nature genetics

Article

https://doi.org/10.1038/s41588-026-02569-z

Pangenomic analyses of rose uncover widespread structure variation and empower genomics-directed breeding

Received: 13 December 2024

Accepted: 9 March 2026

Published online: 16 April 2026

Check for updates

Xiaoni Zhang[1,2,3,4,8], Lan Lan[1,8], Yingxue Yang[1,3,8], Huilin Guan[2,8], Dan Peng[1], Heling Jiang[1], Quanshu Wu[2], Rui Huang[1,5], Xuezhu Liao[1], Shengnan Lin[1], Desheng Gong[1], Bingyao Huang[2], Clement Bellot[6,7], Judit Szécsi[6,7], Manzhu Bao[2], Mohammed Bendahmane[2,6,7] ✉, Xiaopeng Fu[2,6] ✉, Zhiqiang Wu[1,3,4,5] ✉ & Weihua Pan[1] ✉

Roses are economically important ornamental plants, with widespread applications in the cut flowers, garden and cosmetics industries. The genomic evolution and diversity of the subgenus *Rosa* remain understudied, limiting exploitation of its diversity in breeding. Here we assembled genomes of 23 accessions, comprising 51 haplotypes that capture the subgenus's high genetic diversity. Extensive introgression across accessions from different sections highlights crossbreeding potential. Pangenome analysis revealed 1,801,537 structural variations, providing insights into the genetic basis and key regulators controlling key traits such as continuous flowering, petal number and discoloration. A key finding was the identification of a *CCD4* homolog as the main regulator of petal discoloration. Additional subgenomic analysis of the allopolyploid *Rosa gallica* and the triploid *Rosa hybrida* 'La France', two important breeding materials of modern roses, revealed their hybrid origins. Overall, this study advances understanding of rose genomics and provides valuable resources for future breeding and trait improvement.

Modern roses are among the most cherished and economically valuable ornamental plants worldwide, with an estimated global market value of approximately US$3 billion (https://oec.world/en/profile/hs/roses#product-complexity). Roses display remarkable variation in features such as flower color, size, form and scent, making them valuable both ornamentally and commercially. Their cultivation now also extends to pharmaceutical, medicinal and cosmetic uses.

The diversity of traits in modern roses results from the intensive interspecific and intraspecific hybridization over a long breeding history, although much of the genetic material of the subgenus *Rosa* remains underused. Chinese roses, key progenitors of modern roses, have been cultivated in China[1] for over 2,000 years and were one of the first genetic resources for rose breeding. *Rosa chinensis* 'Old Blush' (OB), *Rosa odorata* 'Parks' Yellow Tea-scented China' and *R. odorata* 'Hume's Blush Tea-scented China' had pivotal roles in initiating the first revolution in rose breeding. European species, like *Rosa gallica* and *Rosa phoenicia*, have also made substantial contributions to modern roses. Several published chromosome-scale genomes, including those of *R. chinensis* OB[2] and 'Chilong Hanzhu' (CH)[3], *Rosa rugosa*[4,5], *Rosa wichuraiana*[6] and *Rosa multiflora*[7], have provided valuable resources for rose functional genomics. However, the limited number of reference genomes limits the ability to fully explore the extensive trait diversity and genomic architecture across the genus *Rosa*.

Pangenomes, incorporating genomic data from a diverse range of individuals, have proven effective in capturing the genetic diversity within a species or genus[8,9]. Comparative analyses of de novo genome assemblies from representative individuals can reveal molecular mechanisms underlying key traits unexplained by small-scale variations[10–12].

A full list of affiliations appears at the end of the paper. ✉e-mail: mohammed.bendahmane@ens-lyon.fr; fuxiaopeng@mail.hzau.edu.cn; wuzhiqiang@caas.cn; panweihua@caas.cn

Structural variants (SVs), in particular, explain 'missing heritability' in expression quantitative trait loci more frequently than anticipated due to their larger phenotypic effect compared to SNPs and short indels[13–15]. Despite their utility, pangenomic resources for roses remain scarce, hindering progress in studies of trait diversity and evolution.

The genus *Rosa* comprises over 200 species, most of which belong to the *Rosa* subgenus. According to the published studies, the *Rosa* subgenus is subdivided into the following ten sections: *Pimpinellifoliae*, *Gallicanae*, *Synstylae*, *Banksianae*, *Laevigatae*, *Bracteatae*, *Cinnamomeae* (also known as section *Rosa*), *Caninae*, *Chinenses* and *Carolinae*, the last of which is native to eastern and central North America[16,17]. An additional section, *Microphyllae*, has been described as native to China[18]. In this study, we assembled 51 haplotype-resolved assemblies from 23 representative Chinese and European/Middle-Eastern *Rosa* accessions. By integrating these newly assembled genomes with the published ones of *R. chinensis* CH[3] and *R. rugosa*[4,5], we constructed a pangenome for the *Rosa* subgenus, encompassing 26 accessions. We identified several genetic variants, including presence-absence variations (PAVs) and inversions, that are associated with ornamental traits. Moreover, extensive introgression among intersection accessions within *Rosa* subgenus underscores the potential of cross-sectional breeding strategies. We also explored the subgenomic features of the tetraploid *R. gallica* and the triploid hybrid *Rosa hybrida* La France, both of which are valuable parental candidates for rose improvement. Collectively, these findings offer a comprehensive foundation for understanding the molecular and genetic mechanisms underlying key ornamental traits and will facilitate accelerated breeding and cultivar development in roses.

Results

De novo genome assembly and annotation of 23 *Rosa* accessions

To support future rose breeding efforts and explore the genomic diversity within the *Rosa* subgenus, we selected 23 *Rosa* accessions spanning nine sections from Asia and Europe[16] (Supplementary Table 1). This diverse panel encompasses (1) nine taxonomically representative rose accessions from the *Chinenses*/*Indicae*, *Banksiae*, *Cinnamomeae*, *Bracteatae*, *Microphyllae*, *Laevigatae*, *Pimpinellifoliae*, *Synstylae* and *Gallicanae* sections recorded in the 'Flora of China'; (2) historically important cultivars with diverse ploidy levels, including the early diploid hybrid 'Pink Noisette', the first triploid modern rose La France and the tetraploid modern rose 'Carola'; and (3) 11 botanical Chinese roses, among which are 3 ancient cultivars (*R. odorata* Parks' Yellow Tea-scented China, *R. odorata* Hume's Blush Tea-scented China and a newly updated haplotype-resolved *R. chinensis* OB genome[2]), previously reported to contribute to modern cultivar diversity[19], and some accessions harboring underused traits of potential breeding value, such as petal discoloration in 'Butterfly', miniature flower size in 'Minima' and sepal overgrowth at the pistil in 'Sai Zhao Jun'.

Using a combination of the HiFi and Hi–C sequencing, we generated 51 haplotype-resolved genome assemblies across the 23 accessions (Fig. 1a and Supplementary Table 1), with haplotype sizes ranging from 355.9 Mb to 639.7 Mb (Table 1 and Supplementary Fig. 1). The assemblies had an average contig N50 of 29.2 Mb, and for 20 diploid accessions, approximately 180 contigs per haplotype were successfully anchored to seven pseudochromosomes. Gene annotation, performed through an integrative approach combining transcriptomic evidence, homology-based prediction and ab initio methods, identified between 24,394 and 40,653 protein-coding genes, with an average of 30,856 genes (Supplementary Table 2). Transposable elements (TEs) accounted for 49.16–66.99% in these haplotypes, with long terminal repeat retrotransposons being the most abundant, mainly Copia (7.8–15.5%) and Gypsy (10.4–15.2%; Supplementary Table 3).

We rigorously evaluated assembly and phasing quality using multiple strategies. First, Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis revealed a high level of completeness, with an average of 97.6% across all haplotypes (including polyploids; Table 1). Second, the long terminal repeat-assembly index averaged 21.06, indicating well-resolved repetitive regions (Supplementary Table 4). Third, Hi–C contact heatmaps confirmed haplotypes-scale contiguity and successful haplotype resolution in both diploid and polyploid accessions (Supplementary Fig. 1). Fourth, *k*-mer spectrum plots indicated minimizing error-mixing assemblies, further supporting the high accuracy of the assemblies (Supplementary Fig. 2). Fifth, the average consensus quality value across all haplotypes was 64.93 (Supplementary Table 4). Sixth, switch error rates were assessed in representative cultivars A02 (diploid) and A13 (tetraploid), yielding low error rates of 1.26% and 3.27%, respectively, indicating high phasing accuracy.

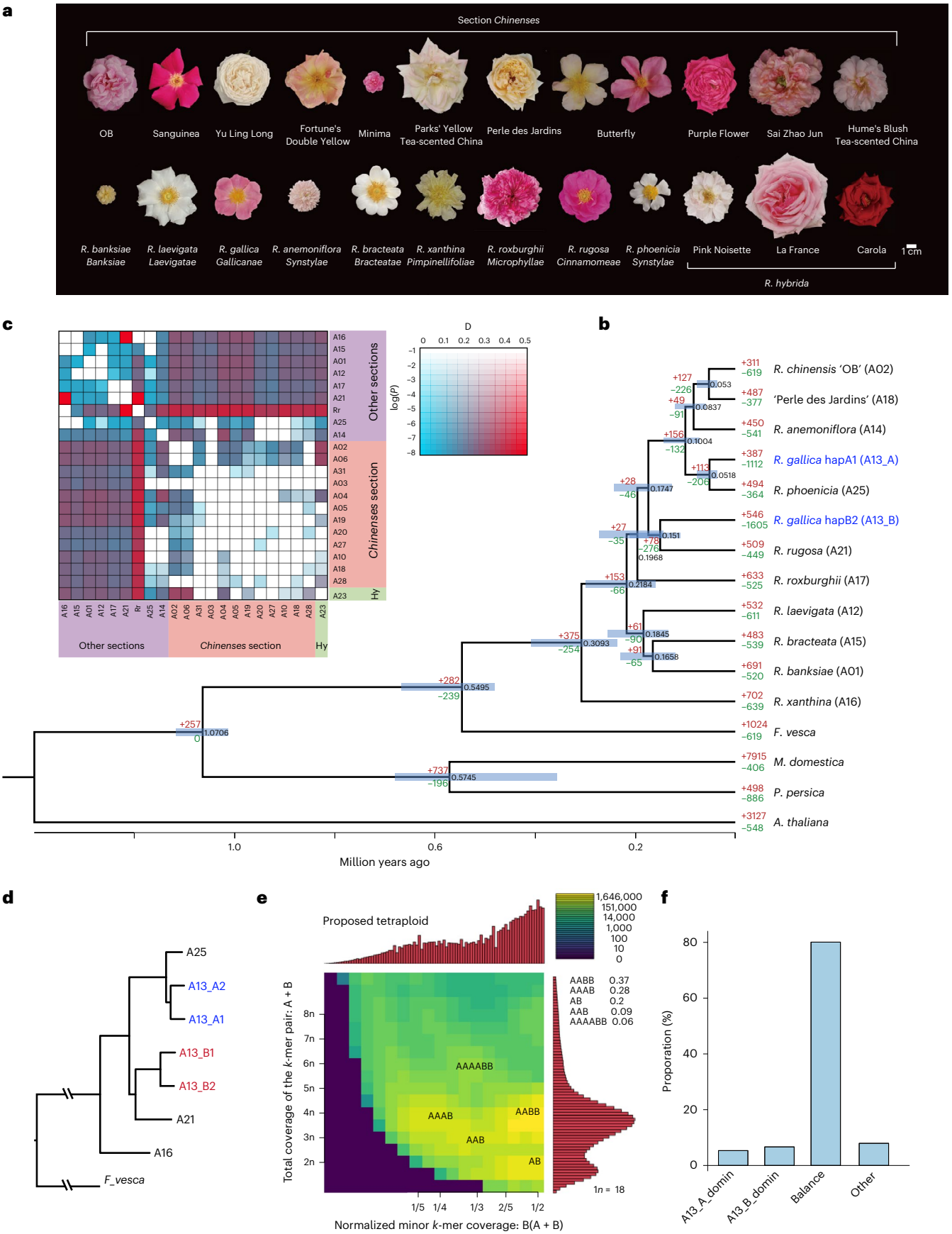
Gene annotations were also evaluated with BUSCO, achieving an average completeness of 97.3% across haplotypes, demonstrating high annotation quality (Supplementary Table 5). Additionally, we compared our haplotype-resolved assembly of OB with two previously published versions, all of which exhibited similarly high BUSCO scores (~98%; Supplementary Table 6). Notably, our new assembly showed improved chromosome lengths and scaffold N50. Although multiple inversions were identified among the three OB assemblies, we confirmed the breakpoints of these variants through HiFi read mapping (Supplementary Fig. 3), demonstrating the structural accuracy and continuity of our assembly. Furthermore, the high concordance between the Bionano mapping and Hi–C scaffolding results clearly demonstrates the high-quality phasing and scaffolding of the newly assembled OB genome (Supplementary Fig. 4).

Phylogenetic evolution

To clarify the evolutionary relationships within the Rosaceae species, we constructed a phylogenetic tree using 1,364 single-copy orthologous genes from 11 newly assembled *Rosa* species, in addition to the genomes of *Fragaria vesca*, *Malus domestica* and *Prunus persica*. *Arabidopsis thaliana* was included as an outgroup (Fig. 1b). Consistent with previous findings[2], *Prunus* and *Malus* diverged earlier from the common ancestor shared by *Rosa* and *Fragaria* within the Rosaceae family. Within the *Rosa* subgenus, *Rosa xanthina*, a representative of the *Pimpinellifoliae* section, diverged first, followed by *R. banksiae*, *R. bracteata*, *Rosa laevigata*, *Rosa roxburghii*, *R. rugosa*, *R. phoenicia*, *Rosa anemoniflora*, *R. odorata* and *R. chinensis* (Fig. 1b). We also observed a considerable variation in the number of expanded and contracted gene families across different species, which likely contributes to species-specific characteristics. Gene Ontology (GO) enrichment analysis of expanded families revealed that genes involved in terpene synthase activity were enriched in the highly fragrant *R. rugosa*, whereas

Fig. 1 | The phylogenetic genomic analysis of the *Rosa* accessions. a, Flowers of *Rosa* genotypes belonging to the nine sections from China and European/Middle-Eastern areas. The flower of A25 was presented from our published paper[40]. **b**, Phylogenetic tree of *Rosa* accessions based on the single-copy genes, displaying the expansion (in red) and contraction (in green) of gene clusters. Nodes highlighted in blue indicate the 95% confidence interval of the estimated divergence times. **c**, Introgression events were detected across accessions. Each colored square represents candidate introgression events between pairs of individuals. Rr represents *R. rugosa*[4]. **d**, Phylogenetic tree of allotetraploid *R. gallica* (A13) constructed using its haplotype genomes. **e**,**f**, Smudgeplots analysis based on *k*-mers (**e**) and expression ratios of different subgenomes (**f**) in *R. gallica*. The *x* axis represents tetrads categorized by different expression biases. The *y* axis indicates the proportion of each type of tetrad to the total tetrads. A13_A_domin refers to tetrads where subgenome A exhibits dominant expression. A13_B_domin refers to tetrads where subgenome B exhibits dominant expression. Balance indicates tetrads with no dominant expression. Other categories (for example, A1_B1_donminate, A1_B2_dominate) include tetrads that do not fit into the three situations above. Hy, hybrids.

https://doi.org/10.1038/s41588-026-02569-z



https://doi.org/10.1038/s41588-026-02569-z

Table 1 | Statistics of genome assemblies of *Rosa* accessions

Species/cultivar	Section/group	Ploidy	Contig N50 (Mb)		Genome size (Mb)		Genome BUSCO (%)		Accessions
			hap1	hap2	hap1	hap2	hap1	hap2	
R. chinensis 'OB'	*Chinenses* DC.	Diploid	31.6	30.6	516.8	538.8	98.5	98.5	A02
R. chinensis 'Sanguinea'	*Chinenses* DC.	Diploid	25.4	28.6	525.9	516.4	98.8	97.7	A03
R. chinensis 'Yu Ling Long'	*Chinenses* DC.	Diploid	22.1	22.5	519.8	530.9	98.0	98.3	A04
R. odorata 'Fortune's Double Yellow'	*Chinenses* DC.	Diploid	15.2	17.3	538.3	515.9	97.9	95.2	A05
R. chinensis 'Minima'	*Chinenses* DC.	Diploid	34.1	33.3	511.3	524.8	97.1	97.6	A06
R. odorata 'Parks' Yellow Tea-scented China'	*Chinenses* DC.	Diploid	45.7	41.4	531.1	520.0	98.6	98.5	A10
R. hybrida 'Perle des Jardins'	*Chinenses* DC.	Diploid	24.1	26.6	520.3	519.2	98.6	98.5	A18
R. chinensis 'Butterfly'	*Chinenses* DC.	Diploid	27.8	38.1	549.5	516.0	98.6	97.8	A19
R. chinensis 'Purple Flower'	*Chinenses* DC.	Diploid	26.7	26.0	520.1	528.6	98.5	98.5	A20
R. chinensis 'Sai Zhao Jun'	*Chinenses* DC.	Diploid	37.9	49.9	534.9	524.3	98.5	98.5	A27
R. odorata 'Hume's Blush Tea-scented China'	*Chinenses* DC.	Diploid	28.6	28.7	536.1	526.1	98.5	98.4	A28
R. banksiae f. *lutea*	*Banksiae* Lindl.	Diploid	30.8	27.4	464.0	462.4	98.2	97.9	A01
R. rugosa Thunb. 'DaGuo'	*Cinnamomeae* DC.	Diploid	36.4	39.8	451.3	447.7	98.7	98.5	A21
R. hybrida 'Pink Noisette'	Hybrid	Diploid	47.7	18.7	543.0	530.1	98.6	98.5	A23
R. bracteata	*Bracteatae* Thory	Diploid	12.5	19.7	511.0	511.4	98.4	98.5	A15
R. roxburghii	*Microphyllae* Crep	Diploid	36.0	26.6	495.4	495.9	96.8	97.0	A17
R. laevigata	*Laevigatae* Thory	Diploid	36.8	47.4	491.8	475.6	98.7	97.6	A12
R. xanthina 'Double Yellow'	*Pimpinellifoliae* DC.	Diploid	20.1	30.2	357.3	355.9	98.4	98.3	A16
R. anemoniflora 'Plena'	*Synstylae*	Diploid	37.4	46.1	541.1	538.2	98.3	98.5	A14
R. phoenicia	*Synstylae*	Diploid	30.0	32.5	639.7	636.2	98.4	98.4	A25
–	–	–	Conitg N50 (Mb)		hap A1	hap A2 / hap B(1) / hap B2	hap A1	hap A2 / hap B(1) / hap B2	–
R. hybrida 'La France'	Hybrid	Triploid	10.7		504.7	487.0 / 449.6 / –	95.8	94.3 / 97.6 / –	A11
R. gallica L.	*Gallicanae*	Tetraploid	5.1		504.1	493.6 / 392.3 / 398.4	96.4	97.1 / 93.3 / 94.4	A13
R. hybrida 'Carola'	Hybrid	Tetraploid	1.2		503.5	478.8 / 459.3 / 482.9	94.5	93.4 / 90.6 / 88.3	A29

genes related to drought response (such as 'oxidoreductase activity' in hapA and 'response to water' in hap B) were enriched in *R. gallica* (A13; Fig. 1b and Supplementary Fig. 5).

Introgression analysis revealed that gene flow predominantly occurred across accessions from different sections, with extensive introgression observed among intersectional accessions involving members of the *Chinenses* section. By contrast, moderate levels of introgression were detected across accessions within other sections, and only minima introgressions occurred within the *Chinenses* section itself (Fig. 1c). Furthermore, demographic inference suggested that the *Chinenses* section experienced a population expansion between approximately 2,000 and 5,000 years ago, while accessions from other sections underwent a continuous decline in population size (Supplementary Fig. 6). Collectively, these findings reveal a complex evolutionary history for the *Rosa* subgenus, shaped by both extensive introgression events and distinct demographic trajectories.

Subgenome features

To further explore the subgenomic features of tetraploid *R. gallica*, we reconstructed haplotype-level phylogenies for all chromosomes in *R. gallica*, a major progenitor of modern roses. Based on *k*-mer signature (Supplementary Fig. 7) and single-copy gene clusters (Supplementary Fig. 8), the phylogenetic analysis revealed that the *R. gallica* haplotypes could be divided into two subgenomes. Subgenome A (haplotypes A1 and A2) is close to *R. phoenicia* (A25), while subgenome B (haplotypes B1 and B2) is more closely related to *R. rugosa* (A21; Fig. 1b,d). To identify the origins of these subgenomes, genomic segments from *R. gallica* were mapped to the 'reference genomes' combined from ten candidate contributors (A02, A18, A14, A25, A21, A17, A12, A15, A01 and A16; Supplementary Fig. 9). In the haplotype A1, the majority of genomic segments mapped to A25_2 (~503.9 Mb, 99.96%), with a small fraction mapping to A25_1 (~0.2 Mb, 0.04%), indicating that A1 likely originated from A25_2. For haplotype B2, the segments were primarily mapped into the A21_2 (~338.26 Mb, 84.9%), with smaller contributions from A21_1 (~36.16 Mb, 9.08%), A25_2 (~23.78 Mb, 5.97%) and A25_1 (~0.2 Mb, 0.05%), supporting A21_2 as the main contributor for B2 (Supplementary Fig. 9). Consistent with these results, Smudgeplot analysis confirmed that of *R. gallica* possesses an AABB genome structure (Fig. 1e). Gene density and gene expression profiles across haplotypes were largely comparable (Fig. 1f and Supplementary Fig. 10). Expression data from four tissues (stems, leaves, roots and flower buds) were analyzed across 9,726 syntenic tetrads in *R. gallica*. Among these, 80.08% showed no significant expression bias, while 5.33% and 6.66% showed subgenome dominance in subgenome A and subgenome B, respectively (Fig. 1f and Supplementary Fig. 11). At the chromosome level, no consistent single-homeolog dominance was observed across chromosomes 1 to 7 (Supplementary Fig. 12). A similar analysis of the triploid hybrid *R. hybrida* La France (A11) indicated an AAB genome structure (Supplementary Fig. 13). Although no consistent single-homeolog dominance at the chromosome level, the number of syntenic triads (analogous to the syntenic tetrads in A13) with single-homeolog

dominance in subgenome A was higher than that in subgenome B (Supplementary Figs. 14 and 15).

To explore intragenomic divergence, we compared the two haplotypes of diploid accession A02 at the levels of genetic variations, gene expression and histone modifications. We identified 62 inversions, 2,167 translocations, 2,040,445 SNPs, 147,497 insertions and 145,378 deletions between the haplotypes. Additionally, SyRi-defined unique regions spanned approximately 108 Mb and 128 Mb in haplotypes 1 and 2, respectively (Supplementary Fig. 16a). Between 33% and 43% of the 12,903 alleles exhibited allele-specific expression across six floral developmental stages (Supplementary Fig. 16b). Genes with allele-specific expression exhibited higher nonsynonymous-to-synonymous substitution (d_N/d_S) ratio than allele-balanced expression genes (Supplementary Fig. 16c), suggesting different selection pressures. Histone modification analysis identified 19,070 and 16,819 genes marked with histone 3 Lys9 acetylation (H3K9ac), and 10,773 and 8,298 genes marked with trimethylation of Lys27 on histone H3 (H3K27me3) in haplotypes 1 and 2, respectively (Supplementary Fig. 16d,e). H3K9ac peaks averaged 587.7 bp and 582.7 bp, and were primarily located near transcription start sites, while H3K27me3 peaks averaged 4,354.5 bp and 4,202.1 bp in haplotypes 1 and 2, respectively (Supplementary Fig. 16f). These results collectively demonstrate substantial genomic, transcriptomic and epigenetic variation across haplotypes, underscoring the value of haplotype-resolved assemblies for capturing the full extent of rose genomic diversity.

Core and dispensable genes

We constructed the rose pangenome using 26 *Rosa* accessions that spanned most regions of Asia and the Middle East. This dataset included 23 newly assembled genomes and 3 previously published assemlies[3–5] (Supplementary Table 1). A total of 55,689 gene clusters were identified, including 16,844 core gene clusters, 4,030 softcore gene clusters, 30,714 dispensable gene clusters and 4,101 private gene clusters (Fig. 2a–c). As the genomes were incorporated, the number of core gene clusters progressively declined and eventually plateaued, suggesting that our pangenome has reached near saturation in gene content (Fig. 2b). Selective pressure analysis showed that the average d_N/d_S ratios of core and softcore genes were lower than those of dispensable genes (Fig. 2d), indicating that core and softcore genes are more evolutionarily conserved. Functional annotation results revealed that approximately 90% core and softcore genes could be annotated, compared to ~50% dispensable genes and only 44% private genes (Fig. 2e). Gene expression analysis demonstrated a descending expression trend—core genes exhibited the highest expression levels, followed by softcore, dispensable and private genes (Fig. 2f and Supplementary Table 7). GO and KEGG enrichment analyses revealed that the core and softcore genes were primarily enriched in essential, such as the cellular metabolic processes (Supplementary Tables 8 and 9), whereas dispensable genes were more frequently associated with functions such as genetic information processing, signaling, cytoskeletal dynamics and cellular structure homeostasis.

Pan-variation information statistics

We conducted a pan-variation analysis across 55 haplotypes, comprising 23 newly assembled accessions and 3 previously published genomes. Using one haplotype of OB (A02_2) as the reference, we identified a nonredundant set of 1,738,269 PAVs, 1,567 inversions and 61,701 translocations (Fig. 3a). The number and type of SVs varied substantially across accessions, particularly among hybrids, members of the *Chinenses* section and accessions from other sections. Notably, minimal variation was observed between *R. chinensis* OB and its closely related taxa, including Minima (A06) and the hybrid Pink Noisette (A23; Fig. 3a). While PAVs were distributed throughout the genome, certain regions exhibited concentrated variation, forming distinct hotspots for PAVs, inversions and translocations (Fig. 3b). GO enrichment analysis of genes located in these hotspot regions revealed functional distinctions—PAV-associated and translocation-associated genes were enriched in inhibitor and protein activity terms, respectively, whereas inversion-associated genes were predominantly involved in biochemical reactions, enzyme activities, molecular binding and related biological processes (Supplementary Table 10).

Large inversions have been demonstrated to suppress recombination by reducing the frequency of crossover events, thereby influencing phenotypic traits[20]. To further investigate this, we constructed a map of large-scale inversions across 55 haplotypes and identified 79 inversions exceeding 1 Mb in size (Fig. 3c), and 1,488 inversions smaller than 1 Mb (Supplementary Fig. 17). The three largest inversions, exclusive to hybrids and *Chinenses* section accessions, included two adjacent inversions on chromosome 2 (~2 Mb and ~15 Mb), and one on chromosome 3 (~4 Mb). These inversions were validated by Hi–C contact maps and HiFi read alignments (Supplementary Fig. 18). Of particular interest, the *MIZU-KUSSEI* gene, which regulates root development[21], is located approximately 3 kb proximal to the breakpoint of the 15-Mb inversion on Chr2 (Supplementary Fig. 19), and the *KSN* allele (*TERMINAL FLOWER 1* (*TFL1*)*-like* gene), associated with continuous flowering (CF)[22], is located at the breakpoint of the 2-kb inversion on Chr3, with its absence correlating strongly with the presence of the inversion (Supplementary Fig. 20).

To assess the functional effect of SVs, we examined their overlap with genic regions. In the *Rosa* pan-variation map, approximately 54.37% inversions overlapped with genes, and 52.9% with exons. For translocations, 13.25% intersected gene regions, while 30.81% PAVs were located within gene regions, and only 5.75% within exons (Supplementary Fig. 21a), suggesting that larger structural variations may have more substantial effects than shorter PAVs. We further identified 25,909 nonredundant genes that overlap with PAVs, and 13,208 and 6,174 genes that overlap with inversions and translocations, respectively. The majority of PAVs were shorter than 100 bp, while most inversions and translocations ranged between 500 and 2,000 bp (Supplementary Fig. 21b). Based on their presence and absence patterns across accessions, most variations were classified as dispensable, with a smaller proportion categorized as private (Supplementary Fig. 21c). Previous studies have linked TEs with structural variation[23,24]. In our pan-variation map, ~30% of all variants contained no TEs, whereas 70% were associated with TE insertions. Among these, ~30% were TE-amplified (Supplementary Fig. 22). Additionally, we constructed a reference-free graph-based pangenome for the 26 accessions. The observed γ values ranged from 0.67 to 0.81, and the large number of graph nodes reflected the high structural complexity and diversity within the *Rosa* subgenus (Supplementary Table 11).

Functional effect of SVs on ornamental traits

Continuous flowering. Seasonal flowering is a key ornamental trait in rose breeding. To explore structural variations associated with this trait, we analyzed gene and promoter-level variations using the *Rosa* pangenome. Accessions within the *Chinenses* section showed similar allelic compositions at the *RcKSN* gene, a locus previously reported to be associated with the CF trait[25]. These accessions harbored genotypes such as $RcKSN^{copia}/RcKSN^{copia}$, $RcKSN^{copia}/RcKSN^{null}$ or $RcKSN^{null}/RcKSN^{null}$. By contrast, all seven once-flowering (OF) accessions from other sections consistently carried out the $RcKSN^{wt}/RcKSN^{wt}$ genotype (Fig. 4a). Notably, an inversion event was consistently observed to co-occur with the presence of the *RcKSN* gene, indicating a tight linkage between the inversion and the CF trait (Fig. 4b). Within the *RcKSN* locus, we identified nine candidate variants by this pangenome that may influence the CF phenotype, including five SNPs, two indels and two SVs involving TE insertions, one of which was previously reported[22] (Fig. 4a). These TE insertions were validated using Illumina sequencing and PCR amplification (Fig. 4c and Supplementary Figs. 23 and 24). Furthermore, one of

https://doi.org/10.1038/s41588-026-02569-z

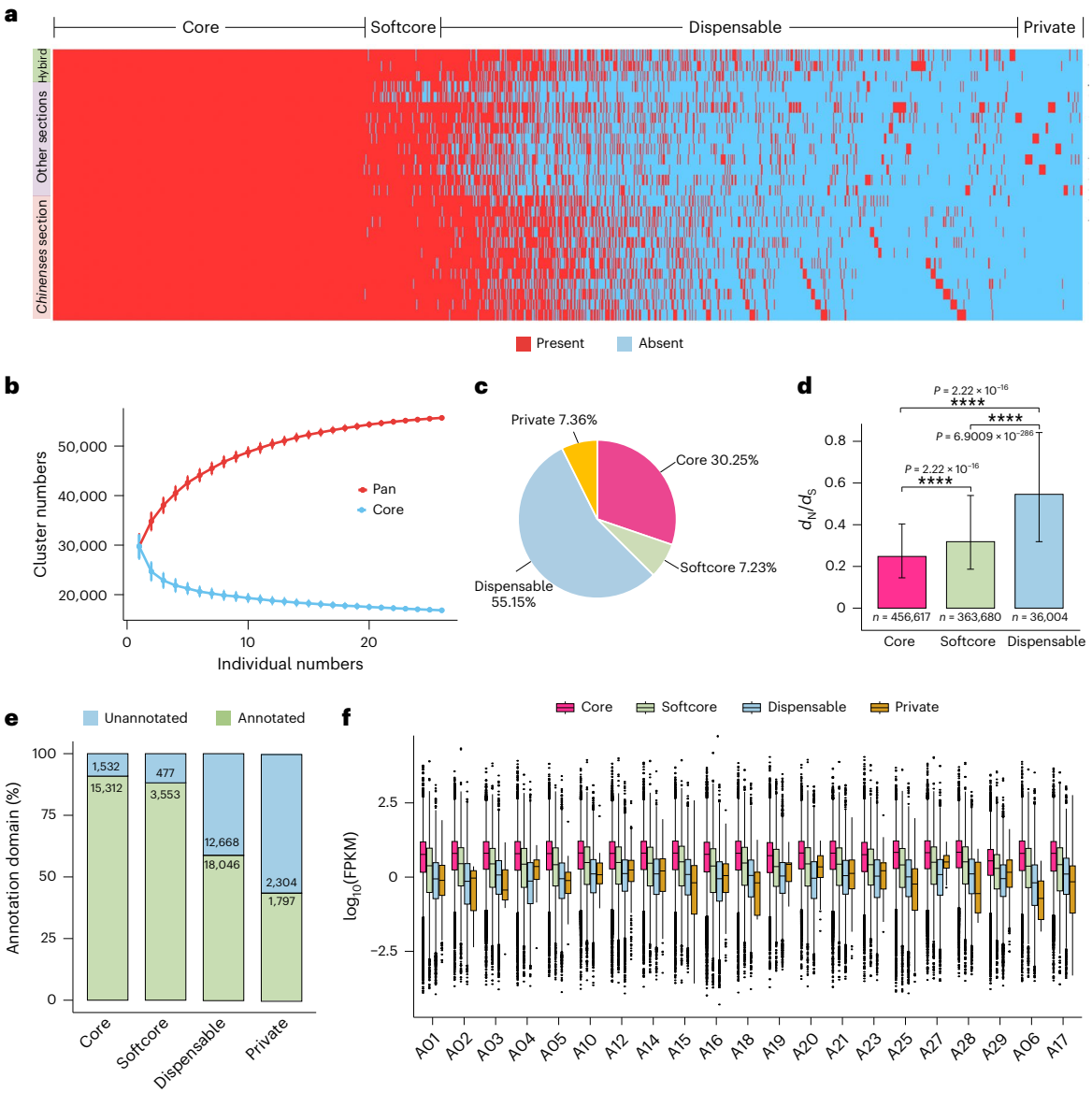


Fig. 2 | The pangenome of the *Rosa* subgenus. a, The presence and absence of different gene clusters presented in different rose individuals. **b**, Variation in gene clusters within the *Rosa* pangenome as additional haplotypes from different individuals were added. **c**, Proportion of different types of gene clusters within the *Rosa* pangenome. **d**, Calculated d_N/d_S values for different types of gene clusters, with asterisks indicating statistical significance, bars representing the sample median and error bars indicating the 25th and 75th percentiles (interquartile range, IQR). ****$P < 0.0001$. **e**, Number of gene clusters annotated by various databases for different gene cluster types. **f**, Expression levels of different gene cluster types (genes with fragments per kilobase transcript per million mapped reads (FPKM) values of 0 were deleted) across various individuals, box plots show the median (center line), IQR (25th–75th percentiles; box), whiskers extending to 1.5× IQR, and individual data points overlaid as dots.

these four SNPs identified by the rose population analysis, located at position 19,143,785 (Chr3_2) in the genome, predominantly exhibits a 'G' in OF accessions and an 'A' in CF accessions. All four SNPs also showed stable associations with the CF trait, suggesting a potential functional role in regulating *KSN* expression (Fig. 4d).

Double flower. Petal number is a crucial determinant of floral morphology and ornamental value. Roses can have simple flowers with five petals, but they can also have double flower (DFs) with a large number of petals (more than ten). A key gene underlying this trait has been characterized in *Rosa* species[26]. Based on the pan-variation information, we confirmed the previously reported ~10-kb TE insertion in the eighth intron of the *APETALA2-like* (*AP2L*) allele, which is known to be associated with DF formation in *Chinenses* section accessions (Fig. 5a). This insertion appears to be specific to DF *Chinenses* roses and was further validated in modern cultivars by PCR (Fig. 5a). Therefore, this TE insertion could be considered as a valuable marker for tracing in modern double-flowered roses derived from the *Chinenses* section.

AP2L homologs across Rosaceae, which are known to regulate double-petal formation[26–28], exhibit diverse variations, but they all retain the capacity to influence microRNA172 (miR172) target binding. For instance, *AP2L* homolog *TARGET OF EAT* (*TOE*)[27] in *Prunus mume* (Mei) has a 49-bp deletion (16 amino acids) that removes the miR172 target site. In peach, the *Prupe.6G242400* allele lacks the entire tenth exon (994-bp deletion) and consequently eliminating the miR172 binding site[28]. We screened our pangenome and found that, in double-petal *R. rugosa*, the *RrAP2L*-encoding sequence was prematurely terminated, resulting in the deletion of the miR172 target site. The SV was further confirmed by the direct mapping results from HiFi data (Supplementary Fig. 25). By contrast, the *AP2L* homologs in

https://doi.org/10.1038/s41588-026-02569-z

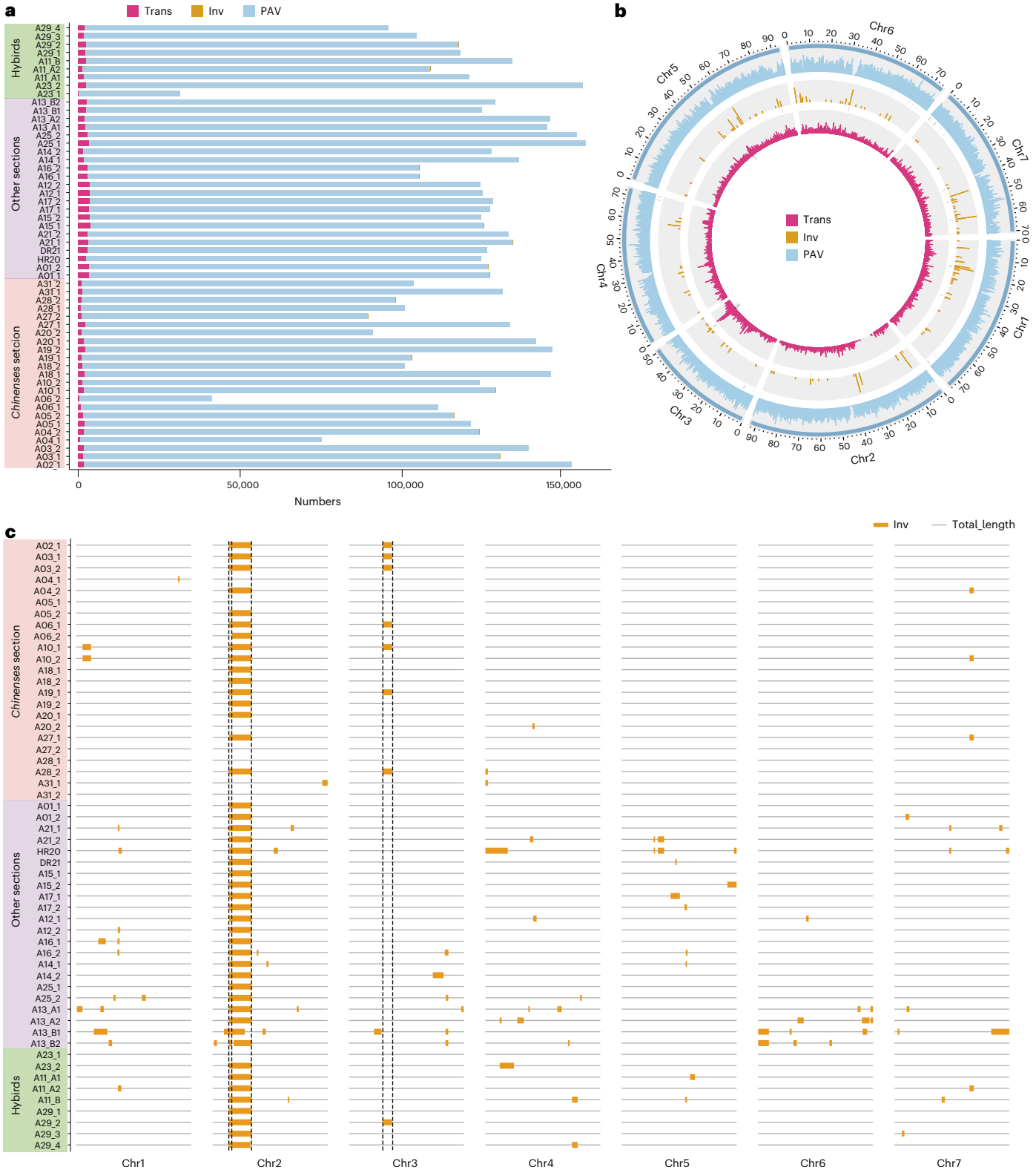


Fig. 3 | The pan-variation of the *Rosa* accessions. a, The identified variation types between different haplotype-resolved genomes and *R. chinensis* OB (A02_2) haplotype. **b**, The circos plot of the total pan-variation in *Rosa* based on the A02_2. From the outermost ring to the innermost, the plot represents chromosome characteristics, PAV distribution rate, PAV hotspot regions, inversion distribution rate, inversion hotspot regions, translocation distribution rates and translocation hotspot regions. The window size was set to 10 kb. **c**, A map of large-scale inversions (>1 Mb) across *Rosa* accessions. Inv, inversion; Trans, translocation.

R. xanthina, *R. banksiae*, *R. bracteata*, *R. anemoniflora*, *R. laevigata* and *R. roxburghii* all have ten exons and the miR172 target site (Fig. 5b), but *R. xanthina*, *R. banksiae*, *R. anemoniflora* and *R. roxburghii* have DFs. We found multiple variations in their introns that may be associated with DF formation in these species, in particular, an SNP at the end of intron 9 (Supplementary Fig. 26a). In *Chinenses* section roses, the

https://doi.org/10.1038/s41588-026-02569-z

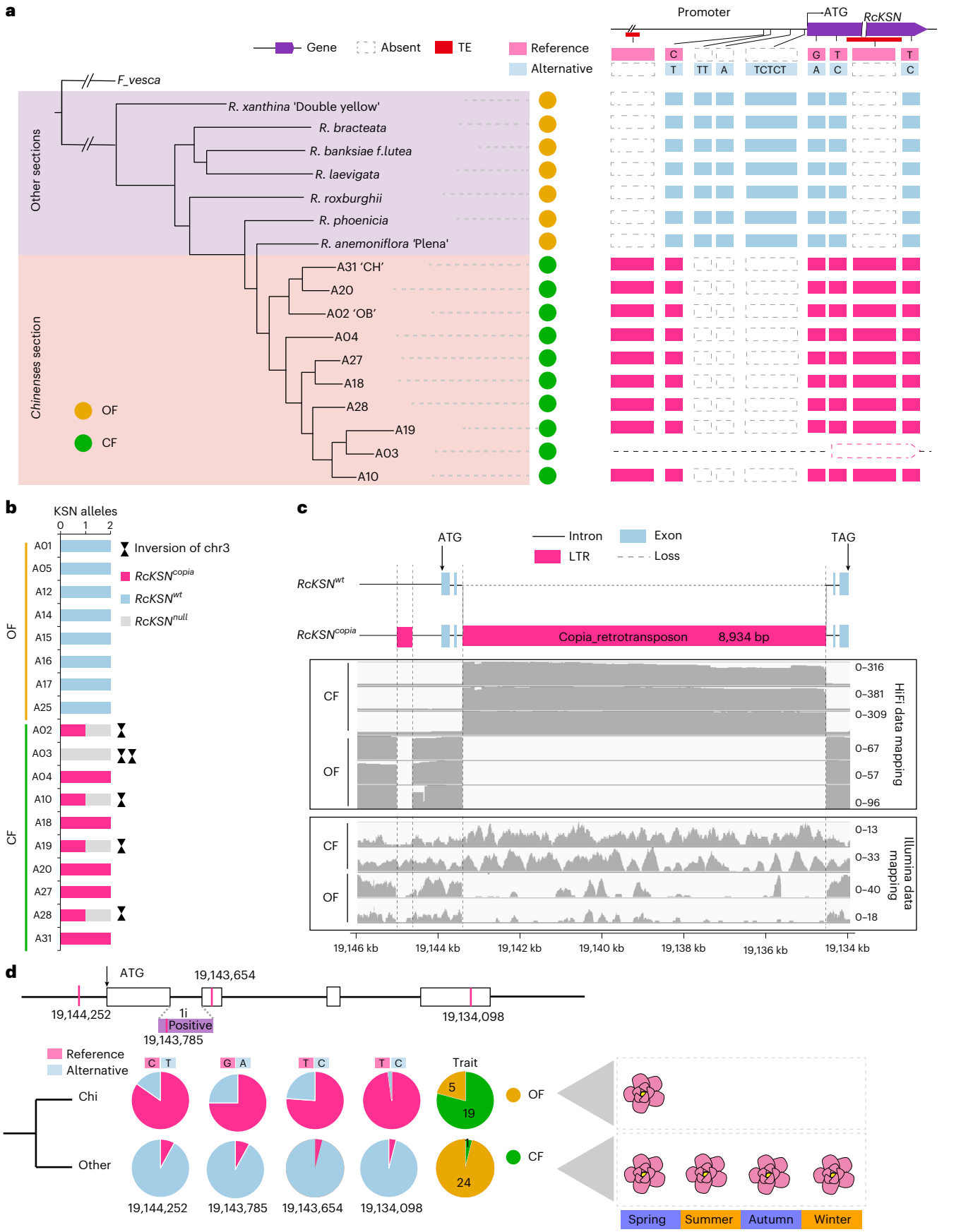


Fig. 4 | Allelic variants in genes associated with CF in *Rosa* accessions. **a**, Phylogenetic relationships of 17 *Rosa* accessions from the *Chinenses* section and other sections constructed by single-copy genes, and the allelic variants in genes associated with CF. **b**, The *RcKSN*^{null} genotype was associated with the inversion of Chr3 in *Rosa* accessions. **c**, The coverage in the two TE insertion regions of the *RcKSN* genotype was illustrated when aligning HiFi and Illumina reads from both OF and CF accessions to the A02 reference genomes. **d**, The number of accessions carrying *KSN* alleles haplotypes. ATG, start codon; LTR, long terminal repeat; TAG, stop codon.

https://doi.org/10.1038/s41588-026-02569-z

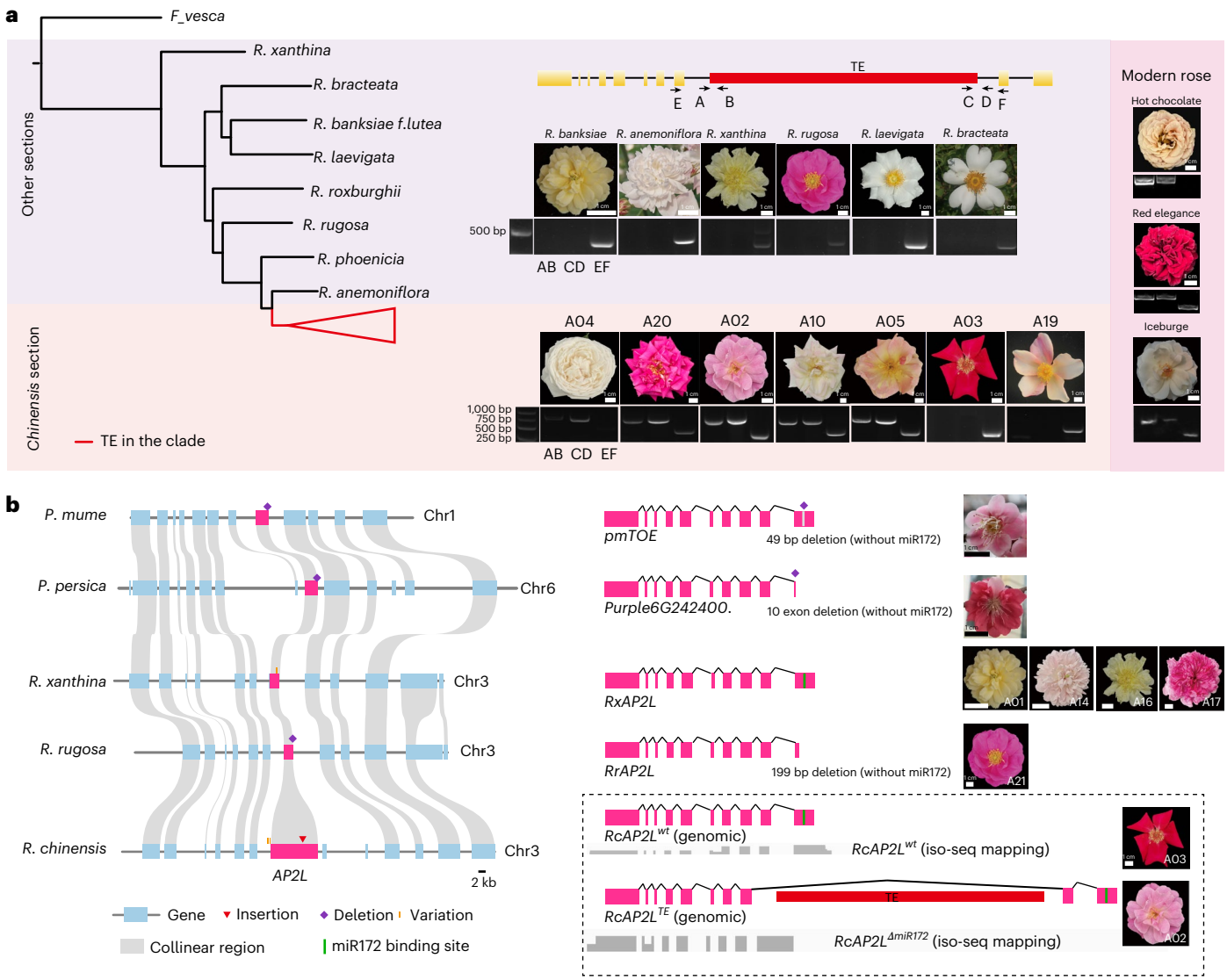


Fig. 5 | Allelic variants in *AP2L* homolog genes associated with double-petal flowers in *Rosa*. **a**, a 10-kb transposon insertion in the eighth intron of the *AP2L* allele was found exclusively in *Chinenses* section of double-petal roses, not in other section roses, which was verified by PCR amplification. The genotypes of the modern double-petal roses are similar to those of *Chinenses* section double-petal roses. Primers used are indicated by arrows following the previous study[26]. **b**, The *AP2L* homologs known to be involved in the formation of double-petal flowers in the Rosaceae family, located within one collinear block, exhibit different variations, yet all can affect miR172 target binding.

RcAP2L gene also has ten exons that include the miR172 target site in both single-petal and double-petal flowers. However, with the specific TE insertion, the *RcAP2L*^{TE} allele can produce a splicing isoform without miR172 binding sites to regulate petal number[26] (Fig. 5b). In addition to this TE insertion, we identified three SNPs and the deletion of 'TAATA' in the promoter of *RcAP2L*^{TE} as associated with the DF trait, which is worthy of further functional characterization (Supplementary Fig. 26b). These in-depth analyses based on our pangenome enhance our understanding of the diversity of variations that occur in the *AP2L* locus and lead to DF. Furthermore, TE insertions in the gene regions of *AP2L* and *KSN* were further confirmed by *Rosa* pan graph we constructed (Supplementary Fig. 27).

Petal discoloration. Petal discoloration is an attractive ornamental trait, and an increasing number of cultivars exhibiting this characteristic are currently being cultivated. The *R. chinensis* Butterfly (A19), which is an early cultivar that is commonly referred to as 'Mutabilis', is known for its distinctive petal discoloration from orange to red (Fig. 6a). We investigated the metabolomic basis of petal discoloration in Mutabilis and determined that cyanidin-3-*O*-glucoside and cyanidin-3,5-*O*-glucoside were the main anthocyanin, and that β-carotene was the main carotenoid (Supplementary Tables 12 and 13). The accumulation of these two anthocyanins gradually increased throughout petal development, whereas β-carotene levels decreased during this process (Fig. 6b and Supplementary Figs. 28 and 29). We further found that *flavonoid 3-O-glucosyltransferase_1* (*3GT_1*), which promotes the stability of anthocyanins[29,30], was highly expressed after flower opening, and the expression of the *carotenoid isomerase* (*CRTISO*) and *lycopene β-cyclase_2* (*LCYB_2*) genes, which are related to β-carotene synthesis, was downregulated (Fig. 6c,d). Notably, *carotenoid cleavage dioxygenase 4* (*CCD4*) was expressed in the early flower bud stage in red petal A31 and pink petal A02. Its expression in A19 was restricted to the blooming flower stage (Fig. 6e–g). To confirm *CCD4* activity, we used engineered *Escherichia coli* that produce β-carotene. After inducing pET28a-CCD4 expression, the cells changed color from orange to white, and β-carotene levels reduced significantly (Fig. 6h,i and Supplementary Fig. 30). These results indicate that *CCD4* has the ability to degrade β-carotene. The promoter elements of the *CCD4*

https://doi.org/10.1038/s41588-026-02569-z

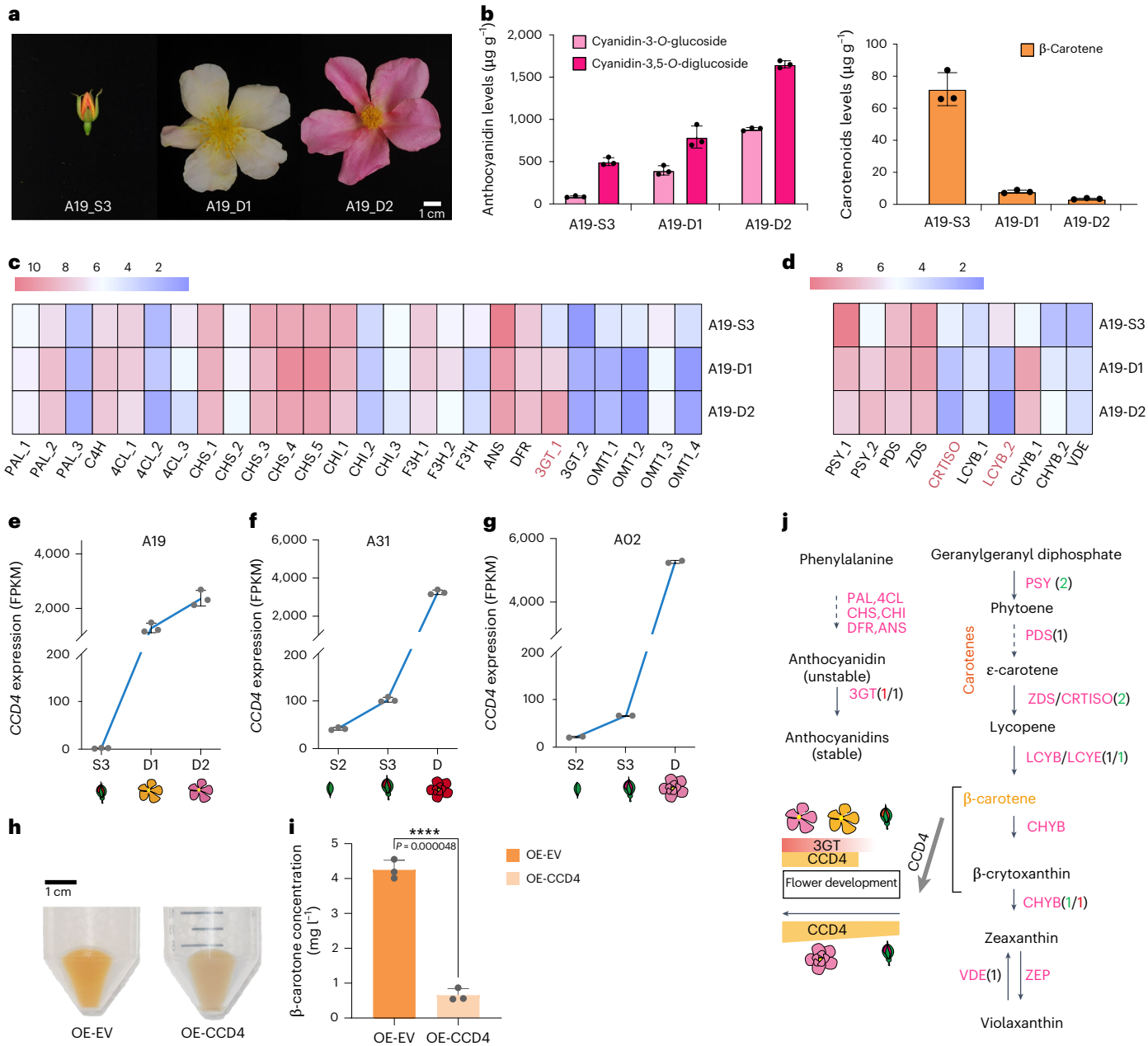


Fig. 6 | The potential regulatory mechanisms underlying petal discoloration in A19. a, Flowers of *R. chinensis* Butterfly (A19) at different floral development stages, including bud about to open (S3), first day of anthesis (D1) and second day of anthesis (D2). **b**, The cyanidin-3-*O*-glucoside and cyanidin-3,5-*O*-glucoside were identified as the main anthocyanins, β-carotene was identified as the primary carotenoid in A19, data are presented as mean ± s.d. ($n = 3$ biological replicates). **c**, Heatmap depicting the expression of genes involved in the anthocyanin biosynthesis pathway. **d**, Heatmap showing the expression of genes associated with the carotenoid synthesis pathway. **e**–**g**, *CCD4* gene expression profiles in A19 (**e**), *R. chinensis* CH (A31) (**f**) and *R. chinensis* OB (A02) (**g**) at different stages of flower development, data are presented as mean ± s.d. ($n = 3$, 3 and 2 biological replicates). **h**,**i**, Discoloration (**h**) and relative content (**i**) changes of *E. coli* cells accumulating β-carotene when expressing *CCD4*. Data are presented as mean ± s.d. ($n = 3$ biological replicates, independent *E. coli* single colonies). ****$P < 0.0001$. **j**, Potential regulatory pathways of petal discoloration. The delayed expression of *CCD4* slowed carotenoid degradation, resulting in discoloration of petals. Red numbers indicate upregulated genes, whereas green numbers indicate downregulated genes. OE-EV, cotransformation with pET28a and pACCARΔ16crtX; OE-CCD4, cotransformation with pET28a-CCD4 and pACCARΔ16crtX.

gene were similar in red/pink petals of accessions A31 and A02, but showed multiple differences when compared with those of A19, which displays petal discoloration. Dual-luciferase assays showed that the A02 promoter exhibited higher transcriptional activity than A19 in both the 260/268 bp and 941/927 bp regions. A19 contained additional 'MYB binding site involved in light responsiveness' and 'light-responsive' elements (Supplementary Figs. 31 and 32), which may delay *CCD4* expression by affecting transcription or introducing light-dependent regulation, contributing to petal discoloration. Based on these results, we propose that the activity of multiple genes that are involved in the regulation of the anthocyanin and carotenoid pathways collectively leads to the delayed carotenoid degradation and the conversion of anthocyanins into stable molecules until flowering, which may be crucial for the discoloration of petals (Fig. 6j).

Discussion

Over the past 150 years, growing demand for new rose varieties has led to a shift from traditional hybridization to marker-assisted and

genomics-based breeding, improving breeding efficiency and enabling the development of modern rose cultivars[31,32]. Nevertheless, rose breeding remains hampered by complex ploidy levels, high heterozygosity, dependence on vegetative propagation, low transformation efficiency[33] and the current reliance on single reference genomes for the discovery of trait-associated variants.

To address these limitations, pangenome approaches combined with de novo domestication and redomestication provide powerful tools to harness underused genetic resources from wild or ancestral germplasms[34,35]. Notably, recent population genomics studies have shown that *Rosa* accessions within the *Chinenses* section harbor the highest level of variation[32,36]. In contrast, modern rose cultivars exhibit reduced genetic diversity, suggesting that reintroduction of Chinese rose germplasm may reinvigorate breeding lines. Our constructed pangenome includes haplotype-resolved assemblies from 26 *Rosa* accessions across nine major taxonomic sections in Asia and Europe, many of which are derived from diverse Chinese roses. These include historically important cultivars exhibiting unique and desirable traits that have been largely lost in modern varieties, such as petal discoloration in *R. chinensis* Butterfly and *R. odorata* 'Fortune's Double Yellow', miniaturized petals in *R. chinensis* Minima and sepal overgrowth at the pistil in *R. chinensis* Sai Zhao Jun. This high-resolution genomic resource allows for precise identification of trait-regulating genes and offers critical guidance for reintroducing lost ornamental traits into breeding programs.

Our pangenomic analyses advance understanding of key genomic loci associated with important phenotypic traits. For example, we revealed extensive allelic diversity at the *AP2L* locus, implicating it in petal doubling. We also uncovered that expression variation of *CCD4* and *3GT* genes is central to petal discoloration. For CF, four SNPs within *RcKSN* were significantly associated with the trait, three of which were specifically identified in this study. One of these SNPs, located in the first intron, was also previously reported[25] to correlate with CF. Moreover, prior research has demonstrated that this intron modulates *KSN* expression and flowering inhibition in *R. rugosa*[37], suggesting a functional role for this variant. Further molecular validation is necessary to confirm these associations and elucidate underlying regulatory mechanisms.

Our analysis of the allotetraploid *R. gallica* revealed two distinct subgenomes. Subgenome A appears closely related to *R. phoenicia* (section Synstylae), supporting earlier findings that *R. gallica* has partial ancestry from the European *Synstylae* section[38]. However, subgenome B shows the highest similarity to *R. rugosa*, which contradicts previous hypotheses suggesting either an extinct lineage near the *Synstylae–Caninae* clade or a trihybrid origin involving a *Caninae* × *Rosa* hybrid[38]. These discrepancies likely reflect limitations in taxonomic sampling. Our dataset includes relatively few wild European rose species, while the other work lacked sufficient representation of *R. rugosa*. Broader sampling of *Rosa* germplasm worldwide will be necessary to definitively trace the evolutionary origins of *R. gallica*.

Inversion can significantly impact gene expression, recombination and linkage disequilibrium, which are essential for plant environmental adaptation and domestication[15,39]. In this study, we constructed a map of large-scale inversions across *Rosa* accessions, providing substantial insights into interspecific genetic variation and applications in molecular breeding. Notably, the *RcKSN* gene was found to be lost at an inversion breakpoint. This finding emphasizes the need to consider SVs, such as inversions, when selecting for the CF trait to enhance breeding precision and efficiency.

In summary, the high-quality pangenome constructed in this study represents a foundational resource for rose genetics and breeding. It provides comprehensive insights into the genomic diversity and evolutionary dynamics within the *Rosa* subgenus and enables detailed investigation of trait inheritance across over 35,000 cultivated varieties. The identification of candidate genes and structural variations associated with ornamental traits opens new avenues for both fundamental research and applied breeding. Additionally, this genomic framework supports the exploration of genetic mechanisms underlying other important agronomic traits such as disease resistance, water-use efficiency and vase life, facilitating a potential second revolution in rose improvement.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-026-02569-z.

References

1. Krüssmann, G. *The Complete Book of Roses* (translated by Krüssmann, G. & Raban, N.) (Timber Press, 1982).
2. Raymond, O. et al. The *Rosa* genome provides new insights into the domestication of modern roses. *Nat. Genet.* **50**, 772–777 (2018).
3. Zhang, X. et al. Haplotype-resolved genome assembly of the diploid *Rosa chinensis* provides insight into the mechanisms underlying key ornamental traits. *Mol. Hortic.* **4**, 14 (2024).
4. Zang, F. Q. et al. A high-quality chromosome-level genome of wild *Rosa rugosa*. *DNA Res.* **28**, dsab017 (2021).
5. Chen, F. et al. A chromosome-level genome assembly of rugged rose (*Rosa rugosa*) provides insights into its evolution, ecology, and floral characteristics. *Hortic. Res.* **8**, 141 (2021).
6. Zhong, M. C. et al. Rose without prickle: genomic insights linked to moisture adaptation. *Natl Sci. Rev.* **8**, nwab092 (2021).
7. Nakamura, N. et al. Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. *DNA Res.* **25**, 113–121 (2018).
8. Qin, P. et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**, 3542–3558 (2021).
9. Chen, S. et al. Gene mining and genomics-assisted breeding empowered by the pangenome of tea plant *Camellia sinensis*. *Nat. Plants* **9**, 1986–1999 (2023).
10. Yan, H. D. et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nat. Genet.* **55**, 507–518 (2023).
11. He, Q. et al. A graph-based genome and pan-genome variation of the model plant *Setaria*. *Nat. Genet.* **55**, 1232–1242 (2023).
12. Liu, Z. et al. Grapevine pangenome facilitates trait genetics and genomic breeding. *Nat. Genet.* **56**, 2804–2814 (2024).
13. Chiang, C. et al. The impact of structural variation on human gene expression. *Nat. Genet.* **49**, 692–699 (2017).
14. Zhou, Y. et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **606**, 527–534 (2022).
15. Tang, D. et al. Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606**, 535–541 (2022).
16. Wissemann, V. Conventional taxonomy (wild roses). *Ref. Mod. Life Sci.* https://doi.org/10.1016/B978-0-12-809633-8.05017-2 (2017).
17. Figueiredo, A. C., Barroso, J. G., Pedro, L. G. & Scheffer, J. J. C. Factors affecting secondary metabolite production in plants: volatile components and essential oils. *Flavour Fragr. J.* **23**, 213–226 (2008).
18. Qiu, X. et al. Phylogenetic relationships of wild roses in China based on nrDNA and matK data. *Sci. Hortic.* **140**, 45–51 (2012).
19. Noh, Y. M., Ait Hida, A., Raymond, O., Comte, G. & Bendahmane, M. The scent of roses, a bouquet of fragrance diversity. *J. Exp. Bot.* **75**, 1252–1264 (2024).
20. Wellenreuther, M. & Bernatchez, L. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* **33**, 427–440 (2018).

21. Miyazawa, Y. et al. Overexpression of *MIZU-KUSSEI1* enhances the root hydrotropic response by retaining cell viability under hydrostimulated conditions in *Arabidopsis thaliana*. *Plant Cell Physiol.* **53**, 1926–1933 (2012).
22. Iwata, H. et al. The *TFL1* homologue *KSN* is a regulator of continuous flowering in rose and strawberry. *Plant J.* **69**, 116–125 (2012).
23. Rech, G. E. et al. Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat. Commun.* **13**, 1948 (2022).
24. Akakpo, R., Carpentier, M. C., Ie Hsing, Y. & Panaud, O. The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytol.* **226**, 44–49 (2020).
25. Soufflet-Freslon, V. et al. Diversity and selection of the continuous-flowering gene, *RoKSN*, in rose. *Hortic. Res.* **8**, 76 (2021).
26. Francois, L. et al. A miR172 target-deficient *AP2-like* gene correlates with the double flower phenotype in roses. *Sci. Rep.* **8**, 12912 (2018).
27. Gattolin, S., Calastri, E., Tassone, M. R. & Cirilli, M. Mutations overlying the miR172 target site of TOE-type genes are prime candidate variants for the double-flower trait in mei. *Sci. Rep.* **14**, 7300 (2024).
28. Gattolin, S. et al. Deletion of the miR172 target site in a TOE-type gene is a strong candidate variant for dominant double-flower trait in Rosaceae. *Plant J.* **96**, 358–371 (2018).
29. Ogata, J., Kanno, Y., Itoh, Y., Tsugawa, H. & Suzuki, M. Plant biochemistry: anthocyanin biosynthesis in roses. *Nature* **435**, 757–758 (2005).
30. Yan, Y. et al. Light-mediated anthocyanin biosynthesis in rose petals involves a balanced regulatory module comprising transcription factors *RhHY5*, *RhMYB114a*, and *RhMYB3b*. *J. Exp. Bot.* **74**, 5783–5804 (2023).
31. Bendahmane, M., Dubois, A., Raymond, O. & Le Bris, M. Genetics and genomics of flower initiation and development in roses. *J. Exp. Bot.* **64**, 847–857 (2013).
32. Zhou, L. et al. Multi-omics analyzes of *Rosa gigantea* illuminate tea scent biosynthesis and release mechanisms. *Nat. Commun.* **15**, 8469 (2024).
33. Anderson, N. O. *Flower Breeding and Genetics: Issues, Challenges and Opportunities for the 21st Century* (Springer-Verlag, 2007).
34. Fernie, A. R. & Yan, J. De novo domestication: an alternative route toward new crops for the future. *Mol. Plant* **12**, 615–631 (2019).
35. Schreiber, M., Jayakodi, M., Stein, N. & Mascher, M. Plant pangenomes for crop improvement, biodiversity and evolution. *Nat. Rev. Genet.* **25**, 563–577 (2024).
36. Zhang, Z. et al. Haplotype-resolved genome assembly and resequencing provide insights into the origin and breeding of modern rose. *Nat. Plants* **10**, 1659–1671 (2024).
37. Bai, M. et al. The intragenic *cis*-elements mediate temperature response of *RrKSN*. *Plant Physiol. Biochem.* **202**, 107983 (2023).
38. Debray, K. et al. Unveiling the patterns of reticulated evolutionary processes with phylogenomics: hybridization and polyploidy in the genus *Rosa*. *Syst. Biol.* **71**, 547–569 (2022).
39. Hu, H. et al. Unravelling inversions: technological advances, challenges, and potential impact on crop breeding. *Plant Biotechnol. J.* **22**, 544–554 (2024).
40. Guan, H. et al. Identification of distinct roses suitable for future breeding by phenotypic and genotypic evaluations of 192 rose germplasms. *Hortic. Adv.* **2**, 5 (2024).

[1]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, State Key Laboratory of Tropical Crop Breeding, State Key Laboratory of Genome and Multi-omics Technologies, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [2]National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, China. [3]Kunpeng Institute of Modern Agriculture at Foshan, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China. [4]College of Landscape Architecture and Forestry, Shandong Key Laboratory for Germplasm Innovation of Saline-alkaline Tolerant Grasses and Trees, Qingdao Agricultural University, Qingdao, China. [5]College of Horticulture, Shanxi Agricultural University, Jinzhong, China. [6]The Sino-French international joint Laboratory for Horticultural Research, Huazhong Agricultural University, INRAE, Université Paris-Saclay, ENS de Lyon, Lyon, France. [7]Laboratoire de Reproduction et Développement des Plantes, INRAE-CNRS-UCBL-ENS de Lyon, Ecole Normale Supérieure, Lyon, France. [8]These authors contributed equally: Xiaoni Zhang, Lan Lan, Yingxue Yang, Huilin Guan. ✉e-mail: mohammed.bendahmane@ens-lyon.fr; fuxiaopeng@mail.hzau.edu.cn; wuzhiqiang@caas.cn; panweihua@caas.cn

Methods

Samples collection and genome sequencing

All *Rosa* accessions were cultivated at the Flower Garden of Huazhong Agricultural University. For de novo assembly, leaves from 23 accessions were collected for whole-genome sequencing. Genomic DNA was extracted and used to construct PacBio SMRTbell libraries with a target insert size of ~15 kb, following the manufacturer's standard protocols. Whole-genome sequencing was performed on the PacBio Sequel II platform using circular consensus sequencing to generate HiFi data with default parameters. For Hi–C sequencing, leaf cells were cross-linked and lysed, and fixed chromatin was digested using the DpnII endonuclease. Subsequently, 150-bp-insert-size libraries were constructed and sequenced on the MGI platform to generate Hi–C reads for each accession. Additionally, transcriptome sequencing was conducted on the MGI platform using extracted RNA from a mixture of tissues, including young stems, flower buds, leaves and young roots to support genome annotation (Supplementary Table 14). For ultralong reads, high-molecular-mass DNA was extracted using a super-long DNA extraction kit developed by Sailgene, yielding reads with an N50 greater than 90 kb. These ultralong reads were generated using the Oxford Nanopore Technologies high-throughput sequencing platform (Supplementary Table 14).

Genome assembly and annotation

Each *Rosa* accession generated HiFi sequencing data of ~32 Gb, and Hi–C data of ~82 Gb. Two hybrids (the tetraploid Carola and the triploid La France) were generated at higher depth, each yielding ~64 Gb of HiFi data (Supplementary Table 14). These genomes were assembled using hifiasm (v0.1.6.3). The resulting contigs that were obtained were aligned to rose plastid genomes using BLASTN (v2.9.0), and contigs with over 95% identities to the plastid genome were filtered out. Chromosome-level scaffolding was performed using Juicer[41] (v1.6) and 3D-DNA[42]. For polyploid accessions, candidate collapsed regions were identified based on sequencing depth and copied to other haplomes following the ALLHiC (v0.9.13) pipeline[43]. Manual curation was performed to enhance the completeness and accuracy of chromosome-level assemblies, aided by tools such as GPhase (v0.1.0-beta; https://github.com/panlab-bioinfo/GPhase), Gap-Aid (v1.1; https://github.com/panlab-bioinfo/Gap-Aid) and Gap-Graph (v1.0.1; https://github.com/panlab-bioinfo/Gap-graph). TEs in each accession were identified using the following Extensive De-Novo TE Annotator (EDTA)[44] with the parameters: --anno 1 --force 1 --debug 1 -sensitive 1. Protein-coding gene annotation was performed using EVidenceModeler[45], which integrates predictions from three approaches—ab initio gene prediction, transcriptome-based evidence and homology-based annotation. Ab initio predictions were generated using AUGUSTUS (v3.3.3)[46], SNAP (https://github.com/KorfLab/SNAP) and GlimmerHMM (v3.0.4)[47]. For homology-based prediction, protein sequences from eight species were used, *A. thaliana* (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001735.4), *Oryza sativa* (GCF_001433935.1_IRGSP-1.0), *Vitis vinifera* (GCF_000003745.3) and five Rosaceae species *R. chinensis* OB, *R. chinensis* OB _DH, *P. persica*. v2, *Pyrus betulifolia* and *F. vesca*_v4 (https://www.rosaceae.org/). These protein sequences were incorporated with Exonerate (v2.2.0) and AUGUSTUS. For transcriptome-based prediction, RNA-seq data were mapped with HiSAT2 (v2.2.1)[48] and assembled with PASA (v2.5.1)[49].

Evaluation of the genome assembly and annotation results

The quality of the genome assemblies was evaluated using multiple complementary approaches. First, BUSCO, based on the embryophyta_odb10 database, was used with default parameters to evaluate both genome assembly completeness and gene annotation quality. Second, *k*-mer analysis toolkit[50] was used to compare the *k*-mer spectra between the HiFi reads and the assemblies. Third, long terminal repeat-assembly index scores were calculated to assess the genome assembly continuity based on repetitive sequences annotation generated by EDTA. Fourth, Hi–C contact heatmaps were visualized to evaluate the structural correctness and phasing accuracy of the assemblies. Fifth, merqury[51] was used to calculate the consensus quality value of each haplotype, providing a quantitative measure of base-level accuracy. Furthermore, the switch error rates were estimated by the calc_switcherr (https://github.com/tangerzhang/calc_switchErr) and polyswitch (https://gitee.com/tanger-lab/polyswitch) pipelines, further validating haplotype phasing accuracy. And Bionano cmaps were aligned to phased assembly haplotypes of OB to identify haplotype 1 and haplotype 2 cmaps using Bionano RefAligner.

Phylogenetic relationship construction and introgression events identification

OrthoFinder (v2.5.1)[52] was used with default parameters to cluster genes and identify single-copy orthologous gene groups. These groups were then aligned with MAFFT[53] with default settings. Species divergence times were estimated using MCMCTree from the PAML package[54], with time calibrations based on *A. thaliana* and *F. vesca* (102–112.5 MYA), *M. domestica* and *P. persica* (34.4–67.2 MYA), *M. domestica* and *F. vesca* (49.2–77.1 MYA), obtained from the TimeTree database[55]. Gene family contraction-expansion analysis was performed with CAFE5 software[56]. HiFi reads from *Rosa* accessions were mapped to the *F. vesca* reference genome with minimap2 (ref. 57), and duplicates were marked with Samblaster[58]. Variant calling was performed using DeepVariant[59], followed by gVCF merging and joint genotyping using GLnexus[60]. Introgression signals were evaluated using the ABBA-BABA (D-statistic) test, implemented in Dsuite[61] based on SNPs that passed quality control as described in a previous report[14].

Subgenome classification

To further investigate the subgenomic features in the tetraploid genome of *R. gallica* (A13), a phylogenetic tree containing multiple accessions (*R. rugosa*, *R. Phoenicia, R. xanthina* and *R. gallica*) and *F. vesca* was constructed using single-copy genes. The ploidy and genome structure of A13 and A11 were estimated and visualized with Smudgeplot[62] based on heterozygous *k*-mer pairs. All the haplotypes from accessions A02, A18, A14, A25, A21, A17, A12, A15, A01 and A16 were merged to generate the 'reference contributor genome'. The haplotypes of A13_A1 and A13_B2 were split into 2-kb fragments and mapped to the 'reference contributor genome' by BLASTN (e value = 1×10^{-5}). Alignments with coverage below 70% were discarded, and the best hits were selected as the candidate donors. A sliding window of 100 kb (without step) was then applied across the A13_A1 and A13_B2 genomes to identify the candidate origin contributor. Additionally, amino acid sequences from each homologous chromosome were extracted and used as input for OrthoFinder. Sequences were aligned and trimmed as previously described, and phylogenetic trees were constructed to compare chromosomes across accessions. Maximum likelihood phylogenetic tree was inferred from the super-matrix alignments using IQ-TREE[63]. Sourmash was used to cluster the homologous chromosomes based on the *k*-mer with the default parameters[64].

The demographic history of *Rosa*'s accessions

To gain a deeper understanding of the demographic history of diploid accessions, the effective population size over time was inferred using the pairwise sequentially Markovian coalescent[65]. Pairwise sequentially Markovian coalescent analysis was performed on each individual and 100 bootstrap replicates were generated to assess estimation robustness. The analysis was conducted using an estimated mutation rate of 4.02×10^{-9} substitutions per site per generation and an average generation time of 1 year.

Homolog expression bias analysis

Total RNA was extracted from A13 and La France (A11) samples, including petals, flower buds, leaves and stems, and sequenced on the

MGI platform, with three replicates for each sample. The reads were aligned using HiSAT2 (v2.2.1; ref. 48) and quantified with StringTie (v2.1.6; ref. 66) to obtain FPKM values. Homologous gene clustering revealed a 1:1:1:1 relationship among the four haplotypes (1:1:1 in triploids), which were defined as tetrads. Only tetrads with a cumulative expression value of >0.5 FPKM across all homologs were retained for subsequent analyses. To enable subgenome-level expression comparison, the expression levels of each subgenome in the tetraploid A13 were normalized as previously described[67] using the following formulas:

$$\text{Expression}_{\text{A13A1}} = \frac{\text{FPKM (A13A1)}}{\text{FPKM (A13A1)} + \text{FPKM (A13A2)} + \text{FPKM (A13B1)} + \text{FPKM (A13B2)}}$$

$$\text{Expression}_{\text{A13A2}} = \frac{\text{FPKM (A13A2)}}{\text{FPKM (A13A1)} + \text{FPKM (A13A2)} + \text{FPKM (A13B1)} + \text{FPKM (A13B2)}}$$

$$\text{Expression}_{\text{A13B1}} = \frac{\text{FPKM (A13B1)}}{\text{FPKM (A13A1)} + \text{FPKM (A13A2)} + \text{FPKM (A13B1)} + \text{FPKM (A13B2)}}$$

$$\text{Expression}_{\text{A13B2}} = \frac{\text{FPKM (A13B2)}}{\text{FPKM (A13A1)} + \text{FPKM (A13A2)} + \text{FPKM (A13B1)} + \text{FPKM (A13B2)}}$$

Based on normalized expression profiles, tetrads in A13 were classified into the following distinct expression patterns: 'X_dominant'—one haplotype exhibits expression of >0.4, and the remaining three are each of <0.2; 'X_X_dominant'—two haplotypes each show expression of >0.3, and the others of <0.2; 'X_suppressed'—one haplotype shows expression of <0.1, while the other three are each of >0.3; 'Balanced'—all other expression patterns not fitting the above criteria. For triploid A11, expression classification followed a strategy detailed in the Supplementary Note. The results were further visualized with the R package.

The rose pangenome

The coding sequences from the 26 accessions were clustered using OrthoFinder with default parameters. The resulting gene groups were classified into core, softcore, dispensable and private categories. Core clusters contain genes present in all individuals, while softcore clusters include genes found in at least 24 individuals. Private clusters contain genes unique to a single individual, and all other clusters are classified as dispensable. Genes from the core, softcore and dispensable clusters were further analyzed to calculate d_N/d_S value with PAML under the M0 model in the codeML program. The expression levels of genes from different clusters were identified using RNA-seq data from various tissues, including young stems, flower buds, leaves and young roots. The graph pangenome was built using PanGenome Graph Builder (PGGB)[68] following the default instructions.

Structure variation detection

All the haplotypes were aligned to the A02_2 genome with minimap2 (-asm 5)[57]. PAVs, translocations and inversions were detected by SyRi[69]. Given the large number of SVs, we adopted a pipeline similar to that used in rice[8]. Based on SyRi's output, the CPL, CPG, DEL, INS, DUP/INVDP, HDR, NOTAL and TDM were converted as PAVs in each accession. INV variants were considered inversion SVs, while TRANS and INVTR variants were categorized as translocation SVs. Only variations longer than 50 bp were considered. SVs were considered identical if their overlapping ratio exceeded 90%, with the overlap covering at least 90% of the total SV length. To examine the correlation between SVs and TEs, SVs longer than 50 bp were classified as TE-amplified if TEs accounted for 80% or more of the total SV length. Variation hotspot regions were defined using a window size of 400 kb and a sliding size of 200 kb. The top 10% of all windows with the highest frequency of SV breakpoints were identified as hotspot regions.

Verification of genetic variation in the *RcKSN* gene

To investigate the genotypes of the *KSN* gene across different accessions, we identified the homologous *RcKSN* genes in all accessions by using the amino acid sequence of *RcKSN* (RcHm_v2.0_Chr3g0473021) as a query in BLASTP. Whole-genome resequencing data from 10 accessions (SRP119986) and 50 accessions (PRJNA704782) were mapped to the A02_2 assembly using BWA (0.7.17-r1188) with default parameters. SNP calling was performed using GATK (Supplementary Table 15). Additionally, HiFi reads from these accessions were mapped to the A02_2 genome using minimap2, and the resulting BAM files were visualized using Integrative Genomics Viewer.

RcAP2L sequence and alternative splicing analysis

Flower buds at stage 0, as previously described[3], were collected from *R. chinensis* A02 (double-petal flower), A03 (single-petal flower), A04 (double-petal flower) and A28 (double-petal flower). Total RNA from these four accessions was sequenced on the MGI platform. Additionally, RNA from A02 and A03 was sequenced using the PacBio Sequel II platform at Kindstar Sequenon Biotechnology (Wuhan). Pair-end RNA-seq reads obtained from the four accessions were mapped to the respective genomes with HiSAT2 and assembled with PASA to identify alternative splice variants of *AP2L*. Iso-seq data were mapped to the reference genome with minimap2 to validate the alternative splicing isoform of *AP2L*.

Genotyping PCR of *RcKSN* and *RcAP2L*

Young rose leaves were collected, and genomic DNA was extracted using the CTAB method. Primers targeting the 3′ and 5′ ends of the transposons in the *KSN* genes were designed (Supplementary Table 16), whereas primers for the *AP2L* transposon were adopted from a previous study[26]. PCR amplification was carried out with an initial denaturation step at 95 °C for 3 min, followed by 32 cycles of 95 °C for 15 s, 56 °C for 20 s and 72 °C for 50 s. A final elongation step was performed at 72 °C for 10 min.

Transcriptomic analysis of different samples

The raw RNA-seq data from all accessions were processed for quality control using Trimmomatic (0.36)[70] to remove low-quality and adaptor-containing reads. The resulting clean reads were then aligned to the reference genome with HiSAT2. Expression levels were normalized using the FPKM method, implemented by StringTie (v2.1.6)[66].

Pigment-related gene expression analysis

Total RNA was extracted from A19 flowers at five different developmental stages, namely S3 (bud about to open), D1 (first day of anthesis with orange petals), D2 (second day of anthesis with pink petals), D3 (third day of anthesis with pink petals) and D4 (fourth day of anthesis with pink petals), with three biological replicates for each stage, and was sequenced (Genome Sequence Archive, CRA006521)[71]. Additionally, transcriptome data of roses at various developmental stages were obtained for comparative expression analysis, including OB with pink petals (PRJNA351281)[72] and CH with red petals, where each stage of both OB and CH included two and three biological replicates. RNA-seq reads were aligned to their respective genomes using HiSAT2, and transcript abundance was quantified as FPKM values using StringTie (v2.1.6).

Pigment extraction and liquid chromatography–mass spectrometry analysis

Flowers of A19 at five floral developmental stages (S3, D1, D2, D3 and D4) were selected for petal discoloration analysis. Petals samples from three stages (S3, D1 and D2) were collected for the detection of anthocyanins and carotenoids. The petals were ground in liquid nitrogen, weighed and subjected to extraction, vertexing, ultrasound treatment and centrifugation to obtain the supernatant for liquid chromatography–mass spectrometry (LC–MS) analysis of pigment

types (Supplementary Note). Metabolite profiling of anthocyanins and carotenoids was performed and quantified using scheduled multiple reaction monitoring with Analyst (v1.6.3) software (Sciex) and Multiquant (v3.0.3) software (Sciex). Detailed protocols are available in the Supplementary Note.

Gene cloning and vector construction

The coding sequence of *CCD4* was amplified from cDNA derived from A19 D1 petals using gene-specific primers (Supplementary Table 17). The resulting PCR products were cloned into the linearized pET28a vector to generate the recombinant construct pET28a-*CCD4*. Detailed methods for cloning and vector construction of various *CCD4* promoter fragment lengths from A19 and A02 are provided in the Supplementary Note.

Heterologous expression of *CCD4* in *E. coli*

The pACCARΔ16crtX plasmid, along with either the pET28a-*CCD4* construct or empty pET28a vector, was cotransformed into *E. coli* BL21 (DE3) cells. Transformants were cultured in lysogeny broth medium at 37 °C in the dark until the optical density at 600 nm (OD_{600}) reached 0.4–0.55. Protein expression was then induced by adding 0.5 mM isopropyl-β-D-1-thiogalactopyranoside, followed by incubation at 37 °C with shaking at 140 rpm for 4 h.

β-Carotene quantification was performed by measuring the absorbance of acetone-extracted β-carotene at 453 nm, as previously described[73]. Cells were collected through centrifugation at 1,340*g* for 10 min, resuspended in 5 ml of acetone and incubated at 55 °C for 15 min in the dark. After centrifugation at 4,100*g* for 10 min, the acetone supernatant containing β-carotene was transferred to a new tube (Centrifuge 5804/5804R, Eppendorf). A standard curve was generated by measuring the OD453 of serially diluted β-carotene standards (Lot N08HB200346, Shanghai YuanYe Biotechnology) using a Mapada UV-1800PC spectrophotometer (Mapada). This standard curve (Supplementary Fig. 29) was then used to calculate β-carotene concentration in each strain. All the promoter construction for CCD4 and the DUAL-LUC assay are provided in the Supplementary Note.

Statistical analysis

P value for significant differences was calculated using the 'stat_compare_means' from the gghubr, leveraging the two-tailed *t* test.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The 51 haplotype-resolved genome assemblies and the associated raw sequencing data for the 23 *Rosa* accessions have been deposited in the NCBI BioProject under accessions PRJNA1194317, PRJNA1400323, PRJNA1400713, PRJNA1400325, PRJNA1400715, PRJNA1400330, PRJNA1400716, PRJNA1400345, PRJNA1400717, PRJNA1400382, PRJNA1400718, PRJNA1400411, PRJNA1400719, PRJNA1400417, PRJNA1400720, PRJNA1400440, PRJNA1400721, PRJNA1400442, PRJNA1400722, PRJNA1400443, PRJNA1400723, PRJNA1400444, PRJNA1400724, PRJNA1400306, PRJNA1400725, PRJNA1400449, PRJNA1400726, PRJNA1400451, PRJNA1400727, PRJNA1400477, PRJNA1400728, PRJNA1400478, PRJNA1400730, PRJNA1400479, PRJNA1400731, PRJNA1400481, PRJNA1400732, PRJNA1400490, PRJNA1400733, PRJNA1400492, PRJNA1400734, PRJNA1400494, PRJNA1400735, PRJNA1400736, PRJNA1400495, PRJNA1400737, PRJNA1400738, PRJNA1400739, PRJNA1400496, PRJNA1400740, PRJNA1400742and PRJNA1400743, and in the National Genomics Data Center under accession PRJCA039059 (https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA039059). The genome annotation data have been deposited in the Figshare database and are available at https://figshare.com/authors/xiaoni_zhang/19517107. All resequencing data were available from NCBI (SRP119986 and BioProject PRJNA704782). The RNA sequence data were available from NCBI (PRJNA351281and PRJNA911454) and the National Genomic Data center Genome Sequence Archive (PRJCA008921; https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA008921). Source data are provided with this paper.

Code availability

All code supporting this study is publicly accessible online, with bash scripts for the Rose pangenome analysis available on GitHub (https://github.com/LanLBio/Rose_pangenome_scripts) and Zenodo (https://doi.org/10.5281/zenodo.18030525)[74].

References

41. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi–C experiments. *Cell Syst.* **3**, 95–98 (2016).
42. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi–C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
43. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi–C data. *Nat. Plants* **5**, 833–845 (2019).
44. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
45. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
46. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).
47. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
48. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
49. Haas, B. J. et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
50. Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J. & Clavijo, B. J. KAT: a k-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574–576 (2017).
51. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
52. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
53. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
54. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
55. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
56. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2021).
57. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
58. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).

59. Poplin, R. et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987 (2018).
60. Yun, T. et al. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. *Bioinformatics* **36**, 5582–5589 (2021).
61. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite—fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
62. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
63. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
64. Pierce, N. T., Irber, L., Reiter, T., Brooks, P. & Brown, C. T. Large-scale sequence comparisons with *sourmash*. *F1000Res.* **8**, 1006 (2019).
65. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
66. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
67. Ramírez-González, R. H. et al. The transcriptional landscape of polyploid wheat. *Science* **361**, eaar6089 (2018).
68. Garrison, E. et al. Building pangenome graphs. *Nat. Methods* **21**, 2008–2012 (2024).
69. Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).
70. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
71. Kong, Y., Wang, H., Lang, L., Dou, X. & Bai, J. Effect of developmental stages on genes involved in middle and downstream pathway of volatile terpene biosynthesis in rose petals. *Genes (Basel)* **13**, 1177 (2022).
72. Han, Y. et al. Comparative RNA-seq analysis of transcriptome dynamics during petal development in *Rosa chinensis*. *Sci. Rep.* **7**, 43382 (2017).
73. Yuan, L. Z., Rouviere, P. E., Larossa, R. A. & Suh, W. Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in *E. coli*. *Metab. Eng.* **8**, 79–90 (2006).
74. Lan, L. Rose_pangenome_scripts. *Zenodo* https://doi.org/10.5281/zenodo.18030525 (2025).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32470678 to W.P.), the Innovation Program of Chinese Academy of Agricultural Sciences and the Chinese Academy of Agricultural Sciences Elite Youth Program (110243160001007 to Z.W.). Additional supports were provided by the National Key R&D Program of China (2021YFD1200205 and 2019YFD1000400 to X.F.), the GuangDong Basic and Applied Basic Research Foundation (2024A1515012230 to X.Z.), the Scientific Research Foundation for Principle Investigator, Kunpeng Institute of Modern Agriculture at Foshan (KIMA-QD2022004 to Z.W.), the Funding of Major Scientific Research Tasks, Kunpeng Institute of Modern Agriculture at Foshan (KIMA-ZDKY2022004 to Z.W.), the Agricultural Science and Technology Innovation Program (CAAS-ZDRW202503 to W.P.) and the Youth Innovation Program of the Chinese Academy of Agricultural Sciences (Y2025QC36 to W.P.). Other supports are from the French National Institute of Research for Agriculture, Alimentation and Environment (INRAE to M. Bendahmane and J.S.) and from Partenariats Hubert Curien France-Chine Cai Yuanpei program (52528RB to M. Bendahmane and X.F.). The authors are grateful to X. Deng and Kirin Holdings for providing the pACCARΔ16crtX plasmid, and to X. Zhang and J. Yu for their valuable guidance on the polyploid genome.

Author contributions

W.P., Z.W., X.F. and M. Bendahmane conceived and managed the project. X.Z., L.L. and Y.Y. performed the bioinformatic analysis. H.G., Q.W., R.H. and B.H. planted the rose accessions and prepared the samples. H.G., Q.W. and R.H. performed the molecular experiments. D.P., H.J., X.L., S.L., J.S., C.B. and D.G. contributed to data analysis. X.Z., L.L., Y.Y. and H.G. analyzed the data, prepared the figures and tables, and wrote the paper. W.P., Z.W., X.F., M. Bendahmane and M. Bao revised the manuscript.

Competing interests

All authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02569-z.

Correspondence and requests for materials should be addressed to Mohammed Bendahmane, Xiaopeng Fu, Zhiqiang Wu or Weihua Pan.

Peer review information *Nature Genetics* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

nature portfolio

Corresponding author(s): Weihua Pan; Zhiqiang Wu; Xiaopeng Fu; Bendahmane Mohammed

Last updated by author(s): Dec 25, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	23 rose accessions were sequenced and generated the High-fidelity (HiFi) long reads and HIC data.
Data analysis	hifiasm v0.16.3, BLASTN v2.9.0, Juicer v1.6,3D-DNA 180922, AllHiC 0.9.13, GPhase v0.1.0-beta, Gap-Aid v1.1,Gap-Graph v1.0.1,EDTA v1.9.4, AUGUSTUS v3.3.3, SNAP (https://github.com/KorfLab/SNAP), GlimmerHMMv3.0.4,Exonerate v2.2.0, HiSAT2 v2.2.1, PASA v2.5.1, EVidenceModeller 1.1.1, Orthofinder v2.5.1, PAML4.8, CAFE5, minimap2, Samblaster V0.1.26, Deepvariant V1.4.0, GLnexus v1.4.1-0-g68e25e5, Dsuite V0.5 r44, Smudgeplot v0.2.2, iqtree V2.1.4-beta, Sourmash V4.3.0, PSMC V0.6.5-r67, StringTie v2.1.6, Mummer, BWA 0.7.17-r1188, Trimmomatic 0.36, Analyst 1.6.3,Multiquant 3.0.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All long read and Hi-C sequencing data have been deposited in the NCBI BioProject under accession PRJNA1194317, PRJNA1400323, PRJNA1400713, PRJNA1400325, PRJNA1400715, PRJNA1400330, PRJNA1400716, PRJNA1400345, PRJNA1400717, PRJNA1400382, PRJNA1400718, PRJNA1400411, PRJNA1400719, PRJNA1400417, PRJNA1400720, PRJNA1400440, PRJNA1400721, PRJNA1400442, PRJNA1400722, PRJNA1400443, PRJNA1400723, PRJNA1400444, PRJNA1400724, PRJNA1400306, PRJNA1400725, PRJNA1400449, PRJNA1400726, PRJNA1400451, PRJNA1400727, PRJNA1400477, PRJNA1400728, PRJNA1400478, PRJNA1400730, PRJNA1400479, PRJNA1400731, PRJNA1400481, PRJNA1400732, PRJNA1400490, PRJNA1400733, PRJNA1400492, PRJNA1400734, PRJNA1400494, PRJNA1400735, PRJNA1400736, PRJNA1400495, PRJNA1400737, PRJNA1400738, PRJNA1400739, PRJNA1400496, PRJNA1400740, PRJNA1400742, PRJNA1400743 and in the National Genomics Data Center under accession number PRJCA039059. All resequencing data were available from NCBI (SRP119986 and BioProject PRJNA704782). The RNA sequence data were available from NCBI (PRJNA351281 and PRJNA911454) and the National Genomic Data center (NGDc) Genome sequence Archive (GSA) (PRJCA008921).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For pan-genome construction, 26 representative rose accessions were selected based on their phylogenetic relationships within the rose germplasm, ensuring a comprehensive representation of genetic diversity.
Data exclusions	No data was excluded
Replication	For pigment detection, all samples were successfully used in this study with three biological replicates each. For RNA sequencing, all three biological replicates were successful aid used in this study, except for the tissue mixed samples used for annotation.
Randomization	Randomization does not directly apply to the genome sequencing and assembly.
Blinding	Blinding is not necessary for genome sequencing and assembly, since the investigators know which rose accessions they were handing

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☒	☐	Public health
☒	☐	National security
☒	☐	Crops and/or livestock
☒	☐	Ecosystems
☒	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	All rose accessions were cultivated in the Flower Garden of Huazhong Agricultural University.
Novel plant genotypes	N/A
Authentication	N/A