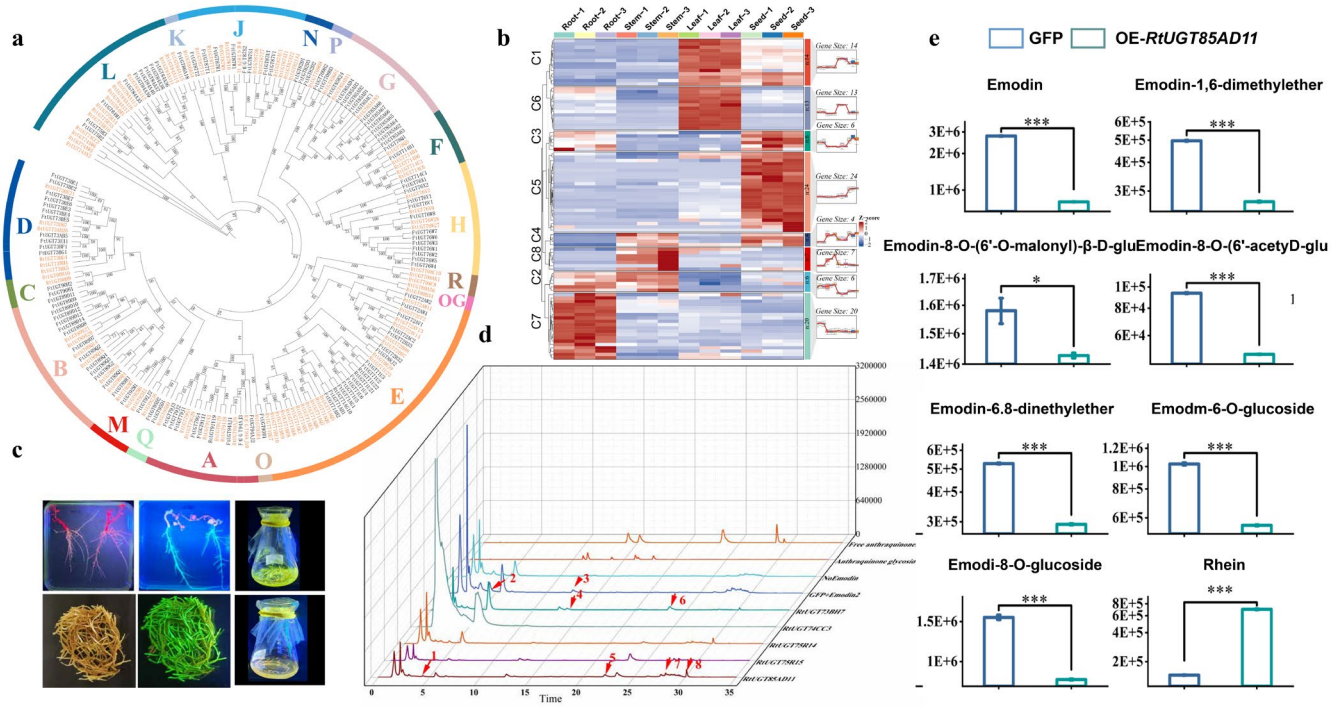

FIGURE 6 | Genome-wide identification of UGT gene family in *R. tanguticum* and preliminary functional validation of key *RtUGT* genes. (a) Phylogenetic tree of 97 *R. tanguticum RtUGT* and 106 *F. tataricum FtUGT* protein family members; (b) Expression analysis of *RtUGT* genes in four tissues; (c) Transgenic hairy roots of *G. inflata* with candidate *RtUGT* genes; (d) Determination of AQs in transgenic hairy root of *G. inflata* with 5 *RtUGT* genes; (e) Comparative analysis of AQs contents between GFP-negative control and *RtUGT85AD11*-overexpressing *G. inflata* hairy roots.

Zhou, et al. 2024) (Table S14). Using this genome as reference, we could integrate the multi-omics (transcriptome and metabolism) to reveal the key genes and regulators in AQs biosynthesis pathway.

3.1 | Metabolic Diversity and Spatiotemporal Distribution of AQs in *R. tanguticum*: Biosynthetic and Pharmacological Insights

In this study, we conducted a comprehensive metabolomic analysis of different tissues of *R. tanguticum* using UPLC-MS/MS, identifying nearly 2000 metabolites across 13 categories. Among them, the high levels of free AQs were detected in leaves and seeds, suggesting that these tissues may also play significant roles in the biosynthesis and storage of these compounds. This finding challenges the traditional view that AQs are primarily synthesised and stored in roots and opens new avenues for exploring the biosynthetic pathways and physiological functions of these metabolites in different plant tissues. Using MALDI-MSI and quantitative detection of 11 AQs, we revealed the heterogeneous spatial distribution of AQs within the roots. Free AQs were predominantly enriched in the periderm and cortex, while AQ glycosides exhibited high concentrations in the vascular cylinder and pith regions. The same phenomenon has also been observed in *R. palmatum* (Shen et al. 2022) and *Centaurea* sp. (Bouafia et al. 2020). This distribution pattern may reflect functional differences of these compounds in various root tissues and indicate complex interactions among biosynthesis, transport and storage mechanisms. Our study provides a comprehensive overview of the metabolic diversity and tissue-specific accumulation patterns in *R. tanguticum*, emphasising the importance of considering both temporal and spatial factors when studying the biosynthesis and accumulation of secondary metabolites. The insights gained from this research not only enhance our understanding of AQs metabolic pathways in *R. tanguticum* but also lay a foundation for further exploration of the pharmacological potential of its metabolites. Future studies may focus on elucidating the biosynthetic pathways and regulatory mechanisms underlying the observed metabolite distribution patterns, as well as exploring the potential applications of these compounds in medicine and agriculture.

3.2 | Molecular Insights Into AQs Biosynthesis in *R. tanguticum*

This study aimed to elucidate the molecular mechanisms underlying AQs biosynthesis in *R. tanguticum* and identify key genes involved in the synthesis process. Unlike well-studied secondary metabolites such as flavonoids and terpenoids, whose biosynthetic pathways are relatively well-characterised, the AQs biosynthetic pathway remains largely unexplored (Mund and Čellárová 2023). To address this, we integrated metabolomic and transcriptomic data to conduct in-depth analyses of the *PKSIII* and *UGT*, which are key gene families in the polyketide pathway and AQ glycosylation modification, respectively (Kang et al. 2020; Yin et al. 2020; Zhao et al. 2023). Our findings preliminarily characterised the role of *RtPKSIII-8* in the AQs biosynthetic pathway, demonstrating its ability to promote the synthesis of aloe-emodin and emodin while inhibiting the production of rhein. This phenomenon may result from the

regulation of key enzymes in the AQs biosynthetic pathway by *RtPKSIII-8*, which directs precursor flux preferentially toward the synthesis of aloe-emodin and emodin, thereby limiting rhein production. Similarly in flavonoid metabolism, the anr mutation reduces proanthocyanidins and via precursor reallocation promotes anthocyanin and flavonol (quercetin) accumulation (Xu et al. 2024). Additionally, this metabolic flux redistribution suggests the possibility of dynamic interconversion among these compounds.

Furthermore, our study revealed that *RtUGT85AD11*, although not directly involved in AQ glycosylation, significantly enhances rhein synthesis and catalyses the glycosylation of multiple flavonoid compounds. This substrate specificity may arise from the preferential recognition and catalysis of flavonoid glycosylation by *RtUGT85AD11*, while remaining inactive toward AQs. However, the precise functions of the *RtPKS* and *RtUGT* genes and their roles in anthraquinone metabolism require further validation through stable genetic transformation in *R. tanguticum* and in vitro enzymatic assays. This study not only provides valuable genetic resources for AQs biosynthesis but also lays a theoretical foundation for further elucidating the AQs biosynthetic pathway.

4 | Conclusions

In this study, we successfully assembled a chromosome-level genome of *R. tanguticum*, with a size of 2.68 Gb, a contig N50 of 233.65 Mb, and a chromosome anchoring rate of 97.94%, demonstrating significant improvements in continuity and completeness. Functional annotation revealed that the specific and expanded gene families in *R. tanguticum* were primarily involved in photosynthesis, energy transfer and the biosynthesis of secondary metabolites, highlighting its efficient energy metabolism and potential for secondary metabolite production. These characteristics likely provided a molecular basis for its survival under stress conditions and its medicinal value. Furthermore, the burst of LTR retrotransposons may represent a key adaptive strategy for *Rheum* species to thrive in extreme environments. Through comprehensive analysis of the PKSIII, UGT and MYB gene families, we preliminarily characterised the functions of *RtPKSIII-8* and *RtUGT85AD11*, offering novel insights into the biosynthetic mechanisms of AQs. These findings not only enrich the genomic resources of *R. tanguticum* but also lay a solid foundation for further elucidating its secondary metabolic pathways and environmental adaptation mechanisms.

5 | Materials and Methods

5.1 | Plant Materials

In this study, five-year-old genetically stable *R. tanguticum* plants cultivated at the Germplasm Resource Nursery in Huzhu County, Haidong City, Qinghai Province, China (N: 36°53′37″; E: 102°07′05″) were selected as experimental materials. To ensure genetic consistency, all samples were collected from the same plant. Fresh healthy leaves were harvested for genomic DNA extraction and whole-genome sequencing, while roots, stems and seeds from the same plant were collected for transcriptome (RNA-seq) analysis. For developmental stage-specific root studies, fresh root samples were systematically collected from one (1Y), two- and three-year-old (1Y, 2Y and 3Y) plants. A standardised sampling protocol was implemented, with transverse sections uniformly taken at 1/4 of the distance from the root tip to ensure positional consistency. Immediately after collection, all samples were flash-frozen in liquid nitrogen and stored at −80°C in ultra-low temperature freezers for long-term preservation. To ensure data reliability, three independent replicates were used for transcriptome analysis and five for metabolome analysis.

5.2 | Metabolite Extraction, Identification, and Quantification

After vacuum freeze-drying the samples using a freeze dryer (Scientz-100F), they were ground into a powder for 1.5 min at a frequency of 30 Hz using a grinder (MM 400, Retsch). A precise amount of 50 mg of the sample powder was weighed and dissolved in 1.20 mL of 70% methanol solution pre-cooled to −20°C. The samples were mixed six times for 30 s each, with a 30-min interval between mixes. Following centrifugation (12 000 rpm/min for 3 min), the supernatant was collected and filtered through a 0.22 μm microporous membrane. The filtered solution was then placed in an injection vial for subsequent UPLC-MS/MS analysis. The widely targeted metabolomics analysis was conducted using the UPLC-ESI-MS/MS system (ExionLC AD, MS, Applied Biosystems 6500 Q TRAP) and the tandem mass spectrometry system from MetWare (Wuhan, China).

5.3 | Analysis of the AQs Content

The AQs content was quantified using an HPLC system (Agilent 1260 Infinity II), with chromatographic-grade methanol, acetonitrile, phosphoric acid and formic acid sourced from Shandong Yu Wang Group. A total of 11 AQs were determined, comprising 5 free AQs, namely, aloe-emodin (481–72-1), rhein (478–43-3), emodin (518–82-1), chrysophanol (481–74-3), physcion (521–61-9) and 6 bound AQs, namely, aloe-emodin-8-O-β-D-glucopyranoside (33037–46-6, AE8G), rhein-8-O-glucopyranoside (34298–86-7, rhein8G), chrysophanol-8-O-beta-D-glucopyranoside (13241–28-6, C8G), emodin-1-O-β-D-glucoside (38840–23-2, E1G), emodin-6-O-β-D-glucopyranoside (34298–85-6, E6G) and emodin-8-O-β-D-glucopyranoside (23313–21-5, E8G), all procured from Sigma (USA). The sample processing, extraction, and HPLC determination methods for five free AQs and six bound AQs were adapted from previous studies on *R. tanguticum* (Zhao, Xiong, et al. 2024).

5.4 | Spatial Metabolomics Based on MALDI-MSI

Transverse sections of 1.0–1.5 cm were excised from the middle part of the main roots of 1Y, 2Y and 3Y *R. tanguticum* plants. The sections were immediately immersed in liquid nitrogen for 20 s and then transferred to −80°C for storage. After equilibrating at −20°C in a cryostat for 1 h, the samples were sectioned at a thickness of 30 μm using a Leica CM1950 cryostat. The sections were transferred to pre-cooled ITO glass slides using

a pre-chilled brush. The tissue sections were thawed to transparency by placing the slides against the back of the hand, and the moisture was evaporated by rubbing the back of the slides, causing the tissue to turn from transparent to white. After sectioning, the ITO slides were vacuum-dried for 30 min, and only intact transverse sections were selected for further analysis. The sections were mounted onto glass slides using double-sided tape for subsequent MS-MSI analysis.

A 15 mg/mL solution of DHB (2,5-dihydroxybenzoic acid) was prepared in 90:10 acetonitrile: water and uniformly sprayed onto the ITO slides using a TM-Sprayer matrix applicator. The instrument parameters were set as follows: temperature at 60°C, flow rate at 0.12 mL/min, pressure at 10 psi, and 30 spray cycles with a drying time of 6 s between cycles. The matrix-coated ITO slides were placed on the target plate of the mass spectrometer. The tissue regions were selected using DataImaging (Bruker) software, and the imaging resolution was set to 50 μm (minimum grid size of 50 μm × 50 μm) with a mass range of 50–1300 Da. Under consistent laser energy, the laser beam irradiated the tissue regions through a raster pattern, continuously scanning the samples. The tissue molecules were desorbed and ionised upon interaction with the matrix, and the released ions were detected by the mass spectrometer to obtain the mass-to-charge ratio (m/z) and peak intensity raw data for each pixel. The raw data were imported into SCiLS Lab software for root mean square normalisation, generating spatial relative intensity information for different m/z values, which were then converted into pixel values for imaging heatmaps.

5.5 | Genome Sequencing

A modified cetyltrimethylammonium bromide (CTAB) method was used to obtain high quality and purified genomic DNA samples, then a SMRT cell sequencing library containing about 15–20 kb cut fragment was constructed and sequenced using PacBio Sequel II sequencing platform. In total, 2225.3 Gb subreads and 128.60 Gb clean data were obtained. Simultaneously, a modified CTAB method was used to obtain high quality and purified genomic DNA samples, then a second generation library that included 150 bp paired-end Illumina libraries (Illumina, San Diego, USA) with 350 bp insertions was constructed, yielding 83.69 Gb clean data, which covered the genome ~31.2×. The resulting short reads were used for genome survey analysis. A Hi-C library was established using the germinated young leaves of the seeds harvested from the same plant, and the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA) was used for sequencing. After the sequencing data of Hi-C was filtered, a total of 377.46 Gb (coverage of ~140.8×) clean data was obtained. The above sequencing was performed in the Wuhan Benagen Technology Co. Ltd. (Wuhan, China).

5.6 | Genome Size Estimation

Genome size estimation was first determined using flow cytometry. 0.2 g of fresh young leaves of maize and *R. tanguticum* were placed in 500 μL Nuclei Extraction buffer respectively, and chopped with a sharp blade then filtered through a 50 μm filter after 60 s. Followed by the addition of 2000 μL of staining buffer with RNase for 30 min in the dark. Nuclei suspension was analysed by Flow Space Flow Cytometer (Sysmex Partec, Muenster, Germany) and the corresponding FloMax software. Meanwhile, a k-mer ($k = 19$) analysis-based approach together with Illumina paired-end short reads was used for a second determination of the genome size. The Jellyfish v2.1.4 was used for calculating a 19-mer frequency (Marçais and Kingsford 2011), and then the genome size and heterozygosity were calculated based on the k-mer frequency results using GenomeScope (http://qb.cshl.edu/genomescope/) (Vurture et al. 2017).

5.7 | RNA Sequencing

RNA sequencing (RNA-seq) was conducted to investigate gene expression in various tissues of *R. tanguticum*, including roots, stems, leaves and seeds. Total RNA was extracted from fresh tissues using the RNAprep Pure Plant Kit (TIANGEN, China), and RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). RNA libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina, USA), followed by paired-end sequencing on the Illumina NovaSeq 6000 platform (Illumina, USA). The raw sequencing reads were evaluated for quality using fastp v0.21.0 (Chen et al. 2018). Clean reads were then aligned to the *R. tanguticum* reference genome using HISAT2 (Kim et al. 2015). Gene expression levels were quantified using featureCounts v2.0.1 (Liao et al. 2014), and differential expression analysis was performed with DESeq2 v1.22.1 (Love et al. 2014). Functional enrichment analysis of differentially expressed genes (DEGs) was carried out using Gene Ontology (GO) and KEGG pathway analysis (Ashburner et al. 2000; Kanehisa et al. 2008). To validate the RNA-seq results, a subset of DEGs was analysed by quantitative real-time PCR (qRT-PCR). First-strand cDNA was synthesised from 1 ng of RNA using the MonScript RTIII All-in-one Mix with dsDNase (Monad, MR05101). The qRT-PCR reaction mixture was prepared using TB Green Premix Ex Taq II (Tli RNaseH Plus) and the fluorescence quantitative PCR reaction was conducted on the ABI 7500 Fast Real-Time PCR System. The relative expression level of the target gene was calculated using the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen 2001) primer sequences for all qRT-PCR assays in this study were detailed in Table S21.

5.8 | Genome Assembly and Quality Assessment

SMRTlink v8.0 was used to perform quality control statistics on the raw PacBio single-molecule sequencing data to obtain valid data. Then, CCS reads were generated by ccs v6.2.0 software filtering, and seqkit v0.16.1 software was used to count the CCS reads information; finally, 128.60 Gb of valid data was obtained. On this basis, genome assembly was performed using the software Hifiasm v0.15.3-r339 (Cheng et al. 2021) with default parameters. At the same time, for highly heterozygous species, the software Purge_haplotigs v1.0.4 (Roach et al. 2018) needs to be used to de-heterozygize the error-corrected genome to obtain a draft genome.

The Hi-C data were used to anchor the genome assembly to the chromosome level. The raw reads were filtered using the software fastp v0.21.0 to obtain high-quality clean data. Use the

software HICUP v0.8.0 (Wingett et al. 2015) to align the clean data to the reference genome to obtain valid HIC data. The draft contig sequences were clustered into chromosome groups by agglomerative hierarchical clustering algorithm using ALLHIC v0.9.8 (Zhang et al. 2019), and the order and orientation of the contigs within the chromosome group were determined. Then, the interaction relationship between the contigs was converted into a specified binary file (i.e., hic file) by 3D-DNA v180419 (Dudchenko et al. 2017) and Jucier v1.6 (Durand et al. 2016), and then review assembly was performed by Juciebox v1.11.08 (Robinson et al. 2018).

The Illumina clean reads were mapped to the reference genome using BWA (Li 2013). The mapping rate and coverage were used to evaluate the integrity of the assembly and the uniformity of sequencing coverage. The software BUSCO v.4.1.2 (Simao et al. 2015) was further used to assess the completeness of the genome assembly.

5.9 | Genome Annotation

Integrate homologous gene comparison, transcriptome data, and de novo prediction methods not reliant on known gene information to forecast protein-coding genes within the rhubarb genome. The de novo prediction approach employs Hidden Markov Models (HMM) to train on known gene sets, thereby constructing species-specific gene models and identifying genes within the genome. Predictions were conducted using three software programs: Augustus v3.3.2 (Stanke et al. 2008), Genscan v1.0 (Burge 1998), and GlimmerHMM v3.0.4 (Delcher et al. 2007). For homologous gene comparison, Exonerate v.2.4.0 (Slater and Birney 2005) aligns the proteins of five related species (*R. acetosa*, *F. esculentum*, *F. tataricum*, *B. vulgaris*, *C. quinoa*) with the genome. In transcriptome data-based prediction, RNA-seq data was reconstructed into transcripts using StringTie v2.1.4 (Kovaka et al. 2019), and TransDecoder v5.1.0 (https://github.com/TransDecoder/TransDecoder) predicts coding frames. The MAKER v2.31.10 (Holt and Yandell 2011) integrates the gene sets predicted by various methods. By filtering, a non-redundant and more comprehensive gene set was formed. The prediction quality was evaluated using the BUSCO v.4.1.2.

Gene function annotation was conducted primarily through two approaches: First, sequence similarity searches were performed using the diamond Blastp v2.0.11.149 (Buchfink et al. 2015) to align the protein sequences from the gene set against the Uniprot (https://www.uniprot.org/) and NR (ftp://ftp.ncbi.nih.gov/blast/db) protein databases, as well as the KEGG (http://www.genome.jp/kegg/) metabolic pathway database. Second, motif similarity searches were conducted using InterProScan v5.52–86 (Blum et al. 2021) to compare resources such as CDD, Gene3D, Panther, Hamap, Phobius and Pfam within the InterPro sub-databases.

Leveraging the structural features of tRNA, tRNA sequences within the genome were detected using tRNAscan-SE v1.23 (Chan et al. 2021). rRNA identification was conducted using BLAST v2.9.0+. Additionally, snRNA and miRNA were recognised using INFERNAL v1.1.2 (Nawrocki and Eddy 2013), which was based on the Rfam database.

5.10 | Analysis of Genome Evolution

To explore the evolutionary history of the *R. tanguticum* genome, 16 plant species with high-quality genome assemblies, representing different evolutionary branches, were chosen (*R. officinale* (A), *R. nobile*, *R. palmatum*, *R. acetosa*, *F. esculentum*, *F. tataricum*, *F. dibotrys*, *B. vulgaris*, *S. oleracea*, *S. tora*, *A. thaliana*, *V. vinifera*, *O. sativa*, *A. trichopoda*, *Z. mays*, *C. canephora*). Gene family clustering was conducted using OrthoFinder v2.3.12 (Emms and Kelly 2019) based on the amino acid sequences of these species, and alignments were performed using Blastp v2.6.0 (Camacho et al. 2009). A comparative analysis was further conducted among *F. tataricum*, *R. officinale* (A), *R. nobile*, *R. palmatum* and *R. tanguticum* to elucidate the shared and unique gene families across these species. The R package clusterProfiler (Wu et al. 2021) was employed to conduct GO (Ashburner et al. 2000) and KEGG (Kanehisa et al. 2008) analyses on shared, unique, expanded and contracted gene families, respectively.

A total of 46 single-copy gene families shared among 17 species were selected to construct a phylogenetic tree. The protein sequences of each single-copy gene family were aligned using MUSCLE v3.8.31 (Edgar 2004), and the alignment results were filtered using the trimAl v1.4.rev22 (Capella-Gutierrez et al. 2009). The filtered alignment results were then merged and concatenated to form a super gene. The RAxML v8.2.10 (Stamatakis 2014) was utilised, employing the PROTGAMMAWAG model, to construct a maximum likelihood (ML) species phylogenetic tree based on the super gene sequence. Based on the topology of the phylogenetic tree and the fossil time node table, the divergence times of the selected species were estimated using the MCMCTree of the PAML v4.9 (Yang 2007) with the following parameters: n sample = 3 000 000; burnin = 8 000 000; seq type = 0; model = 4.

Select *F. tataricum*, *R. officinale* (A), *R. nobile*, *R. palmatum* and *R. tanguticum* for whole-genome duplication (WGD) analysis. Use blast v2.6.0+ (Camacho et al. 2009) to align protein sequences of different species, then use MCScanX (Wang et al. 2012) to analyse genomic collinearity blocks; finally, use the yn00 module in PAML v4.9 (Yang 2007) to calculate the synonymous mutation frequency (Ks), nonsynonymous mutation frequency (Ka) and the ratio of nonsynonymous to synonymous mutation rates (Ka/Ks) for collinear gene pairs, and use ggplot2 v2.2.1 to plot the density.

5.11 | Characterisation of Gene Families in Polyketide Production

To identify the gene families relevant to this study, we first collected the corresponding sequence data from publicly accessible databases such as NCBI and UniProt, with a particular focus on species known for AQs production. The key enzyme gene families studied, along with their respective Pfam IDs, are as follows: DAHPS (Pfam ID: PF02656), DHQS (Pfam ID: PF01092), SDH (Pfam ID: PF00248), SMK (Pfam ID: PF01764), EPSP (Pfam ID: PF01767), CS (Pfam ID: PF01772), ICS (Pfam ID: PF02826), PHYLLO (Pfam ID: PF05482), MenE (Pfam ID: PF04839), MenB (Pfam ID: PF06145), PKS (Pfam ID: PF02801), UGT (Pfam ID: PF00201), CYP450 (Pfam ID: PF00067), MYB

(Pfam ID: PF00249) and MYC (Pfam ID: PF00646). The gene families involved in this study were identified through BLASTp searches, and the results were further refined using Hidden Markov Models (HMMs) specific to the respective gene family domains (Eddy 2011). Multiple sequence alignments were performed using Clustal Omega (Sievers et al. 2011), followed by phylogenetic analysis with MEGA X (Kumar et al. 2018) to explore the evolutionary relationships among the genes. Functional domains were characterised using the Pfam (Mistry et al. 2021) and SMART (Letunic et al. 2012) databases to classify the genes into their respective families. Visualising the gene structures of gene family members in TBtools (Chen, Chen, Zhang, et al. 2020) using genomic annotation GFF files. Finally, comparative genomics was employed to assess the evolution of these gene families across different species. This comprehensive bioinformatics approach provides valuable insights into the diversity of these gene families and their potential roles in metabolic pathways.

5.12 | Subcellular Localization Analyzes

For subcellular localization analysis, key target genes were fused with GFP (green fluorescent protein), and recombinant plasmids containing the fusion constructs were transformed into *Agrobacterium tumefaciens* GV3101 strain. The transformed *Agrobacterium* was then infiltrated into tobacco leaves using the Agrobacterium-mediated transformation method (Yoo et al. 2007). Tobacco protoplasts were isolated from fresh leaves by enzymatically digesting the cell walls with cellulase and macerozyme. The protoplasts were then transformed with plasmids encoding the fluorescent protein-tagged target proteins via PEG-mediated transformation. After 12–24 h of expression, protoplasts were analysed using confocal laser scanning microscopy (CLSM) to observe the localization of the proteins. Fluorescence patterns were compared with organelle-specific markers for cytoplasm and nuclei. Colocalization was further confirmed by co-transfecting protoplasts with plasmids expressing organellar markers. Image analysis was performed to quantify the overlap of fluorescence signals, enabling precise identification of the protein localization in subcellular compartments.

5.13 | Transient Expression of Four *RtPKSIII* in *R. tanguticum*

To investigate the function and localization of four *RtPKSIII* genes and their potential impact on AQs biosynthesis in *R. tanguticum*, recombinant plasmids carrying the *RtPKSIII* genes fused with GFP were constructed and introduced into *Agrobacterium tumefaciens* GV3101. The transformed *Agrobacterium* was used for Agrobacterium-mediated infiltration into *R. tanguticum* leaves. After 48–72 h of expression, the subcellular localization of the RtPKSIII-GFP fusion proteins was observed using a Leica SP8 confocal microscope, with fluorescence detected at 488 nm excitation and 507–532 nm emission. The transient expression system also allowed for the analysis of AQs production by evaluating the changes in metabolic profiles. Primer sequences for constructing the *RtPKSIII* overexpression vector were detailed in Tables S19 and S20.

5.14 | Stable Transformation of *RtUGT* Into *G. inflata* Hairy Roots

Stable genetic transformation of *RtUGT* into *G. inflata* hairy roots was performed to investigate its impact on metabolite synthesis. Recombinant plasmids carrying the *RtUGT* genes were constructed and introduced into *Agrobacterium tumefaciens* strain K599, followed by co-cultivation with surface-sterilised *G. inflata* hypocotyls to induce hairy root formation. Transgenic roots were confirmed through molecular analysis, including PCR amplification of the *RtUGT* and rolB genes using gene-specific primers, followed by RT-qPCR to quantify gene expression levels. Transgenic roots were cultured in liquid MS medium for 4–8 weeks, and metabolite profiling was conducted using HPLC to assess the effect of *RtUGT* expression on AQs biosynthesis. Statistical analysis of metabolite data was performed using ANOVA and the Tuk**ey** HSD test to compare differences in metabolite accumulation between transgenic and wild-type roots.

Author Contributions

G.Z., S.Z. and Y.Z. designed and supervised the work; T.W. and J.L. performed the research; Z.Y, and F.Z. analysed and investigated the data; S.Z. and G.Z. wrote and revised the paper.

Acknowledgements

This work was supported by the Major Science and Technology Special Project of Qinghai Province (2023-SF-A5), Central Forestry Reform and Development Fund 2022 and Central Forestry Reform and Development Fund 2023. We are particularly grateful to Dr. Jiangyi Zeng of the South China Botanical Garden, Chinese Academy of Sciences, for her valuable assistance and dedicated support.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in PRJCA037598National Genomics Data Center at https://ngdc.cncb.ac.cn/gsub/submit/bioproject/subPRO055376.

References

Ashburner, M., C. A. Ball, J. A. Blake, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25: 25–29.

Blum, M., H. Y. Chang, S. Chuguransky, et al. 2021. "The InterPro Protein Families and Domains Database: 20 Years on." *Nucleic Acids Research* 49: D344–D354.

Bouafia, M., A. Benarfa, N. Gourine, and M. Yousfi. 2020. "Seasonal Variation of Fatty Acid Composition, Tocopherol Content and Antioxidant Activity of Lipid Extracts From Centaurea sp." *Food Bioscience* 37: 100728.

Buchfink, B., C. Xie, and D. H. Huson. 2015. "Fast and Sensitive Protein Alignment Using DIAMOND." *Nature Methods* 12: 59–60.

Burge, C. B. 1998. "Modeling Dependencies in Pre-mRNA Splicing Signals." In *Computational Methods in Molecular Biology*, 129–164. Elsevier.

Camacho, C., G. Coulouris, V. Avagyan, et al. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10: 421.

Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon. 2009. "trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses." *Bioinformatics* 25: 1972–1973.

Chan, P. P., B. Y. Lin, A. J. Mak, and T. M. Lowe. 2021. "tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes." *Nucleic Acids Research* 49: 9077–9096.

Chen, C., H. Chen, Y. Zhang, et al. 2020. "TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data." *Molecular Plant* 13: 1194–1202.

Chen, J. Q., Y. Y. Chen, H. J. Tao, et al. 2020. "An Integrated Metabolomics Strategy to Reveal Dose-Effect Relationship and Therapeutic Mechanisms of Different Efficacy of Rhubarb in Constipation Rats." *Journal of Pharmaceutical and Biomedical Analysis* 177: 112837.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. "Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor." *Bioinformatics* 34: i884–i890.

Chen, X., Y. Wu, C. Chen, et al. 2021. "Identifying Potential Anti-COVID-19 Pharmacological Components of Traditional Chinese Medicine Lianhuaqingwen Capsule Based on Human Exposure and ACE2 Biochromatography Screening." *Acta Pharmaceutica Sinica B* 11: 222–236.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. "Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs With Hifiasm." *Nature Methods* 18: 170–175.

Cheng, L. T., Z. L. Wang, Q. H. Zhu, M. Ye, and C. Y. Ye. 2025. "A Long Road Ahead to Reliable and Complete Medicinal Plant Genomes." *Nature Communications* 16: 2150.

Dai, L. X., X. Miao, X. R. Yang, et al. 2021. "High Value-Added Application of Two Renewable Sources as Healthy Food: The Nutritional Properties, Chemical Compositions, Antioxidant, and Antiinflammatory Activities of the Stalks of *Rheum officinale* Baill. and *Rheum tanguticum* Maxim. Ex Regel." *Frontiers in Nutrition* 8: 770264.

Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg. 2007. "Identifying Bacterial Genes and Endosymbiont DNA With Glimmer." *Bioinformatics* 23: 673–679.

Dudchenko, O., S. S. Batra, A. D. Omer, et al. 2017. "De Novo Assembly of the *Aedes aegypti* Genome Using Hi-C Yields Chromosome-Length Scaffolds." *Science* 356: 92–95.

Durand, N. C., M. S. Shamim, I. Machol, et al. 2016. "Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments." *Cell Systems* 3: 95–98.

Duval, J., V. Pecher, M. Poujol, and E. Lesellier. 2016. "Research Advances for the Extraction, Analysis and Uses of Anthraquinones: A Review." *Industrial Crops and Products* 94: 812–833.

Eddy, S. R. 2011. "Accelerated Profile HMM Searches." *PLoS Computational Biology* 7: e1002195.

Edgar, R. C. 2004. "MUSCLE: Multiple Sequence Alignment With High Accuracy and High Throughput." *Nucleic Acids Research* 32: 1792–1797.

Emms, D. M., and S. Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20: 238.

Ghimire, G. P., N. Koirala, R. P. Pandey, H. J. Jung, and J. K. Sohng. 2015. "Modification of Emodin and Aloe-Emodin by Glycosylation in Engineered *Escherihia coli*." *World Journal of Microbiology and Biotechnology* 31: 611–619.

Gui, M., Y. F. Zhang, Z. Y. Xiao, et al. 2007. "Inhibitory Effect of Emodin on Tissue Inhibitor of Metalloproteinases-1 (TIMP-1) Expression in Rat Hepatic Stellate Cells." *Digestive Diseases and Sciences* 52: 200–207.

Gürses, A., M. Açıkyıldız, K. Güneş, and M. S. Gürses. 2016. "Classification of Dye and Pigments." In *Dyes and Pigments*, 31–45. Springer International Publishing.

Holt, C., and M. Yandell. 2011. "MAKER2: An Annotation Pipeline and Genome-Database Management Tool for Second-Generation Genome Projects." *BMC Bioinformatics* 12: 491.

Kanehisa, M., M. Araki, S. Goto, et al. 2008. "KEGG for Linking Genomes to Life and the Environment." *Nucleic Acids Research* 36: D480–D484.

Kang, S. H., R. P. Pandey, C. M. Lee, et al. 2020. "Genome-Enabled Discovery of Anthraquinone Biosynthesis in *Senna tora*." *Nature Communications* 11: 5875.

Karppinen, K., J. Hokkanen, S. Mattila, P. Neubauer, and A. Hohtola. 2008. "Octaketide-Producing Type III Polyketide Synthase From *Hypericum perforatum* Is Expressed in Dark Glands Accumulating Hypericins." *FEBS Journal* 275: 4329–4342.

Kim, D., B. Langmead, and S. L. Salzberg. 2015. "HISAT: A Fast Spliced Aligner With Low Memory Requirements." *Nature Methods* 12: 357–360.

Kon, R., N. Ikarashi, C. Nagoya, et al. 2014. "Rheinanthrone, a Metabolite of Sennoside A, Triggers Macrophage Activation to Decrease Aquaporin-3 Expression in the Colon, Causing the Laxative Effect of Rhubarb Extract." *Journal of Ethnopharmacology* 152: 190–200.

Kovaka, S., A. V. Zimin, G. M. Pertea, R. Razaghi, S. L. Salzberg, and M. Pertea. 2019. "Transcriptome Assembly From Long-Read RNA-Seq Alignments With StringTie2." *Genome Biology* 20: 278.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis Across Computing Platforms." *Molecular Biology and Evolution* 35: 1547–1549.

Lande, P. S., V. S. Adhao, J. P. Ambhore, K. P. Gaikwad, C. S. Chandak, and L. P. Joge. 2023. "Anticancer Action of Naturally Occurring Emodin for the Controlling of Cervical Cancer." *Exploration of Targeted Anti-Tumor Therapy* 4: 690–698.

Letunic, I., T. Doerks, and P. Bork. 2012. "SMART 7: Recent Updates to the Protein Domain Annotation Resource." *Nucleic Acids Research* 40: D302–D305.

Li, H. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs With BWA-MEM. arXiv: Genomics."

Li, P., M.-X. Yan, P. Liu, et al. 2024. "Multiomics Analyses of Two Leonurus Species Illuminate Leonurine Biosynthesis and Its Evolution." *Molecular Plant* 17: 158–177.

Li, Y., Z. Niu, M. Zhu, et al. 2023. "Multi-Omics Data Provide Insight Into the Adaptation of the Glasshouse Plant *Rheum nobile* to the Alpine Subnival Zone." *Communications Biology* 6: 906.

Li, Y., Z. Wang, M. Zhu, et al. 2023. "A Chromosome-Scale Rhubarb (*Rheum tanguticum*) Genome Assembly Provides Insights Into the Evolution of Anthraquinone Biosynthesis." *Communications Biology* 6: 867.

Li, Y., Y. Zhao, X. Niu, et al. 2024. "Distinguishment of Different Varieties of Rhubarb Based on UPLC Fingerprints and Chemometrics." *Journal of Pharmaceutical and Biomedical Analysis* 241: 116003.

Liang, W., Z. Weimei, Y. Chen, et al. 2022. "Quality Evaluation of Different Varieties of Rhubarb Based on Multicomponents and Bioactivity: Application to Quality Control in the Production of Rhubarb Decoction Pieces." *Biomedical Chromatography* 36: e5368.

Liao, B., X. Shen, L. Xiang, et al. 2022. "Allele-Aware Chromosome-Level Genome Assembly of *Artemisia annua* Reveals the Correlation Between ADS Expansion and Artemisinin Yield." *Molecular Plant* 15: 1310–1328.

Liao, Y., G. K. Smyth, and W. Shi. 2014. "featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics* 30: 923–930.

Liu, A.-J., X. Ke-Bo, L. Yu-Yu, et al. 2020. "Biosynthesis of Polyketides by Two Type III Polyketide Synthases From *Aloe barbadensis*." *Journal of Asian Natural Products Research* 22: 434–443.

Livak, K. J., and T. D. Schmittgen. 2001. "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method." *Methods* 25: 402–408.

Love, M. I., W. Huber, and S. Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With DESeq2." *Genome Biology* 15: 550.

Malik, E. M., and C. E. Muller. 2016. "Anthraquinones as Pharmacological Tools and Drugs." *Medical Research Reviews* 36: 705–748.

Marçais, G., and C. Kingsford. 2011. "A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-Mers." *Bioinformatics* 27: 764–770.

Mistry, J., S. Chuguransky, L. Williams, et al. 2021. "Pfam: The Protein Families Database in 2021." *Nucleic Acids Research* 49: D412–D419.

Mizuuchi, Y., S. P. Shi, K. Wanibuchi, et al. 2009. "Novel Type III Polyketide Synthases From *Aloe arborescens*." *FEBS Journal* 276: 2391–2401.

Mund, N. K., and E. Čellárová. 2023. "Recent Advances in the Identification of Biosynthetic Genes and Gene Clusters of the Polyketide-Derived Pathways for Anthraquinone Biosynthesis and Biotechnological Applications." *Biotechnology Advances* 63: 108104.

Nawrocki, E. P., and S. R. Eddy. 2013. "Infernal 1.1: 100-Fold Faster RNA Homology Searches." *Bioinformatics* 29: 2933–2935.

Roach, M. J., S. A. Schmidt, and A. R. Borneman. 2018. "Purge Haplotigs: Allelic Contig Reassignment for Third-Gen Diploid Genome Assemblies." *BMC Bioinformatics* 19: 460.

Robinson, J. T., D. Turner, N. C. Durand, H. Thorvaldsdottir, J. P. Mesirov, and E. L. Aiden. 2018. "Juicebox.Js Provides a Cloud-Based Visualization System for Hi-C Data." *Cell Systems* 6: 256–258.e1.

Shen, N., Y. Chen, F. Guo, Y. Cui, M. Wei, and Z. Cheng. 2022. "Influence of Different Drying Methods Coupled With Different Process Modes on Physicochemical Qualities and Anthraquinones Contents of *Rheum palmatum* L." *LWT* 170: 114021.

Shen, N., J. Ren, Y. Liu, et al. 2023. "Natural Edible Pigments: A Comprehensive Review of Resource, Chemical Classification, Biosynthesis Pathway, Separated Methods and Application." *Food Chemistry* 403: 134422.

Shkryl, Y. N., G. N. Veremeichik, V. P. Bulgakov, and Y. N. Zhuravlev. 2011. "Induction of Anthraquinone Biosynthesis in *Rubia cordifolia* Cells by Heterologous Expression of a Calcium-Dependent Protein Kinase Gene." *Biotechnology and Bioengineering* 108: 1734–1738.

Sievers, F., A. Wilm, D. Dineen, et al. 2011. "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Molecular Systems Biology* 7: 539.

Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31: 3210–3212.

Slater, G. S., and E. Birney. 2005. "Automated Generation of Heuristics for Biological Sequence Comparison." *BMC Bioinformatics* 6: 31.

Stamatakis, A. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30: 1312–1313.

Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler. 2008. "Using Native and Syntenically Mapped cDNA Alignments to Improve de Novo Gene Finding." *Bioinformatics* 24: 637–644.

Tsagkaris, A. S., A. Louckova, J. Polak, and J. Hajslova. 2023. "Identifying Edible Plants With High Anti-Obesity Potential: In Vitro Inhibitory Effect Against Pancreatic Lipase Followed by Bioactive Compound Metabolomic Screening." *Food Bioscience* 56: 103453.

Vurture, G. W., F. J. Sedlazeck, M. Nattestad, et al. 2017. "GenomeScope: Fast Reference-Free Genome Profiling From Short Reads." *Bioinformatics* 33: 2202–2204.

Wang, D., X. H. Wang, X. Yu, et al. 2021. "Pharmacokinetics of Anthraquinones From Medicinal Plants." *Frontiers in Pharmacology* 12: 638993.

Wang, P., J. Wei, X. Hua, et al. 2024. "Plant Anthraquinones: Classification, Distribution, Biosynthesis, and Regulation." *Journal of Cellular Physiology* 239: e31063.

Wang, Y., C. Liu, S. Tang, J. Tian, Y. Wang, and Y. Yang. 2023. "Thermodynamics, Kinetics and Structure-Activity Relationship of Hydroxyanthraquinones Scavenging Free Radicals." *Food Bioscience* 53: 102705.

Wang, Y., H. Tang, J. D. Debarry, et al. 2012. "MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity." *Nucleic Acids Research* 40: e49.

Wingett, S., P. Ewels, M. Furlan-Magaril, et al. 2015. "HiCUP: Pipeline for Mapping and Processing Hi-C Data." *F1000Research* 4: 1310.

Wu, T., E. Hu, S. Xu, et al. 2021. "clusterProfiler 4.0: A Universal Enrichment Tool for Interpreting Omics Data." *Innovation (Cambridge)* 2: 100141.

Xie, K., R. Chen, J. Li, et al. 2014. "Exploring the Catalytic Promiscuity of a New Glycosyltransferase From *Carthamus tinctorius*." *Organic Letters* 16: 4874–4877.

Xu, H., S. Li, J. Liu, et al. 2023. "Bioactive Compounds From Huashi Baidu Decoction Possess Both Antiviral and Anti-Inflammatory Effects Against COVID-19." *Proceedings of the National Academy of Sciences of the United States of America* 120: e2301775120.

Xu, P., M. Li, C. Ma, et al. 2024. "Loss-Of-Function Mutation in Anthocyanidin Reductase Activates the Anthocyanin Synthesis Pathway in Strawberry." *Molecular Horticulture* 4: 33.

Yang, X., L. Dai, F. Yan, et al. 2024. "The Phytochemistry and Pharmacology of Three Rheum Species: A Comprehensive Review With Future Perspectives." *Phytomedicine* 131: 155772.

Yang, Z. 2007. "PAML 4: Phylogenetic Analysis by Maximum Likelihood." *Molecular Biology and Evolution* 24: 1586–1591.

Yang, Z., G. Liu, G. Zhang, et al. 2021. "The Chromosome-Scale High-Quality Genome Assembly of *Panax notoginseng* Provides Insight Into Dencichine Biosynthesis." *Plant Biotechnology Journal* 19: 869–871.

Yin, Q., X. Han, Z. Han, et al. 2020. "Genome-Wide Analyses Reveals a Glucosyltransferase Involved in Rutin and Emodin Glucoside Biosynthesis in *Tartary buckwheat*." *Food Chemistry* 318: 126478.

Yoo, S. D., Y. H. Cho, and J. Sheen. 2007. "Arabidopsis Mesophyll Protoplasts: A Versatile Cell System for Transient Gene Expression Analysis." *Nature Protocols* 2: 1565–1572.

Zhang, H., Q. He, L. Xing, et al. 2024. "The Haplotype-Resolved Genome Assembly of Autotetraploid Rhubarb *Rheum officinale* Provides Insights Into Its Genome Evolution and Massive Accumulation of Anthraquinones." *Plant Communications* 5: 100677.

Zhang, T., L. Zhou, Y. Pu, et al. 2024. "A Chromosome-Level Genome Reveals Genome Evolution and Molecular Basis of Anthraquinone Biosynthesis in *Rheum palmatum*." *BMC Plant Biology* 24: 261.

Zhang, X., S. Zhang, Q. Zhao, R. Ming, and H. Tang. 2019. "Assembly of Allele-Aware, Chromosomal-Scale Autopolyploid Genomes Based on Hi-C Data." *Nature Plants* 5: 833–845.

Zhao, H., Y. He, K. Zhang, et al. 2023. "Rewiring of the Seed Metabolome During *Tartary buckwheat* Domestication." *Plant Biotechnology Journal* 21: 150–164.

Zhao, S., F. Xiong, J. Li, et al. 2024. "Metabolomic Characteristics and Anthraquinones Accumulation Patterns of *Rhubarb* in Different Tissues and Roots From Different Developmental Stages." *Food Bioscience* 62: 105426.

Zhao, X., F. Yan, Y. M. Li, et al. 2024. "Comparative Transcriptome Analysis and Identification of Candidate R2R3-MYB Genes Involved in Anthraquinone Biosynthesis in *Rheum palmatum* L." *Chinese Medicine* 19: 23.

Zhuang, T., X. Gu, N. Zhou, L. Ding, L. Yang, and M. Zhou. 2020. "Hepatoprotection and Hepatotoxicity of Chinese Herb Rhubarb (Dahuang): How to Properly Control the 'General (Jiang Jun)' in Chinese Medical Herb." *Biomedicine & Pharmacotherapy* 127: 110224.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Classification (a) and relative content (b) of metabolites in four tissues of *R. tanguticum*. **Figure S2:** Karyotype analysis of *R. tanguticum*. **Figure S3:** Estimate of the genome size of *R. tanguticum* by flow cytometry. (A) Fluorescence histogram of the standard *Zea mays* (the genome size is 2655.3 M (1C DNA), and the peak is at 129.90). (B) Fluorescence histogram of the *R. tanguticum* (the peak is at 135.17). **Figure S4:** Estimate of the genome size and complexity of *R. tanguticum* by *K*-mers method. 19-mers depth distribution of 83.69 Gb raw sequence data from five libraries with insert size of 350 bp. **Figure S5:** The Hi-C map of *R. tanguticum* genome assembly. **Figure S6:** (b) Syntenic analysis of *R. tanguticum*, *V. vinifera* and *A. thaliana*. **Figure S7:** GO (a) and KEGG (b) enrichment analyses revealed the association between the unique gene families in the *R. tanguticum* genome and diverse biological functions. **Figure S8:** Gene networks underlying AQs production in *R. tanguticum*. (a) Gene expression clustering tree and module division of WGCNA; (b) Correlation heatmap between gene modules and 14 key AQs. **Figure S9:** Analysis of chromosomal localization, tandem duplication events, and cis-acting regulatory elements of the PKSIII gene family in *R. tanguticum*. **Figure S10:** Quantitative real-time PCR (qRT-PCR) analysis of four candidate genes in the PKSIII gene family of *R. tanguticum* regulating AQs biosynthesis: (a) Expression levels in four tissues (root, stem, leaf, seed); (b) Validation via transient overexpression in leaves. **Figure S11:** Tandem duplication events and segmental duplication events of *RtUGT* gene family members in *R. tanguticum*. **Figure S12:** The relative expression of 10 *RtUGT* genes in different tissues (root, stem, leaf, seed). **Figure S13:** The subcellular localization of the 5 candidate *RtUGT* genes. **Figure S14:** qRT-PCR analysis of *RtUGT* genes in transgenic hairy root of *G. inflata*. **Figure S15:** Comparison of differentially expressed flavonoid metabolite contents in transgenic *G. inflata* hairy roots overexpressing *RtUGT85AD11*. **Figure S16:** Phylogenetic tree of the R2R3-MYB subfamily in the MYB gene families of *R. tanguticum* and *A. thaliana*. **Figure S17:** Correlation analysis between the R2R3-MYB subfamily members and the PKS III gene family members in *R. tanguticum*. **Figure S18:** Correlation analysis between the R2R3-MYB subfamily members and the PKS III gene family members in *R. tanguticum*. **Figure S19:** The protein–protein interaction simulation of *RtMYB* TFs with the PKS and UGT gene families. **Figure S20:** qRT-PCR analysis of four candidate genes in the PKSIII gene family of *R. tanguticum* regulating AQs biosynthesis Validation via transient overexpression in leaves. **Table S1:** pbi70607-sup-0002-TablesS1-S21.xlsx.