Multi-omics reveals divergent regulation of anthocyanin glycosylation and gibberellin biosynthesis underlying leaf variegation in Saxifraga stolonifera

Jianhang Zhang

52181300002@stu.ecnu.edu.cn

Shaoxing University

jiecheng Li

East China Normal University

Hai Xing

Shaoxing Cash Crop Technology Extension Center

Feng Zhang

Nanchang University

Chao Wang

Shaoxing University

Fangping Tang

Shaoxing University

Hongqing Li

East China Normal University

Shuai Liao

Key Laboratory of National Forestry and Grassland Administration on East China Plant Conservation and Utilization, Shanghai Chenshan Botanical Garden

---

Research Article

Keywords: leaf variegation, anatomy, Saxifraga stolonifera, gibberellins, transcriptomics

Posted Date: May 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9460163/v1

Additional Declarations: No competing interests reported.

Abstract

Saxifraga stolonifera Curtis is a shade-tolerant, variegated-leaf herb with ornamental and medicinal value that naturally displays three stable leaf phenotypes: green, white-variegated, and purple-variegated. These phenotypes simultaneously exhibit both air space-type and pigment-type variegations. However, the key pathways and regulatory networks underlying their formation remain largely unknown. Here, we generated a 2.01 Gb high-quality chromosome-scale genome assembly (2n = 36). Comparative genomic analyses revealed a recent whole-genome duplication specific to *S. stolonifera* following the shared γ-triplication event, along with 155 expanded gene families enriched in flavonoid and phenylpropanoid biosynthesis pathways. Metabolomic profiling identified 58 anthocyanin-related compounds, among which 10 key pigments including cyanidin-3-O-glucoside, peonidin-3-O-galactoside, and quercetin-3-O-glucoside, were responsible for the purple patches. Their accumulation corresponded with the up-regulation of the anthocyanidin 3-O-glucosyltransferase gene *SsBZ1* (*Sst12G009310*, *Sst14G006050*). Hormone and transcriptome analyses showed that white-variegated leaves accumulate high levels of bioactive gibberellins GA4 and GA7, driven by increased expression of gibberellin 3β-dioxygenase *SsGA3ox* (*Sst05G017510*). Weighted gene co-expression network analysis (WGCNA) further identified a GA-associated module (salmon) with *SsGA3ox* as the hub gene, promoting cell expansion and generating air spaces between epidermal and palisade tissues along the veins. Collectively, our high-quality genome, metabolome, and transcriptome resources demonstrate that *SsGA3ox*-mediated gibberellin biosynthesis drives air space-type leaf variegation, whereas *SsBZ1*-controlled anthocyanin glycosylation produces pigment-type leaf variegation. These findings provide an integrative omics framework for dissecting leaf variegation mechanisms in *Saxifraga* and other ornamental plants.

Introduction

Variegated leaf plants represent a distinct group of colored-leaf species characterized by stable, heritable patterns formed by differentially pigmented leaf regions. These features make them valuable model systems for studying plant chromatology, mechanisms of adaptation to low-light environments, and chloroplast development[1, 2]. These variegated leaves can be classified into five types: chlorophyll type, air space type, epidermis type, pigment type, and appendages type[3]. Among these, the air space type and pigment type are the most prevalent and often co-occur on the same leaf, in species such as *Actinidia lanceolata*, *Erythronium dens-canis*, and *Saxifraga stolonifera*[3, 4]. Variegated leaves contribute to multiple adaptive functions, including acclimation to shade conditions, enhanced cold tolerance, defense against herbivores, and facilitation of self-pollination[2, 5–8]. In addition, variegated plants possess significant ornamental value due to their visually striking foliage.

Although naturally occurring leaf variegation has been studied for decades, previous research has largely focused on morphological classification and ecological functions; genomics, transcriptomics, and other omics studies have been limited by comparison. To date, transcriptomic studies of variegated leaves have primarily focused on the types of variegations associated with chlorophyll deficiency or red-leaf phenotypes. These investigations have elucidated the molecular mechanisms underlying chlorophyll

degradation or anthocyanin accumulation in variegated tissues, as demonstrated in *Ananas comosus* var. *bracteatus*[9], *Epipremnum aureum* 'Marble Queen'[10], *Clivia miniata* var. *variegata*[11], *Alternanthera bettzickiana*[12], *Camellia sinensis* 'Zijuan'[13], and *Perilla frutescens*[14]. In transcriptome analyses of air space-type variegation in *Trifolium pratense* and *Primulina pungentisepala*, key pathways related to photosynthesis, redox regulation, cell-wall modification, and nitrogen metabolism have been examined[2, 15]. Anthocyanins, a major class of water-soluble flavonoid pigments, play a critical role in determining the coloration of plant organs and tissues[16, 18]. Under high-light conditions, the flavonoid biosynthesis pathway is activated, resulting in the accumulation of photoprotective and antioxidative flavonoids, particularly flavonols and anthocyanins, in leaf tissues[19]. The spatially restricted accumulation of anthocyanins is primarily responsible for pigment-type variegation[1, 20]. For example, in *Corydalis hemidicentra*, insertion of a 254-bp transposon into the *bHLH35* gene enhances anthocyanin biosynthesis, conferring an environmentally adaptive gray phenotype[21]. Despite these advances, the developmental dynamics and transcriptional regulatory networks governing both air space-type and pigment-type variegation remain poorly understood[2].

Saxifraga stolonifera is a shade-tolerant perennial exhibiting variegated leaves with both ornamental and medicinal properties, and it displays diverse leaf variegation patterns that support adaptation across variable environments[1, 22, 23]. In its natural habitat, this species commonly shows three stable leaf phenotypes: green-leaf type, white-variegated type, and purple-variegated type. Here, we report a high-quality genome assembly of *S. stolonifera* generated by integrating Illumina short-read sequencing, PacBio long-read sequencing, and high-throughput chromosome conformation capture (Hi-C) technologies. Furthermore, we performed comprehensive metabolomics and transcriptomic profiling of young leaves, first-stage adult leaves, and fully mature leaves across the three phenotypes. A weighted gene co-expression network analysis (WGCNA) was also conducted to identify key genes and regulatory modules associated with changes in hormone-mediated metabolites during leaf variegation. By integrating morphological and physiological traits with plant hormone profiles, metabolomics data, and transcriptomic landscapes, this study provides a high-quality reference genome and offers in-depth insights into the molecular mechanisms underlying leaf variegation. These resources establish a foundation for future genetic, genomic, and functional studies within the genus *Saxifraga*.

Materials and methods

Plant materials

The green-leaf (G) and purple-variegated (P) phenotypes of *S. stolonifera* were collected from Tianmu Mountain, Zhejiang, China (30°18'30" N, 119°23'47" E), while the white-variegated (W) phenotype was obtained from Doupeng Mountain, Guizhou, China (26°20'08" N, 107°17'34" E). The formal identification of *S. stolonifera* was performed by Dr. Jianhang Zhang. Voucher specimens were deposited in herbarium of East China Normal University (HSNU). As S. *stolonifera* is not a protected species, no collection permit was required. Starting in September 2020, all three phenotypes were cultivated in a greenhouse at the Biological Station of East China Normal University using a 1:3 mixture of nutrient soil and vermiculite as

substrate. On April 20, 2021, healthy, uniformly growing individuals of each phenotype were selected for experimentation. From each plant, two leaves were sampled: one first-stage adult leaf and one fully developed adult leaf of similar size and growth status. The smallest leaf exhibiting typical morphological characteristics was designated as the first-stage adult leaf. On June, 22, 2025, the young leaves lacking characteristic pigmentation patterns were sampled. In this study, the second emerging leaf was defined as the young leaf, the third as the first-stage adult leaf (showing distinct patterning), and the sixth or seventh leaf as the mature adult leaf. Leaf developmental stages are illustrated in Fig. 1. After harvest, leaves were wrapped in aluminum foil, rapidly frozen in liquid nitrogen, and stored at − 80°C for downstream analyses. In April 2023, individuals of all three phenotypes were transplanted from East China Normal University to the greenhouse at Shaoxing University. On May, 7, 2024, root tips were collected for chromosome karyotype analysis[24], and whole plants of the green-leaf phenotype at the flowering stage were sampled for genome sequencing.

Macroscopic leaf morphology and phenotypic stability

Eighteen plants, six representing each of the three leaf phenotypes with uniform growth characteristics, were selected for morphological analysis. To assess the adaptability of the green-leaf and purple-variegated types to varying light conditions, these individuals were exposed to distinct light intensities within the same greenhouse: full sunlight, shading with a single-layer shade net (reducing light intensity to approximately 25% of full sunlight), and shading with a double-layer shade net (reducing light intensity approximately 10%). Substrate composition, watering regimen frequency, and fertilization protocols were standardized across all experimental groups. Weekly visual documentation of leaf morphology and color variation was performed using a Canon EOS 800D camera, while leaf surface features were examined using a KEYENCE VHX-5000 stereoscope (Japan).

Leaf structure and ultrastructure

Fully developed leaves of the three *S. stolonifera* phenotypes exhibiting healthy and uniform growth were used for structural and ultrastructural analyses. Leaves were manually sectioned, and thin, intact slices were carefully selected to prepare temporary water-sealed slides. Cell morphology, dimensions, and arrangement; chloroplast number, distribution, and spatial organization; and the spatial arrangement of anthocyanin-containing cells were observed under bright-field mode at 4 ×, 10 ×, and 20 × magnification using an Echo RVL-100-M inverted integrated fluorescence microscope (Discover ECHO, US). Chloroplast autofluorescence was assessed in fluorescence mode via the CY5 channel to determine chloroplast abundance and intracellular localization. For scanning electron microscopy, leaf blades were longitudinally dissected along both sides of the midvein, preserving a central segment approximately 1-2cm wide to include both regular and variegated regions. Observations were conducted using a Hitachi S-4800 cold-field emission scanning electron microscope (Japan), operated at 1 kV.

Measurements of the content of chlorophyll and anthocyanins

The chlorophyll content was measured using a SPAD-502 Plus meter (Konica Minolta, Japan) and expressed as relative SPAD values. A total of 480 leaves were sampled from 40 uniformly growing plants representing the three leaf phenotypes. One leaf each from the first-stage adult and fully adult stages, matched for size and developmental status, was selected per plant. For fully adult leaves, measurements were taken either over the main vein (white-variegated area) or interveinal areas (purple-variegated area). Due to the difficulty of distinguishing variegated and non-variegated areas on first-stage adult leaves, measurements were restricted to the main vein area. Data were processed using Microsoft Excel (2019), and differences in the chlorophyll content among leaf phenotypes were analyzed by one-way ANOVA using R v4.1.2.

The anthocyanin content was quantified using the MetWare platform (Wuhan, China; http://www.metware.cn/) and the AB Sciex QTRAP 6500 LC-MS/MS system. Nine *S. stolonifera* plants, three per phenotype, with similar growth were selected. One first-stage adult leaf and one fully adult leaf, matched for size and developmental stage, were collected from each plant, yielding nine samples in total. Samples were immediately frozen in liquid nitrogen and stored at -80 ℃ until analysis. The results of hierarchical clustering analysis (HCA) of both samples and metabolites were visualized as dendrograms accompanied by heatmaps. HCA was performed using the R package pheatmap with normalized metabolite signal intensities (unit variance scaling) represented on a color scale. Differentially abundant metabolites (DAMs) were identified based on pairwise comparisons between groups: adult leaves of the purple-variegated type (p) versus green-leaf type (g); white-variegated type (w) versus g; and p versus w. DAMs were considered significant if they met the following criteria: absolute Log_2(FC) ≥ 1 (i.e., fold change ≥ 2 or≤ -2), P-value ≤ 0.05, and presence rate ≥ 1 across all samples. Identified metabolites were annotated using the KEGG COMPOUND database (http://www.kegg.jp/kegg/compound/).

Detection of the phytohormone content

Phytohormone levels were quantified using an ultra-high-performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS) system (ExionLC™ AD UHPLC-QTRAP 6500+, AB SCIEX Corp., Boston, MA, USA) at Novogene Co., Ltd. (Beijing, China). On 22 June 2025, young leaves and first-stage adult leaves were sampled. Nine healthy plants, three per phenotype (G, P, W), with consistent growth were selected. From each plant, one young leaf (second leaf) and one first-stage adult leaf (third leaf), comparable in size and developmental stage, were harvested, resulting in 18 samples. All samples were flash-frozen in liquid nitrogen and stored at -80℃ prior before analysis. HCA was performed on all detected plant hormones, with hormone concentrations normalized and clustered accordingly. Differentially abundant metabolites were identified through pairwise comparisons using the thresholds of fold change (FC) > 1.2 or FC < 0.833 and P-value < 0.05.

Genome assembly and annotation

Genome sequencing was conducted using the Illumina platform at Biomarker Technologies (Beijing, China), integrating PacBio HiFi, Hi-C, and RNA-seq data. For genome annotation, mixed RNA samples were collected from four healthy tissues (root, stem, leaf, and flower) and subjected to RNA sequencing. Genomic DNA was extracted from *S. stolonifera* leaves to construct 150-bp paired-end libraries, which were sequenced on the Illumina HiSeq X platform. Circular consensus sequencing (CCS) reads generated by PacBio HiFi were assembled using Hifiasm (version 0.19.9-r616; https://github.com/chhylp123/hifiasm) with key parameters '-n 3 -I = 0' and default settings. Chromosome-level scaffolding was performed using HapHiC (https://github.com/zengxiaofei/HapHiC) with default parameters. Genome completeness was evaluated using BUSCO (https://github.com/metashot/busco) in genome mode (-m genome). Jellyfish v2.3.05[25] was used to count k-mers (K = 31), and GCE v1.0.2[26] was applied to estimate genome size and heterozygosity from k-mer frequency distributions.

Transposable elements (TEs) were identified using a combined homology-based and de novo approach. A de novo repeat library was constructed using RepeatModeler2 v2.0.1[27], which incorporates RECON v1.0.8[28] and RepeatScout v1.0.6[29]. Full-length LTR retrotransposons were initially detected using LTRharvest v1.5.10[30] and LTR_finder v1.07[31], and then refined with LTR_retriever. A non-redundant, species-specific TE library was generated by merging the de novo library with the Dfam v3.5 database. Final TE sequences were classified via RepeatMasker v4.12[32] using homology searches. Tandem repeats were annotated using TRF v4.09[33] and MISA v2.1[34].

Gene annotation was performed using the Makerpipeline[35]. To minimize annotation errors, gene predictions were performed using three complementary approaches: de novo prediction, transcript-based assembly, and homology-based alignment. Homology-based predictions utilized protein sequences from *Arabidopsis thaliana*, *Bergenia scopulosa*, *Chrysosplenium sinicum*, *Liquidambar formosana*, and *Vitis vinifera*, which were mapped to the *S. stolonifera*genome using GeMoMa v1.7[36]. De novo gene prediction was carried out using Augustus v3.2.3[37] and SNAP[38]. RNA-seq reads were aligned and assembled using HISAT2 v2.1.0[39] and StringTie v2.1.4[40], respectively, and used as evidence for gene prediction with GeneMarkS-T v5.1[41]. Additionally, PASA v2.4.1[42] was employed to predict gene models from RNA-Bloom v2.0.0[43] assembled unigenes and PacBio/ONT full-length transcripts. All predicted models were consolidated using EVM v1.1.1[44] and subsequently updated with PASA. The completeness of the annotated gene set was evaluated using BUSCO v5.2.2[45] in protein mode (-m proteins).

Comparative genomic analysis

A comparative genomic analysis was conducted involving *S. stolonifera* and 24 other plant species (Table S1). Protein sequences were clustered into gene families using OrthoFinder v2.4.0[46]. Functional annotations of gene families were derived from the Pfam V33.1 database[47]. Unique gene families in

each species were identified through GO and KEGG enrichment analyses. Single-copy orthologous genes present in at least 80.0% of species (n = 799) were aligned using MAFFT v7.205 (--localpair --maxiterate 1000)[48]. A maximum-likelihood phylogenetic tree was inferred using iqtree v2.2.0 (JTT + F+I+G4, 1000 bootstrap replicates)[49], with *Amborella trichopoda* designated as the outgroup. Divergence times were estimated using the MCMCTree program in the PAML v4.9i package under default settings. Gene family expansions and contractions (family-wide P-values and Viterbi P-values < 0.05) were analyzed using CAFÉ v4.2.1[50]. The evolutionary dynamics of the *S. stolonifera* genome were further explored by calculating the synonymous substitution rate (Ks) for collinear gene pairs using WGDI v0.71[51].

RNA sequencing and bioinformatic analysis

Total RNA was extracted from 38 samples, including one blind test sample from each of the P1 and W2 groups, comprising first-stage adult and fully adult leaves of the three leaf phenotypes. RNA purity and integrity were assessed using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and the Agilent RNA Nano 6000 Assay Kit on the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). Qualified libraries were subjected to next-generation sequencing (NGS) on the Illumina platform at Biomarker Technologies (Beijing, China). Additional transcriptome data for young leaves, representing the three phenotypes before characteristic patterns become visible, were collected in 2025 and sequenced on the T7 platform at Novogene Co., Ltd. (Beijing, China).

Clean reads were aligned to the *S. stolonifera* reference genome using HISAT2[39]. Read counts were quantified using the subread package in R (featureCounts), and gene expression levels were estimated using TPM (transcripts per million) via StringTie[40]. In the KEGG database, KO (KEGG Orthology) identifiers denote functionally conserved orthologous gene groups. Based on whole-genome functional annotation, all genes associated with the phenylpropanoid biosynthesis (ko00940), flavonoid biosynthesis (ko00941), and anthocyanin biosynthesis (ko00942) pathways were identified[16, 52–55]. Genes involved in diterpenoid biosynthesis (map00904) were also annotated [56].

Protein domains of candidate genes were identified using HMMER via the Quick Gene Family Identification plugin in TBtools-II[57], and genes with inconsistent domain architectures were filtered. DEGs were identified from transcriptome data using DESeq2[58] with the following criteria: false discovery rate (FDR) < 0.05 and $|\log_2$ fold change$| \geq 1$. The Venny tool (accessed 27 October 2025; https://www.omicshare.com/tools/) was used to visualize overlapping DEGs. After discarding genes with low relative expression (FPKM > 1 in more than 90% of the samples), the WGCNA plugin (Format =normalized count, Normalized method = raw, Sample percentage = 0.9, Expression Cutoff = 1, Filter Method = MAD, Reserved genes Num. =20,000; R^2 coutoff = 0.8, Recommended = 6, Customized = 8; min Module Size = 30, module cuttree height = 0.25, select max blocksize = 20,000; x axis label angle = 45, KME cutoff = 0.2, Choose method = 2) in TBtools-II was used to identify modules of highly correlated genes attributable to plant hormones based on the filtered FPKM data[57, 59].

Results

Genome assembly and annotation of S. stolonifera

To establish a genomic framework for dissecting leaf variegation mechanisms, we generated a high-quality chromosome-scale genome assembly of S. stolonifera. A green-phenotype individual of *S. stolonifera* from the Yonglai village population on Qingliangfeng Mountain in Jixi County, Anhui Province, China, was selected for sequencing (Fig. 2a). Karyotype analysis confirmed diploidy with 2n = 36 chromosomes (Fig. 2b). K-mer analysis estimated a genome size of ~ 1.63 Gb with 0.29% heterozygosity and 57.6% repetitive sequence, based on the 31-mer frequency distribution (Fig. S1). To achieve a high-quality genome assembly, we integrated multiple sequencing technologies by integrating Illumina short-read (116.53 Gb), PacBio HiFi long-read (99.65 Gb), and Hi-C (325.13 Gb) sequencing data (Table S1). The initial assembly yielded 2,139 contigs, which were subsequently anchored into 18 pseudo-chromosomes using Hi-C interaction maps, accounting for 94% of the assembly (Fig. 2c, d; Tables 1, S1, S2). The final genome assembly spans 2.01 Gb with a scaffold N50 of 3,535,068 bp and a BUSCO completeness score of 98.5%, indicating high contiguity and completeness (Table 1; Table S3).

Table 1
Statistics for the final genome assembly of *S. stolonifera*

Genome information	***S. stolonifera* (PacBio + Hi-C)**
PacBio HiFi data (Gb)	99.66
Hi-C clean data (Gb)	325.13
Assembly size (Gb)	2.01
Number of contigs	2,139
Scaffold N50 (bp)	3,535,068
Scaffold max (bp)	14,862,665
Anchor ratio (%)	94.00%
GC content	34.12%
BUSCO (%)	98.50%

Repetitive sequences constitute a major fraction of eukaryotic genomes and primarily include tandem repeats and interspersed repeats[60, 61]. To annotate repetitive elements in the *S. stolonifera* genome, both de novo and homology-based prediction methods were employed. A total of 1,835,418 repetitive sequences were identified, spanning 822,048,413 bp (~ 73.49% of the assembled genome) (Table S4). Retroelements were the most abundant repeat class, accounting for 44.99% of the genome, followed by DNA transposons (14%) and tandem repeats (14.5%) (Table S4).

Gene models for the *S. stolonifera* genome were predicted through an integrative approach combining ab initio, homology-based, and transcriptome-supported predictions. This strategy yielded a total of 37,191 protein-coding genes, with an average gene length of 3,914.91 bp and an average of 4.95 exons per gene (Fig. S2; Table S5). Of these, 36,664 genes were successfully annotated, resulting in an annotation rate of 98.58% (Tables S6, S7). TrEMBL analysis indicated that 36,585 genes (98.37%) were functionally annotated, with 28,614 encoding metabolic enzymes as classified by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Table S7). The completeness and quality of the annotation were evaluated using BUSCO. Among the 1,614 conserved orthologs assessed, 1588 (98.39%) were complete, including 86.18% single-copy and 12.21% duplicated genes, indicating high annotation completeness (Table S8). This high-quality genome provides a robust reference for subsequent functional and evolutionary analyses.

Genome evolutionary and whole-genome duplication analysis of S. stolonifera

To place *S. stolonifera* in an evolutionary context, we compared its genome with 24 representative plant species (Table S9). A total of 4,130 shared gene families were identified across all species, while 402 gene families were unique to *S. stolonifera*, enriched in pathways related to aminoacyl-tRNA biosynthesis, riboflavin metabolism, phenylpropanoid biosynthesis, and phenylpropanoid / flavonoid biosynthesis (Fig. 3a, b; Fig. S3; Tables S10, S11). These pathway enrichments suggest potential lineage-specific adaptations in secondary metabolite production in *S. stolonifera*. Phylogenomic analysis using 799 single-copy orthologs confirmed that Saxifragaceae is sister to Grossulariaceae, with *S. stolonifera* diverging from other Saxifragaceae species (*Bergenia scopulosa*, *Tiarella polyphylla*, and *Chrysosplenium sinicum*) approximately 59.39 million years ago (46.83–71.99 MYA) (Fig. 3c). Gene family evolution analysis revealed 155 expanded and 17 contracted gene families in *S. stolonifera* (Fig. 3c; Tables S12–S14). Notably, expanded families were significantly enriched in flavonoid and phenylpropanoid biosynthesis pathways (Fig. S4a; Table S13). Anthocyanin accumulation not only reduces photodamage[19, 62] but also underlies purple pigmentation in plants[2, 13]. Thus, the expansion of these gene families appears functionally linked to anthocyanin biosynthesis in *S. stolonifera*.

Whole-genome duplications (WGDs) are recognized as key drivers of plant genome evolution[63, 64]. Analysis of synonymous substitution rates (Ks) among paralogous gene pairs revealed two distinct peaks at ~ 0.32 and ~ 1.37, indicating two whole-genome duplication (WGD) events (Fig. 3d; Fig. S5a), suggesting two independent WGD events. The ancient peak (Ks ≈ 1.37) corresponds to the core eudicot gamma whole-genome triplication, whereas the recent peak (Ks ≈ 0.32) represents a lineage-specific WGD that occurred after the divergence of *S. stolonifera* from other Saxifragaceae species (Fig. 3e; Figs. S5b, S6)).

Microsynteny analyses further support two rounds of WGD along the *S. stolonifera* lineage, with a 4:1 gene copy ratio observed between the *S. stolonifera* and *Vitis vinifera* genomes (Fig. 3d, e). Higher chromosomal collinearity was observed between *S. stolonifera* and the three Saxifragaceae species (*T. polyphylla*, *B. scopulosa*, and *C. sinicum*), compared with *R. nigrum* (Grossulariaceae) (Fig. S5b),

reinforcing the earlier divergence of *S. stolonifera* from *R. nigrum* and its closer relationship with the other Saxifragaceae members. Comparative genomic analyses between *S. stolonifera* and *B. scopulosa* or *C. sinicum* revealed syntenic depth ratios of 4:3 and 5:3, respectively (Fig. S6), indicating a shared WGD. Together, these findings provide strong evidence for the shared gamma-WGT event across Saxifragaceae and confirm an independent, lineage-specific WGD in *S. stolonifera* following this ancient duplication.

Morphological and physiological basis of leaf variegation

Transverse sections of the green-leaf type (G), white-variegated type (W), and purple-variegated type (P) leaves were prepared (Fig. 4; Fig. S7–S9). These sections revealed that air spaces were prevalent within the epidermal and palisade tissues along the central vein in both the white-variegated and purple-variegated types (Fig. 4, b, c, e, f, h, i, k, l; Fig. S7–S8). Notably, in the white-variegated type, air spaces were observed among palisade cells (Fig. 4b, e, h, k; Fig. S7). In contrast, epidermal and palisade tissue cells in the green-leaf type were densely packed (Fig. 4a, d, g, j; Fig. S9). The chlorophyll content did not differ significantly among the three leaf phenotypes, nor across different regions of individual leaves or developmental stages (Fig. S10; Table S15). Therefore, the appearance of white veins primarily results from air spaces located between the epidermal and palisade tissues along the central vein in the variegated types.

In transverse sections, the spatial distribution of red pigments varied across leaf phenotypes. In the white-variegated type, pigment-containing cells were scattered throughout the mesophyll and hypoepidermal tissues (Fig. S7c), whereas in the green-leaf type, they were predominantly confined to the hypoepidermis (Fig. S9c). Neither the green-leaf nor the white-variegated type exhibited purple patches on the adaxial surface (Fig. 4a–c; Fig. S7a, S9a). Conversely, the purple-variegated type displayed accumulation of red pigments in the lower layers of the palisade tissue, leading to distinct purple patches visible on the upper leaf surface (Fig. S8a–e). Collectively, the prominent white veins in the white-variegated type arise from intercellular air spaces between the epidermis and palisade tissue, coupled with altered cellular organization of the palisade layer. The conspicuous purple patches in the purple-variegated type result from the localized accumulation of red pigments in sub-palisade cells, while the associated white venation shares the same structural origin as in the white-variegated type. According to the classification of leaf variegations by Zhang *et al*.[1], the white veins in the white-variegated type represent an air space-type variegation, whereas the purple-variegated type combines both pigment-based and air space-type mechanisms.

Metabolomic profiling identifies anthocyanin glycosides as key pigments for purple-variegated type leaves

To identify the metabolites responsible for purple-variegated type, we profiled anthocyanins in leaves of the three phenotypes using LC-MS/MS. A total of 58 anthocyanin-related compounds were detected (Table S16). Comparative analysis revealed 10 differentially accumulated metabolites (DAMs) enriched

in the anthocyanin and flavonol biosynthesis pathways (ko00942, ko00944), including cyanidin-3-O-glucoside, peonidin-3-O-galactoside, and quercetin-3-O-glucoside (Fig. 5a; Table S17). Notably, these compounds accumulated at significantly higher levels in purple-variegated type compared to green-leaf type, white-variegated type leaves (Fig. 5a; Table S17). All identified pigments are glycosylated forms of anthocyanidins, suggesting that glycosylation plays a critical role in their stabilization and accumulation. These results pinpoint anthocyanin glycosides as the chemical basis of the purple patches.

Plant hormone profiling identifies for white-variegated type leaves

Plant hormones profiling detected 46 hormones across eight classes (Fig. 5b; Table S18). Comparative analysis identified 35 differentially accumulated hormones, with the most pronounced differences observed between purple-variegated type and green-leaf type leaves (Table S19). Strikingly, both purple-variegated type and white-variegated type (W and P) accumulated significantly higher levels of multiple gibberellins (GA4, GA7, GA9, GA15, GA24) compared to green-leaf type leaves, in both young and first-stage adult leaves (Fig. 5c; Table S19). Among these, GA4 and GA7 are bioactive forms known to promote cell expansion[65]. The elevated GA levels in variegated leaves point to a potential role for gibberellin in mediating air space formation.

Transcriptomic analysis identifies SsBZ1 and SsGA3ox as key regulators of pigment type and air space type leaf variegation

To identify genes associated with the formation of air space-type and pigment-type variegation, RNA sequencing was performed on leaves across the three developmental stages of the three leaf phenotypes in *S. stolonifera*. An average of 7.32 Gb clean reads per sample was obtained (Table S20), with an average mapping rate of 85%. A total of 18,923 differentially expressed genes (DEGs) were identified across all comparisons, with the highest number observed in the W2 vs. G2 comparison (6,209 DEGs; Fig. S11a; Table S21). Among these, 4,457 DEGs were commonly differentially expressed in at least two of the three stage specific comparisons between the purple-variegated and green-leaf types (YP-vs-YG, P1-vs-G1, and P2-vs-G2; Fig. S11b; Table S21). Similarly, 4,895 DEGs were shared across at least two comparisons between the white-variegated and green-leaf types (YW-vs-YG, W1-vs-G1, and W2-vs-G2; Fig. S11d; Table S21). Integrating these results with our high-quality *S. stolonifera* genome, we focused on key regulatory genes involved in anthocyanin and gibberellin biosynthesis, reconstructing the anthocyanin (Fig. 6) and gibberellin (Fig. 7a) pathways associated with leaf variegation.

We annotated 63 enzymatic genes involved in the anthocyanin biosynthesis pathway and visualized their expression patterns across three developmental stages of the three leaf phenotypes (Table S22). By reconstructing anthocyanin synthesis in *S. stolonifera* leaves, we found that *Ss4CL* (*Sst04G018150*, *Sst04G018170*, *Sst04G019880*, *Sst14G006060*), *SsCHS* (*Sst12G013330, Sst08G008480*), and *SsBZ1* (*Sst14G006050, Sst12G009310*) were expressed at relatively high levels in fully developed leaves of the purple-variegated type (P2) (Fig. 6; Fig. S11 c; Table S22, S23). Notably, anthocyanidin 3-O-glucosyltransferase genes (*SsBZ1*: *Sst12G009310*, *Sst14G006050*) showed elevated expression in both

the white-variegated (W2) and purple-variegated (P2) leaf types, coinciding with the significant accumulation of key pigmentation metabolites: cyanidin-3,5-O-diglucoside, cyanidin-3-O-glucoside, peonidin-3-O-galactoside, peonidin-3-O-glucoside, peonidin-3,5-O-diglucoside, and quercetin-3-O-glucoside (Fig. 5a; Table S17).

In gibberellin biosynthesis, 24 enzymatically active genes were identified, and their expression profiles were analyzed across the three leaf phenotypes (Table S24). Reconstruction of the gibberellin pathway in *S. stolonifera* revealed that gibberellin 3β-dioxygenase (*SsGA3ox*: *Sst05G017510*) was highly expressed in all three phenotypes, particularly in the white-variegated type (groups W1 and W2), likely contributing to sustained high concentrations of bioactive GA4 (Fig. 7a; Fig. S11 e; Table S25). These expression patterns are consistent with the significant accumulation of GA4, the primary active gibberellin, in these tissues (Fig. 5b, c, 7a; Table S19). Collectively, these results suggest that the upregulated expression of GA 3-oxidases (*GA3ox*), which maintains elevated levels of active gibberellin, is a key factor underlying white vein formation.

WGCNA identifies a gibberellin-associated co-expression module with SsGA3ox

To gain further insights into the regulatory mechanisms underlying gibberellin content fluctuations during leaf variegation, weighted gene co-expression network analysis (WGCNA) was performed. A total of 9,960 DEGs were filtered and retained based on Fragments Per Kilobase of transcript per Million mapped reads (FPKM) > 1 in more than 90% of the samples and showing significant differential expression across the three developmental stages of the three leaf phenotypes (Fig. S12). This analysis revealed 23 co-expression modules (Fig. 7b, c, color-coded). Among these gene co-expression subnetworks (Fig. 7c), several exhibited strong correlations with the synthesis and activation of the five gibberellins. Notably, the two bioactive gibberellins, GA4 and GA7, showed highly significant positive correlations with the salmon-coded module, with correlation coefficients of 0.9 and 0.94, respectively. Gene expression patterns within this module were characterized by high transcript levels in the white-variegated and purple-variegated leaf types (Fig. 7a; Fig. S11 d, e). Within the salmon-coded module, the key gene *SsGA3ox* (*Sst05G017510*) displayed significantly elevated expression in white-variegated leaves, particularly in fully expanded adult leaves (Fig. 7a; Fig. S13). These findings support the hypothesis that upregulated *SsGA3ox* expression maintains elevated levels of active gibberellins, thereby promoting rapid leaf expansion and contributing to the formation of intercellular air spaces between the epidermal and palisade tissues along the veins in air space type-leaf variegation.

Discussion

Here, we present a high-quality chromosome-scale genome assembly for *S. stolonifera*. Ks distribution analyses provided compelling evidence for a shared gamma WGT event among Saxifragaceae species and revealed an independent WGD event in *S. stolonifera* after the gamma-WGT. Comparative analysis of gene families across 25 species identified expanded gene families in *S. stolonifera* that are highly

enriched in phenylpropanoid biosynthesis, fatty acid metabolism, stilbenoid diarylheptanoid and gingerol biosynthesis, and flavonoid biosynthesis. Additionally, we found that *GA3ox* genes in *S. stolonifera* play a significant role in maintaining elevated levels of active gibberellins. Collectively, these findings suggest that high concentrations of active gibberellins may have been crucial for the development of air spaces between the epidermal and palisade tissues along leaf veins.

Localized anthocyanin accumulation determines the pigment type of S. stolonifera

Bioinformatic analysis of *S. stolonifera* leaves indicated that the up-regulated expression of the anthocyanidin 3-O-glucosyltransferase gene (*SsBZ1*: *Sst12G009310*, *Sst14G006050*) was a primary contributor to the formation of the purple-variegated phenotype. Anthocyanidin 3-O-glucosyltransferase is a key glycosyltransferase in plants[66]. Glycosylation plays a critical role in secondary metabolite biosynthesis by enhancing the stability, solubility, subcellular localization, and biological activity of conjugated compounds[66–69]. Enhanced expression of this gene promotes the glucosylation of anthocyanidins into stable anthocyanins, thereby preserving their coloration[55]. Previous studies have shown that the anthocyanidin 3-O-glucosyltransferase gene is up-regulated in anthocyanin-rich, non-green leaves, such as those of *Alternanthera bettzickiana*[12], *Camellia sinensis* 'Zijuan'[13], and *Perilla frutescens*[14]. Similarly, in transcriptomic analyses of pigment-type variegation in *Begonia masoniana*, the *UFGT* gene (anthocyanidin 3-O-glucosyltransferase), involved in anthocyanin glycosylation is also up-regulated[17].

In this study, expression of *SsBZ1* (*Sst12G009310*, *Sst14G006050*) was significantly up-regulated in the purple-variegated type compared with the green-leaf type (Fig. 6; Table S22). The content of visible anthocyanins was markedly higher in the purple-variegated type than in the green-leaf type, consistent with the elevated expression of *SsBZ1* (Fig. 5a, 6; Fig. S11 c; Table S17). Therefore, up-regulation of anthocyanidin 3-O-glucosyltransferase likely underlies the formation of the purple-variegated phenotype.

Furthermore, compared with the white-variegated type, visible anthocyanins in the purple-variegated type were localized specifically within palisade tissue cells between the veins rather than scattered in the mesophyll cells (Fig. 4; Fig. S7–9). This spatial pattern explains why both types contain similar levels of visible anthocyanins (Fig. 5a), yet only the purple-variegated type exhibits distinct purple patches. The localized accumulation of anthocyanins in palisade cells constitutes the structural basis for the observed pigmentation pattern and has also been documented in other pigment-variegated plants, including *Nephelaphyllum pulchrum*, *Streptolirion volubile*, and *Vriesea saundersii*[1]. In conclusion, up-regulation of the anthocyanidin 3-O-glucosyltransferase gene enhances anthocyanin accumulation, forming the molecular foundation of pigment-type variegation in *S. stolonifera*. However, the molecular mechanisms governing the specific localization of anthocyanins in palisade tissue cells between veins remain unclear.

High gibberellin levels promote air spaces formation

Based on the chromosome-level genome sequence, combined with gene co-expression and plant hormone profiling, we reconstructed the gibberellin biosynthesis pathway and its regulatory network in *S. stolonifera*. GAs are phytohormones that regulate multiple aspects of plant development[70], including leaf expansion[71], morphogenesis[72], and final leaf size[73]. Elevated GA concentrations accelerate cell division and leaf elongation rates[74–76]. Hormone profiling revealed that increased levels of bioactive GAs, particularly GA4 and GA7, play a pivotal role in the formation of air space-type leaf variegation in *S. stolonifera* (Fig. 5c). During early leaf development, significant accumulation of multiple gibberellins, including GA4, GA7, GA9, GA15, and GA24, was observed in both the purple-variegated and white-variegated types (Fig. 5c; Table S19).

GA3ox catalyzes the final step in gibberellin biosynthesis, converting precursor GAs into the bioactive forms GA1 and GA4[77–82]. These enzymes are also known as gibberellin 3β-dioxygenases. Transcriptomic data demonstrated that the gibberellin 3β-dioxygenase gene (*SsGA3ox: Sst05G017510*), which mediates this conversion, was highly expressed in white-variegated leaves, especially at the fully mature stage (Fig. 7a; Fig. S11d e, S13). This sustained expression likely maintains elevated levels of active gibberellins, promoting rapid cell division and leaf expansion, key processes facilitating physical separation between epidermal and palisade tissue cells and resulting in air space formation along veins. Supporting this, WGCNA identified a co-expression module (salmon-colored) strongly correlated with GA4 and GA7 levels (Fig. 7c), and genes within this module, including *SsGA3ox*, were highly expressed in variegated leaves (Fig. S12, S13). Together, these results indicate that gibberellin mediated cell expansion and tissue patterning form the structural basis of air space-type variegation. This mechanism enhances our understanding of leaf morphological diversity and provides a physiological and molecular explanation for the development of non-pigmented white veins in variegated leaves of *S. stolonifera*.

In sum, the high-quality reference genome of *S. stolonifera* presented here serves as a valuable resource for studying evolutionary dynamics and phenotypic diversity within the genus *Saxifraga*. Moreover, comprehensive transcriptomic, anthocyanin metabolic, and plant hormone profiling analyses provide novel insights into the molecular mechanisms underlying both pigment-type and air space-type variegation. The data generated in this study represent essential resources for future functional genomics and genetic investigations in *Saxifraga*.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors agree to the publication of this manuscript upon its acceptance.

Conflict of interest

The authors declare no competing interests.

Availability of data and materials

The data that support the findings of this study have been deposited in the CNSA (accessed on 28 November 2024, https://db.cngb.org/cnsa/) of CNGBdb with accession code CNP0006535.

Acknowledgments

We extend our sincere thanks to Dr. Biao Xiong (Guizhou University) for guiding some data analyses. Our gratitude goes to Quangang Xue, Jiayu Jin, and Boyang Xie (Shaoxing University) for their kind help with the materials planting. We thank TopEdit (www.topeditsci.com) for linguistic assistance during the preparation of this manuscript. We also appreciate the assistance of Dr. Bin Dong (Zhejiang A&F University) for his help in improving our manuscript.

AI declaration

During the preparation of this work, the author(s) used [DeepSeek] solely for language polishing and to improve readability. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Author contributions

H.L., S.L., J.L., and J.Z. conceived and designed the project; H.L., J.Z., and S.L. provided financial support; S.L., J.L., and J.Z. collected the samples; S.L., J.L., and J.Z. performed the experiments; F.Z., H.X., F.T., J.Z., and S.L. analyzed the data; F.Z., H.X., F.T., and C.W. contributed to project discussion. H.L., J.Z., and S.L. wrote the manuscript. All authors contributed critically to the drafts and gave final approval for publication. All authors have read and agreed to the published version of the manuscript.

Funding

The project was supported by a grant from the National Natural Science Foundation of China (32300178) to S.L., the Startup Fund for Shaoxing University (13011001002/218) to J.Z., Biodiversity Investigation, Observation and Assessment Program (2019-2023) of Ministry of Ecology and Environment of China to H.L.

Supplemental data

Supplementary data are available at *BMC Genomics* online.

References

1. Zhang JH, Zeng JC, Wang XM, et al. A revised classification of leaf variegation types. Flora. 2020;272:151703.

2. Zhang JH, Li JC, Zou L, et al. Transcriptome analysis of air space-type variegation formation in *Trifolium pratense*. Int J Mol Sci. 2022;23:7794–7794.
3. Zhang JH. Classification of leaf variegation types in angiosperm and variegation related genes of *Trifolium pretense*. East China Normal Univ. 2022.
4. La Rocca N, Pupillo P, Puppi G et al. *Erythronium dens-canis* L. (Liliaceae): An unusual case of change of leaf mottling. *Plant Physiol Biochem.*2014; 74: 108–117.
5. Smith AP. Ecology of a leaf color polymorphism in a tropical forest species: habitat segregation and herbivory. Oecologia. 1986;69:283–7.
6. Sheue CR, Pao SH, Chien LF, et al. Natural foliar variegation without costs? The case of *Begonia*. Ann Bot. 2012;109:1065–74.
7. Pao SH, Chesson P, Peng CI et al. Characteristics and ecological significance of natural foliar variegation from *Begonia formosana*. In: The 99th Annual Meeting, 10th-15th (2014) August. Sacramento, California, America. Ecological Society of America.
8. Shelef O, Summerfield L, Lev-Yadun S, et al. Thermal benefits from white variegation of *Silybum marianum* leaves. Front Plant Sci. 2019;10:688.
9. Li X, Kanakala S, He YH, et al. Physiological characterization and comparative transcriptome analysis of white and green leaves of *Ananas comosus* var. *bracteatus*. PLoS ONE. 2017;12:e0169838.
10. Sun YH, Hung CY, Qiu J, et al. Accumulation of high OPDA level correlates with reduced ROS and elevated GSH benefiting white cell survival in variegated leaves. Sci Rep. 2017;7:1–16.
11. Wang QM, Cui JG, Dai HY et al. Comparative transcriptome profiling of genes and pathways involved in leaf-patterning of *Clivia miniata* var. *variegata*. *Gene*. 2018; 677: 280–8.
12. Li WJ, Li HG, Shi LS, et al. Leaf color formation mechanisms in *Alternanthera bettzickiana* elucidated by metabolite and transcriptome analyses. Planta. 2022;255:59.
13. Xie H, Zhu JY, Wang H, et al. An enhancer-transposable element from purple leaf tea varieties underlies the transition from evergreen to purple leaf color. Plant Commun. 2025;6(2):101176.
14. Liu XN, Zhai YN, Liu JY, et al. Comparative transcriptome sequencing analysis to postulate the scheme of regulated leaf coloration in *Perilla frutescens*. Plant Mol Biol. 2023;112:119–42.
15. Chen J, Li Y, He D, et al. Cytological, physiological and transcriptomic analysis of variegated Leaves in *Primulina pungentisepala* offspring. BMC Plant Bio. 2022;22:419.
16. Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. Plant J. 2008;54:733–49.
17. Yang T, Xue ZZ, Li N. Reference genes selection and validation in *Begonia masoniana* leaves of different developmental stages. Acta Hortic Sin. 2021;48:2251–61.
18. Zhu J, Wang YZ, Wang QY, et al. The combination of DNA methylation and positive regulation of anthocyanin biosynthesis by MYB and bHLH transcription factors contributes to the petal blotch formation in Xibei tree peony. Hortic Res. 2023;10:uhad100.

19. Araguirang GE, Richter AS. Activation of anthocyanin biosynthesis in high light-what is the initial signal? New Phytol. 2022;236:2037–43.
20. Cui WH, Guan KY. Diversity of leaf variegation in Chinese Begonias. Plant Divers Resour. 2013;35:119–27.
21. Zhang H, Zhang P, Niu Y, et al. Genetic basis of camouflage in an alpine plant and its long-term co-evolution with an insect herbivore. Nat Ecol Evol. 2025;9:628–38.
22. Kong LY, Ou GJ, Wang CH. Application of *Saxifraga stolonifera* Curt. to landscape architecture. Shaanxi For Sci Technol. 2014;2:77–8.
23. Tang SM, Cai WQ, Zhang DM, et al. Effects of light intensites on morphological and physiological idexes of three ornamental cultivars of *Saxifraga stolonifera*. Guihaia. 2023;43:699–711.
24. Li GF, He Y, Qin SH, et al. Optimization of chromosome preparation and karyotype analysis of polyploid watermelons. J South Agric. 2017;48:663–8.
25. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70.
26. Liu B, Shi Y, Yuan J, et al. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. arXiv. 2012. 10.48550/arXiv.1308.2012.1308.2012v2.
27. Flynn JM, Hubley R, Goubert C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. PNAS. 2020;117:9451–7.
28. Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. 2002;12:1269–76.
29. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005; Suppl 1: i351.
30. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinforma. 2008;9:18.
31. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:W265–8.
32. Tarailo-Graovac M, Chen NS. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protocols Bioinf 2009; 4: 4.10. 11–14.10. 14.
33. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
34. Beier S, Thiel T, Münch T, et al. Misa-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.
35. Cantarel BL, Korf I, Robb SMC. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18:188–96.
36. Keilwagen J, Wenk M, Erickson JL, et al. Using intron position conservation for homology-based gene prediction. Nucleic acids res. 2016;44:e89–89.

37. Stanke M, Keller O, Gunduz I, et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34:W435–9.
38. Korf I. Gene finding in novel genomes. BMC Bioinforma. 2004;5:59.
39. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.
40. Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11:1650–67.
41. Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. 2015;43:e78.
42. Haas BJ, Delcher AL, Mount SM, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31:5654–66.
43. Nip KM, Chiu R, Yang C, et al. RNA-Bloom enables reference-free and reference-guided sequence assembly for single-cell transcriptomes. Genome Res. 2020;30:1191–200.
44. Haas BJ, Salzberg SL, Zhu W, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7.
45. Simão FA, Waterhouse RM, Ioannidis P, et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
46. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome boil. 2019;20:1–14.
47. Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9.
48. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol. 2009;537:39–64.
49. Nguyen LT, Schmidt HA, von Haeseler A, et al. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
50. Han MV, Thomas GWC, Lugo-Martinez J, et al. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 2013;30:1987–97.
51. Sun PC, Jiao BB, Yang YZ, et al. A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol Plant. 2022;15:1841–51.
52. Meng Y, Zhang H, Fan Y, et al. Anthocyanins accumulation analysis of correlated genes by metabolome and transcriptome in green and purple peppers (*Capsicum annuum*). BMC Plant Biol. 2022;22:358.
53. Wang Q, Zhu J, Li B, et al. Functional identification of anthocyanin glucosyltransferase genes: a Ps3GT catalyzes pelargonidin to pelargonidin 3-O-glucoside painting the vivid red flower color of Paeonia. Planta. 2023;257:56.
54. Geng Z, Liu M, Wang Y, et al. Transcriptomic and metabolomic analyses reveal CmMYB308 as a key regulator in the pink flower color variation of ‘Dante Purple’ chrysanthemum. Plant Cell Rep.

2024;43:157.
55. Xia XM, Du HL, Hu XD, et al. Genomic insights into adaptive evolution of the species-rich cosmopolitan plant genus *Rhododendron*. Cell Rep. 2024;43:114745.
56. Liu FH, Zhao Y, Wang XR, et al. Regulatory roles of GA and sugar metabolism genes with metabolomic profiling during seed dormancy release in *Gleditsia sinensis*. BMC Plant Biol. 2025;25:1231.
57. Chen C, Wu Y, Li J, et al. TBtools-II: a one for all, all for one bioinformatics platform for biological big-data mining. Mol Plant. 2023;16:1733–42.
58. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
59. Wang X, Liang S, Yang WQ, et al. MetMiner: A user-friendly pipeline for large-scale plant metabolomics data analysis. J Integr Plant Biol. 2024;66:2329–45.
60. Bendich AJ, Rogers SO. The biological and evolutionary consequences of competition between DNA sequences that benefit the cell and DNA sequences that benefit themselves. Nucleic Acids Res. 2025;53:gkaf589.
61. Wang F, Jiang Z, Gao J, et al. Analysis of the *Rehmannia chingii* geneome identifies RcCYP72H7 as an epoxidase in iridoid glycoside biosynthesis. Nat Commun. 2025;16:6035.
62. Solovchenko AE, Merzlyak MN. Screening of visible and UV radiation as a photoprotective mechanism in plants. Russ J Plant Physiol. 2008;55:719–37.
63. Jiao YN, Wickett NJ, Ayyampalayam S, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473:97–100.
64. Adams K. Genomic clues to the ancestral flowering plant. Science. 2013;342:1456–7.
65. Hauvermale AL, Ariizumi T, Steber CM. Gibberellin signaling: a theme and variations on DELLA repression. Plant physiol. 2012;160:83–92.
66. Yin RH, Messner B, Faus-Kessler T, et al. Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation. J Exp Bot. 2012;63:2465–78.
67. Li Y, Baldauf S, Lim EK, et al. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. J Biol Chem. 2001;276:4338–43.
68. Meßner B, Thulke O, Schäffner AR. Arabidopsis glucosyltransferases with activities toward both endogenous and xenobiotic substrates. Planta. 2003;217:138–46.
69. Bowles D, Lim EK, Poppenberger B, et al. Glycosyltransferases of lipophilic small molecules. Annu Rev Plant Biol. 2006;57:567–97.
70. Salazar-Cerezo S, Martínez-Montiel N, García-Sánchez J, et al. Gibberellin biosynthesis and metabolism: A convergent route for plants, fungi and bacteria. Microbiol Res. 2018;208:85–98.
71. Li J, Gao HW, Jiang JS, et al. Overexpression of the *Galega orientalis* gibberellin receptor improves biomass production in transgenic tobacco. Plant Physiol Biochem. 2013;73:1–6.

72. Gao LW, Lyu SW, Tang J, et al. Genome-wide analysis of auxin transport genes identifies the hormone responsive patterns associated with leafy head formation in Chinese cabbage. Sci Rep. 2017;7:42229.
73. Mu XH, Chen QW, Wu XY, et al. Gibberellins synthesis is involved in the reduction of cell flux and elemental growth rate in maize leaf under low nitrogen supply. Environ Exp Bot. 2018;150:198–208.
74. Wenzel CL, Williamson RE, Wasteneys GO, et al. Gibberellin-induced changes in growth anisotropy precede gibberellin-dependent changes in cortical microtubule orientation in developing epidermal cells of barley leaves. Kinematic and cytological studies on a gibberellin-responsive dwarf mutant, M489. Plant Physiol. 2000;124:813–22.
75. Nelissen H, Rymen B, Jikumaru Y, et al. A local maximum in gibberellin levels regulates maize leaf growth by spatial control of cell division. Curr Biol. 2012;22:1183–7.
76. Zhang XJ, He LL, Zhao BL, et al. *Dwarf and Increased Branching 1* controls plant height and axillary bud outgrowth in *Medicago truncatula*. J Exper Bot. 2020;71:6355–65.
77. Lester DR, Ross JJ, Davies PJ, et al. Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. Plant Cell. 1997;9:1435–43.
78. Itoh H, Ueguchi-Tanaka M, Sentoku N, et al. Cloning and functional analysis of two gibberellin 3β-hydroxylase genes that are differently expressed during the growth of rice. PNAS. 2001;98:8909–14.
79. Dalmadi A, Kaló P, Jakab J, et al. Dwarf plants of diploid *Medicago sativa* carry a mutation in the gibberellin 3-beta-hydroxylase gene. Plant Cell Rep. 2008;27:1271–9.
80. Hedden P, Thomas SG. Gibberellin biosynthesis and its regulation. Biochem J. 2012;444:11–25.
81. Chen Y, Hou M, Liu L, et al. The maize DWARF1 encodes a gibberellin 3-oxidase and is dual localized to the nucleus and cytosol. Plant Physiol. 2014;166:2028–39.
82. Phillips AL, Huttly AK, Alarcón-Reverte R, et al. *GIBBERELLIN 3-OXIDASE* genes regulate height and grain size in bread wheat. J Exp Bot. 2025;76:3345–58.

Figures

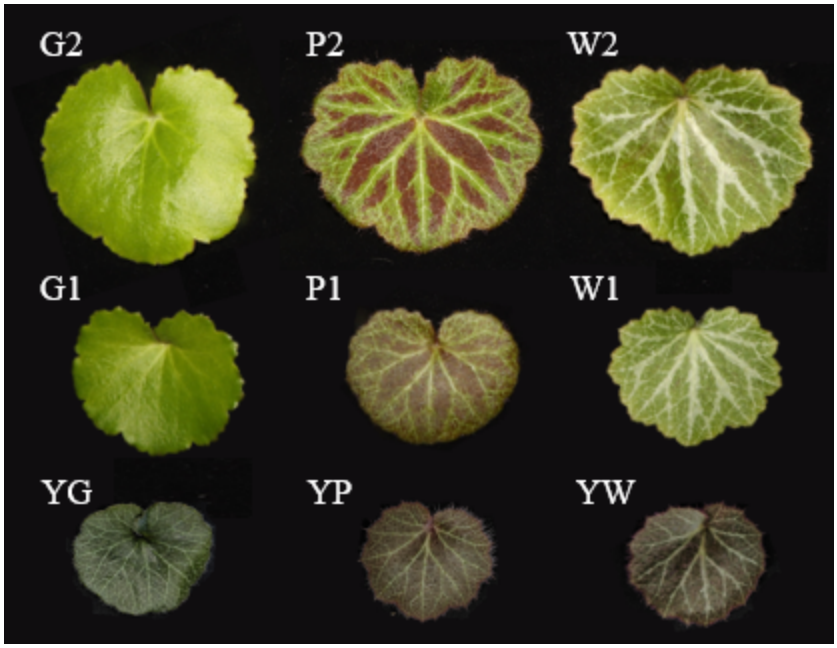

Figure 1

Status of sampled leaves of *S. stolonifera*. **G1**, first-stage adult leaf of the green-leaf type; **G2**, fully adult leaf of the green-leaf type; **YG**, young leaf of the green-leaf type; **P1**, first-stage adult leaf of the purple-variegated type; **P2**, fully adult leaf of the purple-variegated type; **YP**, young leaf of the purple-variegated type. **W1**, first-stage adult leaf of the white-variegated type; **W2**, fully adult leaf of the white-variegated type; **YW**, young leaf of the white-variegated type.

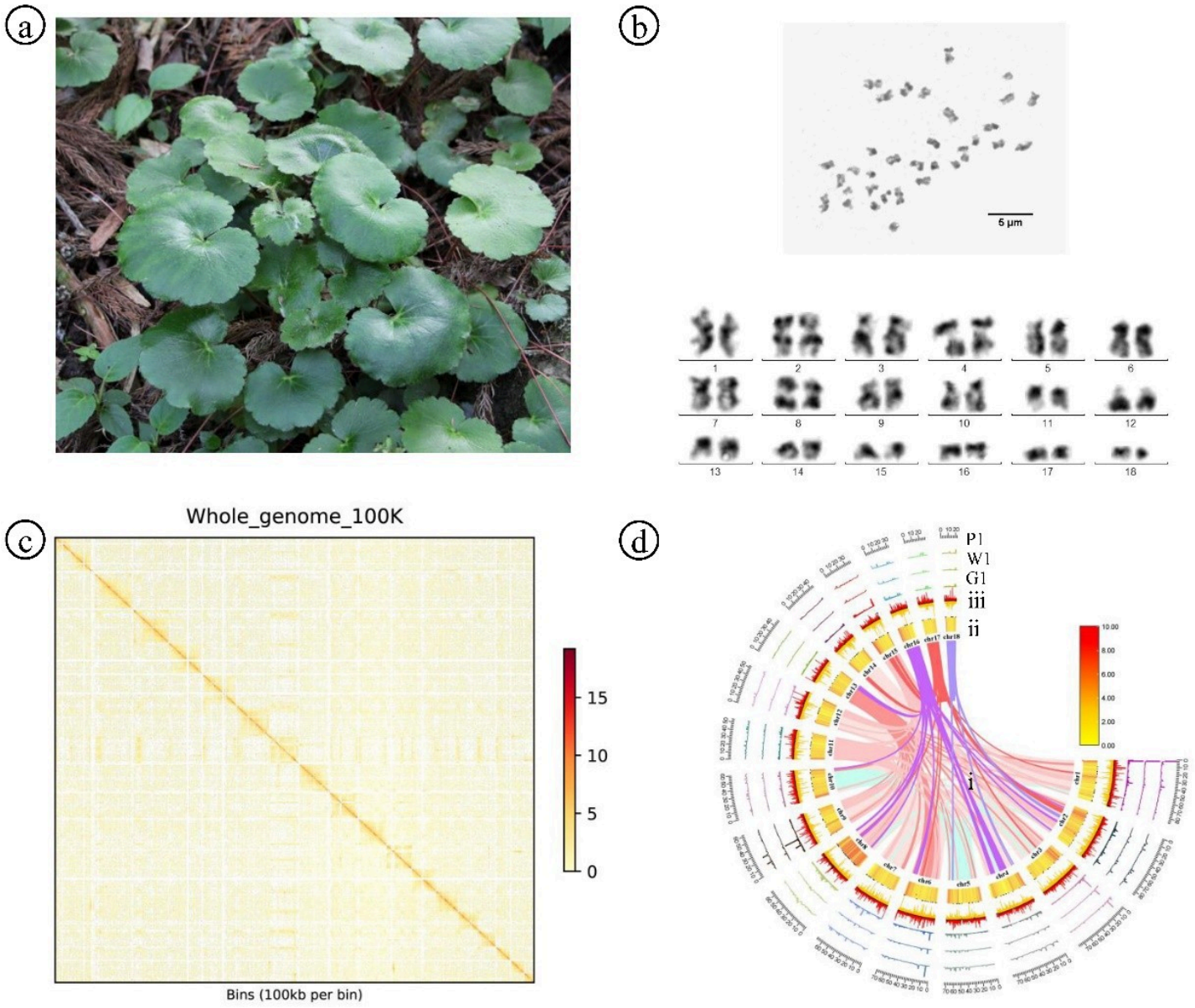

Figure 2

Overview of *S. stoloniferag*enome assembly. **(a)** Photographs of *S. stolonifera*. **(b)** Chromosome counts in *S. stolonifera* root tip cells. **(c)** Hi-C interaction heatmap of *S. stolonifera*. **(d)** Circos plot of the *S. stolonifera* genome assembly. , syntenic blocks; , gene density and unknown base (N) ratio; , GC contentand skew; G1, W1, and P1, sequencing coverage for three phenotypes (G, green-leaf type; W, white-variegated type; P, purple-variegated type).

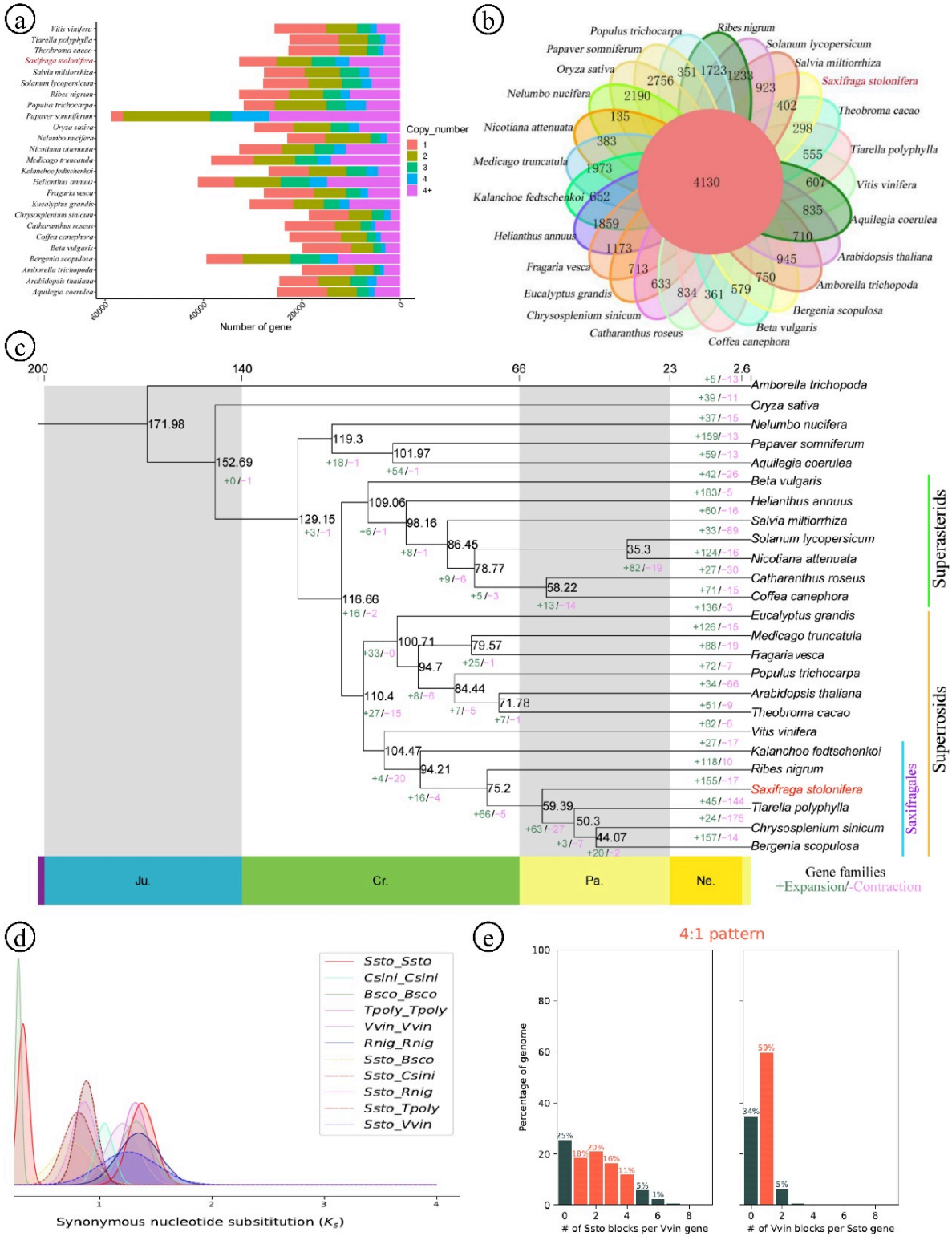

Figure 3

Comparative genomic and evolutionary analysis of *S. stolonifera*. **(a)** Copy number distribution of all gene families in 25 selected angiosperm species. **(b)** Gene family cluster petal diagram, with the central circle representing common gene families and the outer petals depicting species-specific gene families.**(c)** Phylogenetic tree of 25 plant species. Numbers represent divergence times at each node (Mya, million years ago). Gene family expansions and contractions are indicated by green and red numbers,

respectively. **(d)** KS distribution analysis. **(e)** Syntenic depth analyses of *Saxifraga stolonifera* and *Vitis vinifera* genomes.

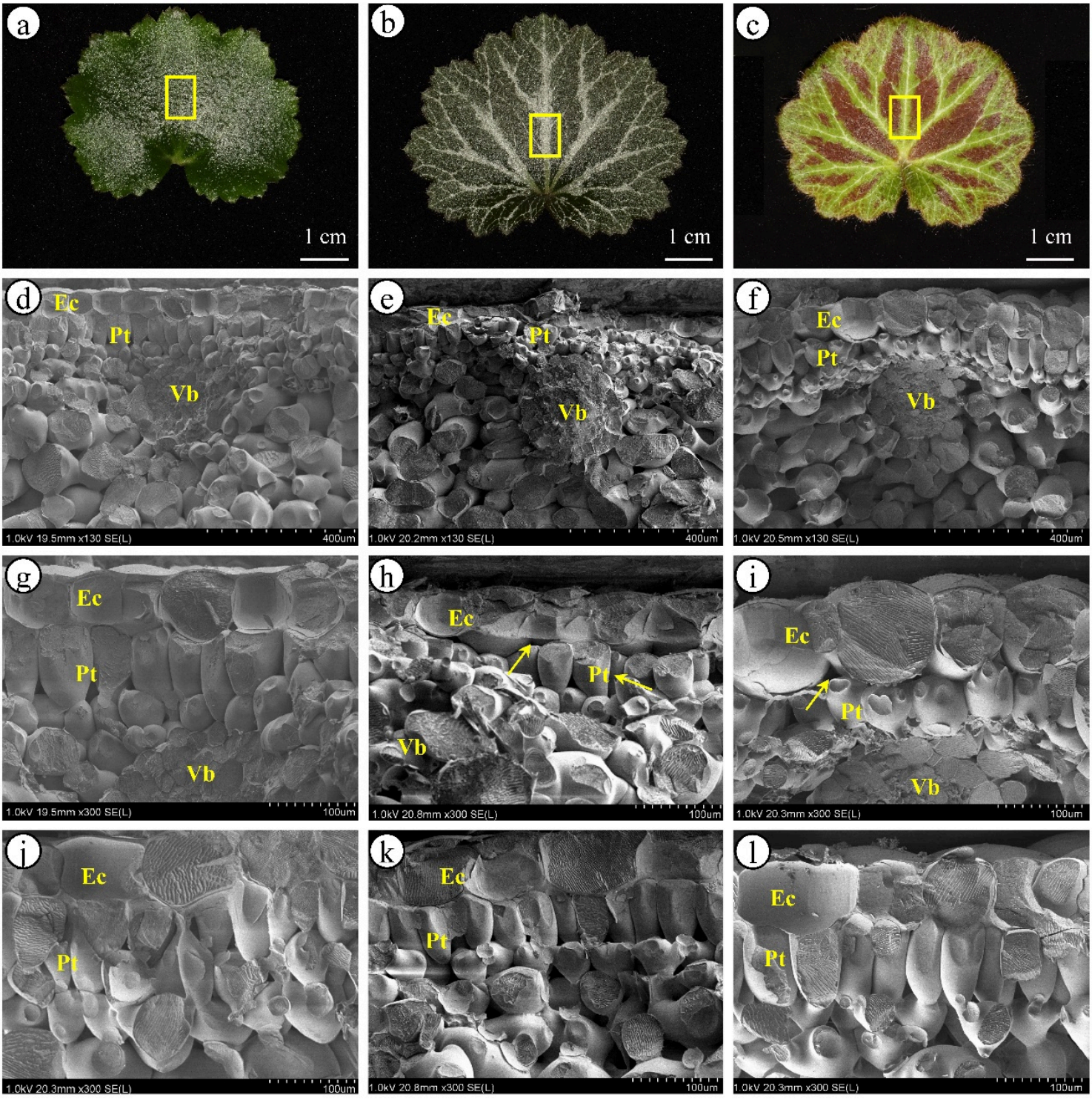

Figure 4

Transverse sections of different *S. stolonifera*leaves under cold-field emission scanning electron microscopy, showing air spaces. **(a–c)** sampling position (from left to right, the green-leaf type G, white-variegated type W, purple-variegated type P, the same below). **(d–f)**leaf transverse section (130×). **(g–**

i)transverse section of the leaf vein area (300×). **(j–l)** transverse section between the leaf vein area (300×). **Ec**, epidermal cells. **Pt**, palisade tissue. **Vb**, vascular bundle. Arrows in h and i indicate air spaces.

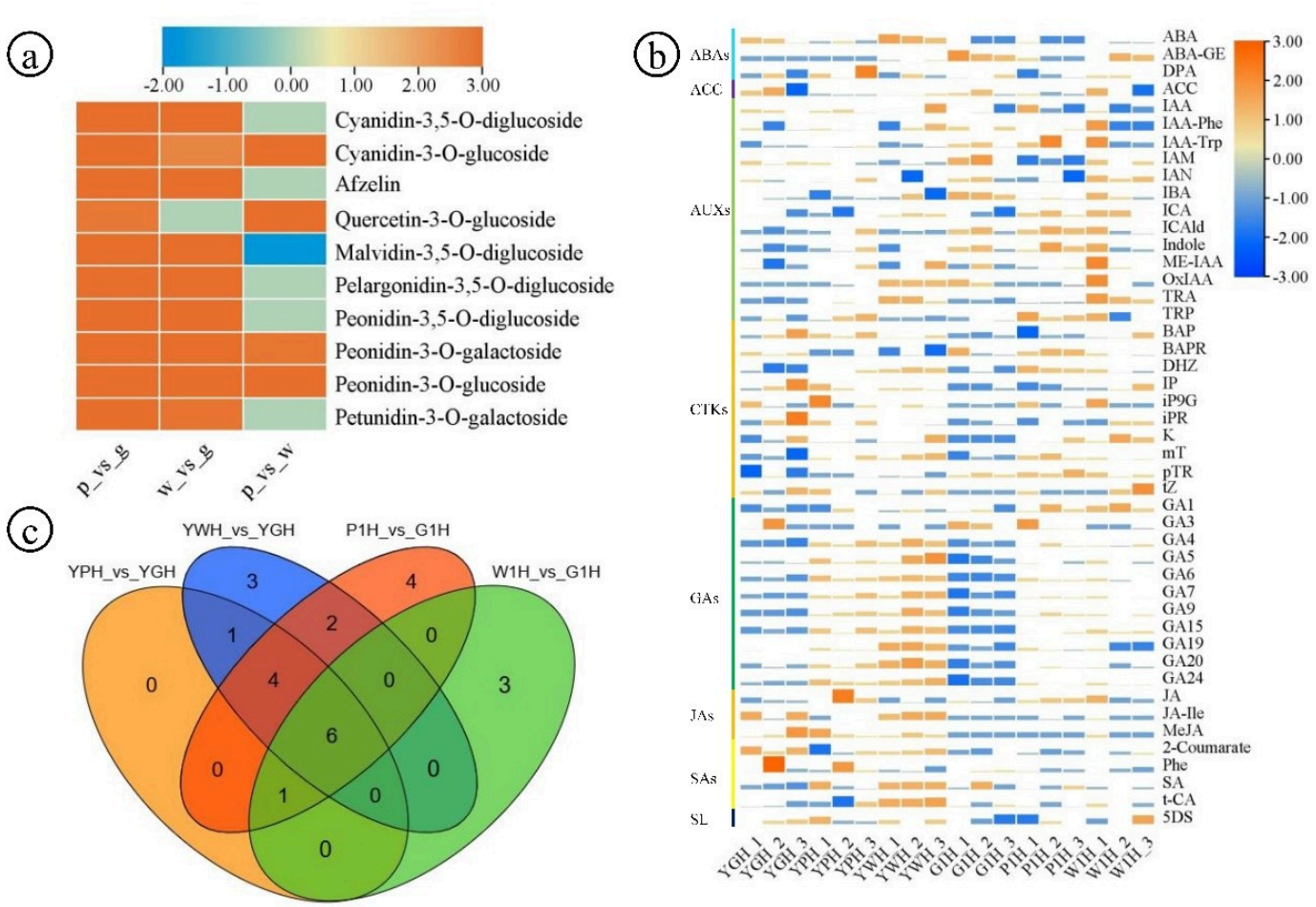

Figure 5

Metabolomic profiling of anthocyanins and plant hormones. **(a)** Heatmaps of differentially abundant metabolites (DAMs) in the three compared phenotypes. **(b)** Heatmaps of plant hormones in three leaf phenotypes. **(c)** Venn analysis of differential plant hormones in four comparison groups. **g**, second-stage adult leaf of the green-leaf type; **p**, second-stage adult leaf of the purple-variegated type; **w**, second-stage adult leaf of the white-variegated type.

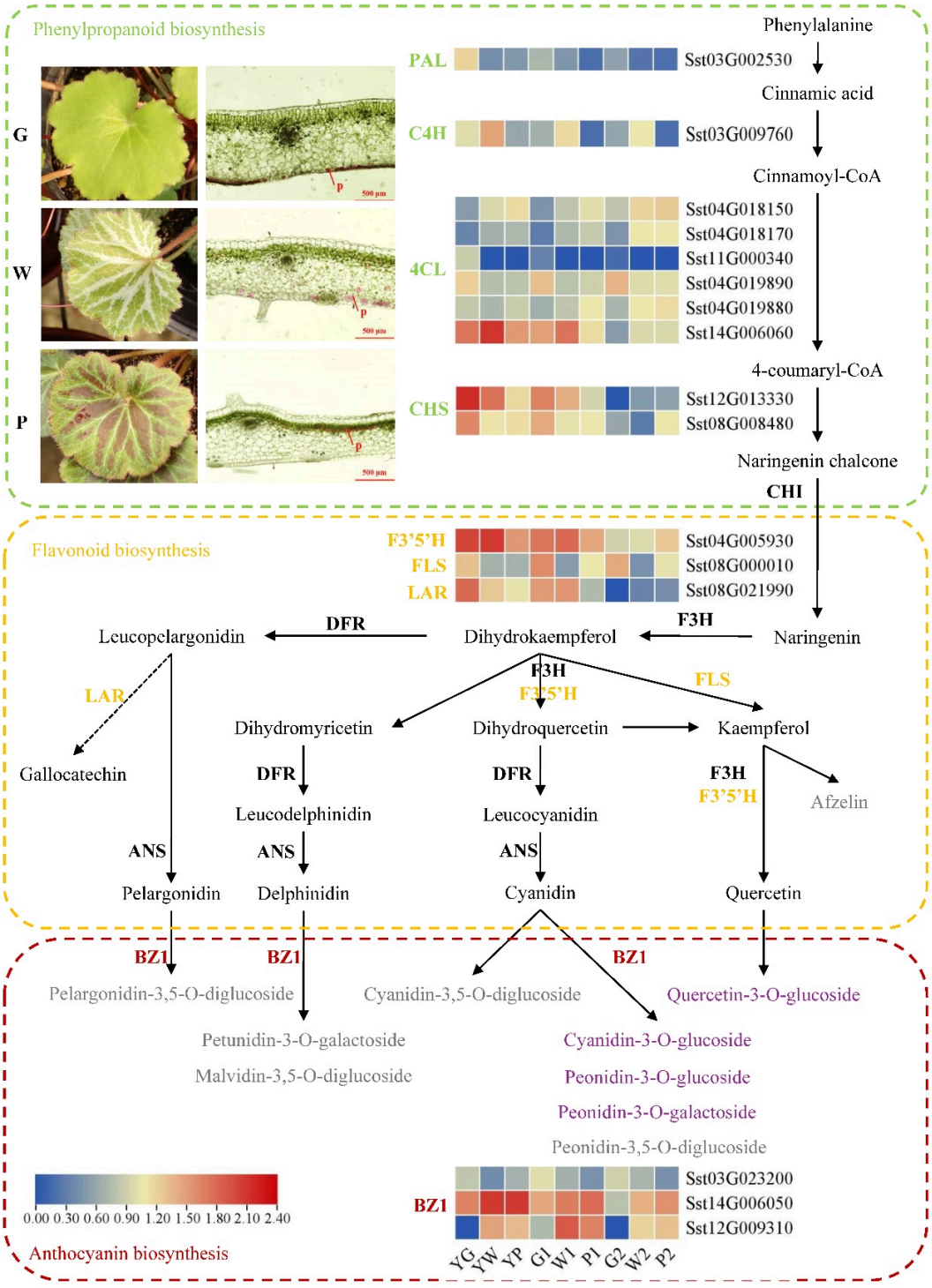

Figure 6

Anthocyanin biosynthetic pathways and gene expression in three leaf phenotypes. Gene expression levels (log_{10} (TPM + 1)) in the three leaf phenotypes are represented by color grading, where TPM = the average value of each group of all samples. **G**, green-leaf type; **P**, purple-variegated type; **W**, white-variegated type.

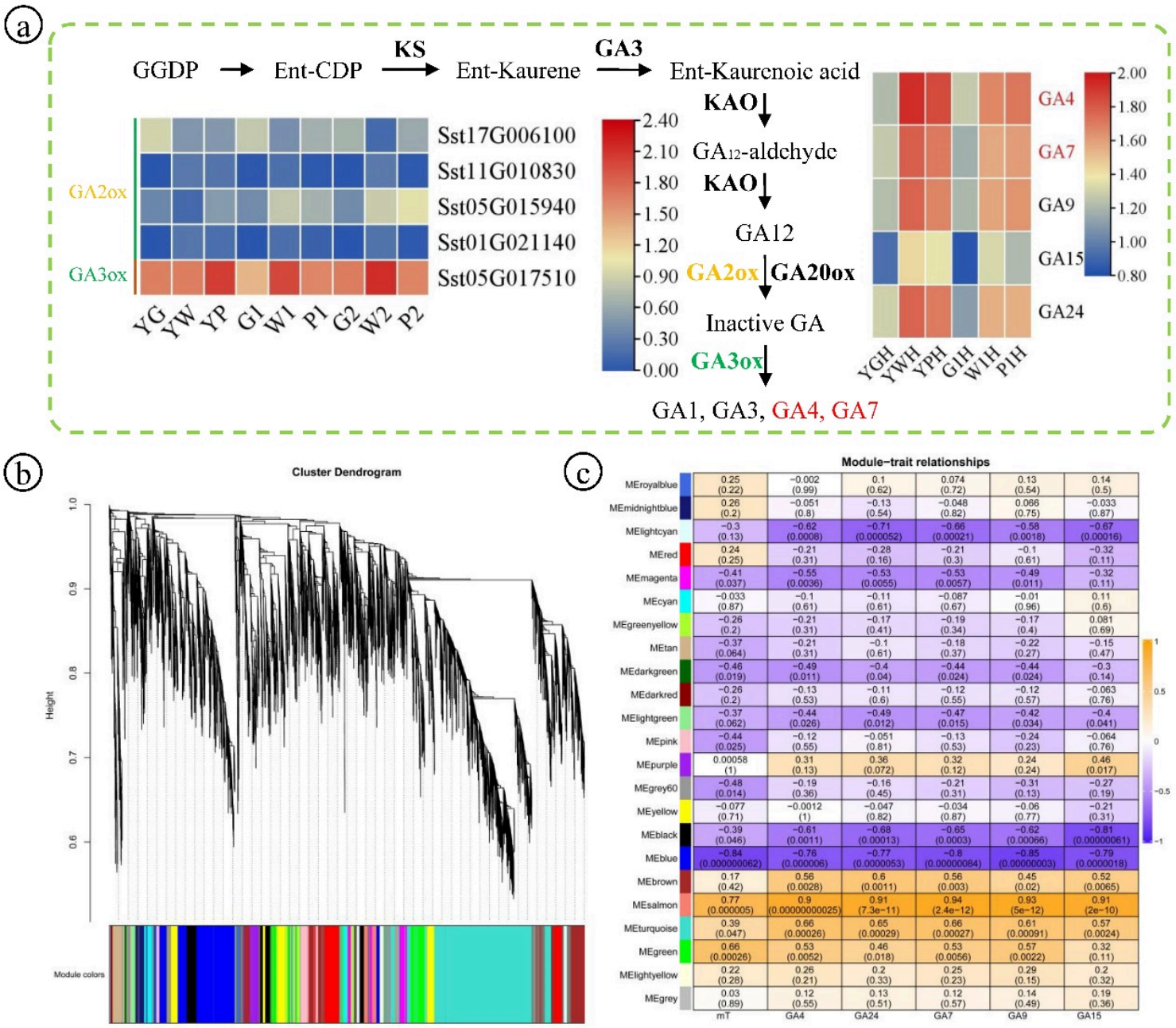

Figure 7

Transcriptomic and plant hormonecorrelation analysis in three leaf phenotypes. **(a)** Map and heatmaps of the gibberellin biosynthesis pathway. **(b)** Dendrogram showing co-expression modules identified by WGCNA across plant hormones in three leaf phenotypes. **(c)**Heatmap displaying the correlations between gene expression modules and gibberellin biosynthesis. Each of the 23 rows corresponds to a specific module (ME) indicated by a distinct color.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 20260419SupplementaryFig.docx
- 20260312SupplementaryTable.xlsx