

Pangenome and resequencing analyses reveal flowering evolution and genetic control in *Cerasus*

Received: 17 October 2025

Accepted: 24 April 2026

Published online: 27 May 2026

 Check for updates

Songtao Jiu^{1,2,24} ✉, Yahui Lei^{3,24} , Linlin Fang^{1,2}, Yue Huang⁴ , Wenbo Chen⁵ , Yan Xu¹, Lei Wang^{1,2}, Zhengxin Lv¹, Xunju Liu^{1,2}, Lu Han¹ , Boyang Liu¹, Chen Zhang⁶, Xiao Dong³, Taoxian Zhang¹, Jiawei Wang⁷ , Xu Zhang^{2,8}, Yuliang Cai⁹, Wei Zheng¹⁰, Fangdong Li^{2,8}, Congli Liu¹¹, Hongwen Li¹², Jijun Zhu¹³, Fei Yu¹⁴, Ming Li¹¹, Jing Wang¹⁵, Long Chen¹⁶, Chunjin Yin¹⁷, Lei Peng¹⁸, Jiyuan Wang¹, Zhuo Zhang¹, Xiaojuan An^{1,2,19}, Lixia Yu²⁰, Ruie Liu^{1,2}, Yan Zhang¹⁴, Li Wang¹, Yaqin Wu¹⁶, Anthony Bernard²¹, Mitsuho Nakagomi²², Shiping Wang¹, Dirlwanger Elisabeth²¹, Dawei Li³, Qi Zhao³, Hongzhang Xue²³, Yang Dong³ , Zongyi Sun⁶ ✉, Shengchang Duan³  ✉ & Caixi Zhang^{1,2} ✉

Prunus subgenus *Cerasus* contains numerous species with ornamental, edible, and medicinal value. However, limited genomic resources have constrained systematic analyses of structural variation and the genetic basis of key phenological traits in this group. Here, we assemble eight genomes from diverse *Cerasus* species. Together with 13 published genomes, we construct a pangenome of 21 accessions representing 17 species. Phenological observations reveal substantial variation in flowering time. Integrating comparative genomics, transcriptomics, and population genetic analyses highlight candidate regulators of flowering time. We find that *AGAMOUS-LIKE 9* (*AGL9*) is strongly associated with flowering progression. Both ectopic expression and transient overexpression of *PavAGL9* can accelerate post-dormancy flowering progression. We reveal that *PavBPC6* binds the *PavAGL9* promoter and represses its transcription, indicating a negative regulatory role. Furthermore, *PavAGL9* interacts physically with *PavSEP1* and *PavPMADS2*, suggesting synergistic roles in floral organ development. Our pangenome resource establishes a comprehensive genomic framework for *Cerasus* and provides insights into the regulation of flowering progression.

The subgenus *Cerasus*, a major lineage within *Prunus* sensu lato, comprises ~70 species with ornamental, edible, and medicinal values. These species are distributed across temperate regions of Eurasia and North America, with centers of diversity in East Asia, including China, Japan, and the Korean Peninsula^{1–3}. Numerous *Cerasus* species, including sweet cherry (*P. avium*), sour cherry (*P. cerasus*), and Chinese cherry (*P. pseudocerasus*), are of considerable economic importance

with extensive cultivation⁴. In China, the cultivation of Chinese cherries dates back to the Zhou Dynasty, when they were regarded as precious offerings in imperial sacrificial ceremonies⁵.

Despite their long cultivation history and economic importance, the taxonomic classification and phylogenetic relationships within the subgenus *Cerasus* remain contentious^{3,6}. Several species traditionally referred to as dwarf cherries (*Microcerasus*), including *P. tianshanica*,

A full list of affiliations appears at the end of the paper. ✉ e-mail: jiusongtao@sjtu.edu.cn; sunzongyi@163.com; duanshengchang@163.com; acaizh@sjtu.edu.cn

P. humilis, *P. tomentosa*, *P. japonica*, *P. glandulosa*, and *P. pumila*, have historically been treated as a part of *Cerasus*^{4,7}. However, phylogenetic studies suggested that these species correspond more closely to sections *Amygdalus*, *Armeniaca*, or *Prunus*^{2,8,9} rather than forming a cohesive lineage within *Cerasus*. Although genome assemblies for several *Cerasus* species have become available, an overall phylogenetic framework within the group remains unresolved, leaving the evolutionary relationships among wild, domesticated, and ornamental species poorly understood.

Phenotypic variation in crop species arises from natural variation and selection for traits favored in domestication^{10,11}. Although short-read sequencing enables large-scale detection of single-nucleotide polymorphisms (SNPs) and small insertions/deletions (Indels), it is limited in detecting structural variations (SVs), including insertions, deletions, inversions, and translocations. Such variants often exert large effects on agronomic traits, such as flowering time, fruit size, stress tolerance, and plant architecture^{12–14}. Recent advances in long-read sequencing and graph-based pangenomes have enabled more accurate detection and comprehensive characterization of SVs, revealing forms of genetic variation that were previously inaccessible^{15–19}. These developments are particularly valuable for dissecting the genomic architecture of perennial woody plants such as cherries, whose highly heterozygous genomes have historically constrained genomic analyses.

De novo genome assemblies are available for several edible and ornamental cherry species^{20,21}, however, wild and geographically diverse *Cerasus* species remain underexplored. Pangenomes have advanced studies in numerous crops, including soybean (*Glycine max*)²², wheat (*Triticum* spp.)²³, barley (*Hordeum vulgare*)²⁴, tomato²⁵, strawberry²⁶, watermelon²⁷, apple²⁸, citrus²⁹, grapevine³⁰, and *Solanum*³¹. These efforts have demonstrated the value of capturing gene presence–absence variation (PAV), SVs, and population-scale diversity to elucidate trait evolution and support genomic-assisted breeding^{3,26,29,30}. These advances underscore the need for a comprehensive *Cerasus* pangenome that captures both cultivated and wild diversity to facilitate comparative genomic and trait dissection.

Here, we assemble eight genetically diverse *Cerasus* genomes and integrate them with 13 published genomes to construct a pangenome, CERASUSpan v1.0, representing the current global *Cerasus* diversity. Using this resource, we generate a variation map of SNPs, Indels, and SVs across 219 cherry accessions and conduct a genome-wide association study (GWAS), together with integrating candidate gene analyses to investigate the genetic basis of flowering time. Collectively, this pangenome resource establishes a comprehensive genomic framework for *Cerasus*, enabling us to elucidate the genetic regulation of flowering phase progression in cherries.

Results

The pangenome reference for *Cerasus* (CERASUSpan v1.0)

Cerasus, comprising ~70 species, is one of the most diverse, ecologically varied, and economically significant groups in *Prunus*⁴. To characterize its genomic landscape, we de novo assembled genomes of eight species, including the edible cherry *P. cerasus*, the ornamental *P. incisa* “Oshidori,” and six additional species (*P. campanulata*, *P. subhirtella*, *P. cerasoides*, *P. serrula*, *P. tianshanica*, and *P. mahaleb*) (Supplementary Data 1). Among these, *P. tianshanica* and *P. cerasus* were tetraploid³². Assembly contig N50 values ranged from 2.37 to 31.13 Mb, with haploid assembly sizes of 234.39–353.73 Mb, consistent with flow cytometry estimates (Supplementary Fig. 1 and Supplementary Data 2 and 3). Assembly qualities were further evaluated using the LTR Assembly Index (LAI), which ranged from 7.17 to 16.44 (Supplementary Data 2). Annotated gene models ranged from 22,864 to 32,588 (Supplementary Data 2). Using Hi-C data, the contigs were anchored into eight pseudochromosomes for all species (Supplementary Fig. 2), comparable to previous assemblies^{6,20,21,33–39}. Centromere regions were

identified, with lengths varying across chromosomes and species (Supplementary Data 4).

For pangenome construction, we integrated these eight de novo assemblies with 13 published genomes (*P. pseudocerasus*, *P. pusilliflora*, *P. conradinae*, *P. serrulata*, *P. fruticosa*, *P. humilis*, *P. campanulata*, *P. speciosa*, *P. spachiana*, and four sweet cherry cultivars “Burlat,” “Tieton,” “Regina,” and “13-33”) (Supplementary Data 1). In total, 21 accessions representing 17 species captured broad phylogenetic and phenotypic diversity (Fig. 1a, b and Supplementary Fig. 3), encompassing both cultivated and wild taxa. Based on use and domestication status, species were classified as wild (W), consumed (C), ornamental (O), and domesticated (D) (Fig. 1a, b). A phylogenetic tree constructed with 310 single-copy orthologs (Fig. 1c) showed that *P. avium* was more distantly related to *P. pseudocerasus* than to *P. cerasus* among edible species (Fig. 1c).

We annotated transposable elements (TEs) across the 21 genomes, ranging from 63.93 to 200.76 Mb (Supplementary Data 5). Seven major TE categories were identified (LINE, LTR, SINE, TIR, non-LTR, non-TIR, and rDNA), with LTR elements being the dominant component, accounting for 13.96–41.86% of total TEs (Fig. 1d and Supplementary Data 5). Consistent with other plant pangenomes^{31,40,41}, species-specific expansions of retrotransposons drove repetitive content and correlated strongly with genome size (Fig. 1e).

The simulated pangenome stabilized when >17 genomes were included, while core gene families stabilized beyond 19 genomes (Fig. 2a). Gene families were classified as core (7126; 12.53%), softcore (7350; 12.93%), dispensable (16,680; 29.33%), and private (25,709; 45.21%) (Supplementary Data 6). Core gene families were enriched in essential biological processes, including lipid oxidation, exocytosis, photosynthesis, and structural molecule activity, whereas dispensable gene families were mainly enriched in shoot development, protein phosphatase inhibitor activity, and signaling receptor activity (Supplementary Fig. 4). Across all gene families (Fig. 2b), gene counts were 219,617 (core), 209,672 (softcore), 225,232 (dispensable), and 28,641 (private) (Supplementary Data 7). Although dispensable and private families comprised 74.54% of the gene families, 62.84% of all genes belonged to core and softcore categories, which together accounted for 25.46% of all families, highlighting a conserved genomic backbone within the *Cerasus* pangenome (Fig. 2c and Supplementary Data 6 and 7).

To identify SVs, 20 genomes were aligned to the *P. avium* “Tieton” (Pavi-T) reference genome. SVs exhibited an uneven distribution along chromosomes (Fig. 2d). In total, we detected 493,836 SVs across the 20 genomes, with counts ranging from 6488 in *P. tianshanica* to 32,510 in *P. fruticosa* (Fig. 2e and Supplementary Data 8). These included 234,733 deletions and 240,638 insertions, whereas duplications (7947), inversions (774), and translocations (9744) were relatively rare (Fig. 2e, Supplementary Fig. 5a and Supplementary Data 8). Variant length distributions showed differences among SV types (Supplementary Fig. 5b–f) and correlated with chromosome length variation (Supplementary Fig. 6a and Supplementary Data 9). To capture structural variation and sequence diversity across *Cerasus* genomes, we constructed a graph-based pangenome using the VG toolkit, which enables representation of multiple haplotypes beyond a single linear reference (see Methods). Functions of genes affected by SVs (within or ±5 kb) varied across gene families, with softcore genes enriched in catalytic activity and transcription regulation, private genes enriched in heme binding, RNA modification, and defense responses (Supplementary Fig. 6b–d). Most deletions occurred along chromosome arms, whereas insertions clustered near or within centromeres, coinciding with high LTR density (Fig. 2f).

Variations in the blooming phenophase of the *Cerasus* species

Dormancy, budding, and flowering are critical developmental processes regulated by temperature, photoperiod, and hormonal cues.

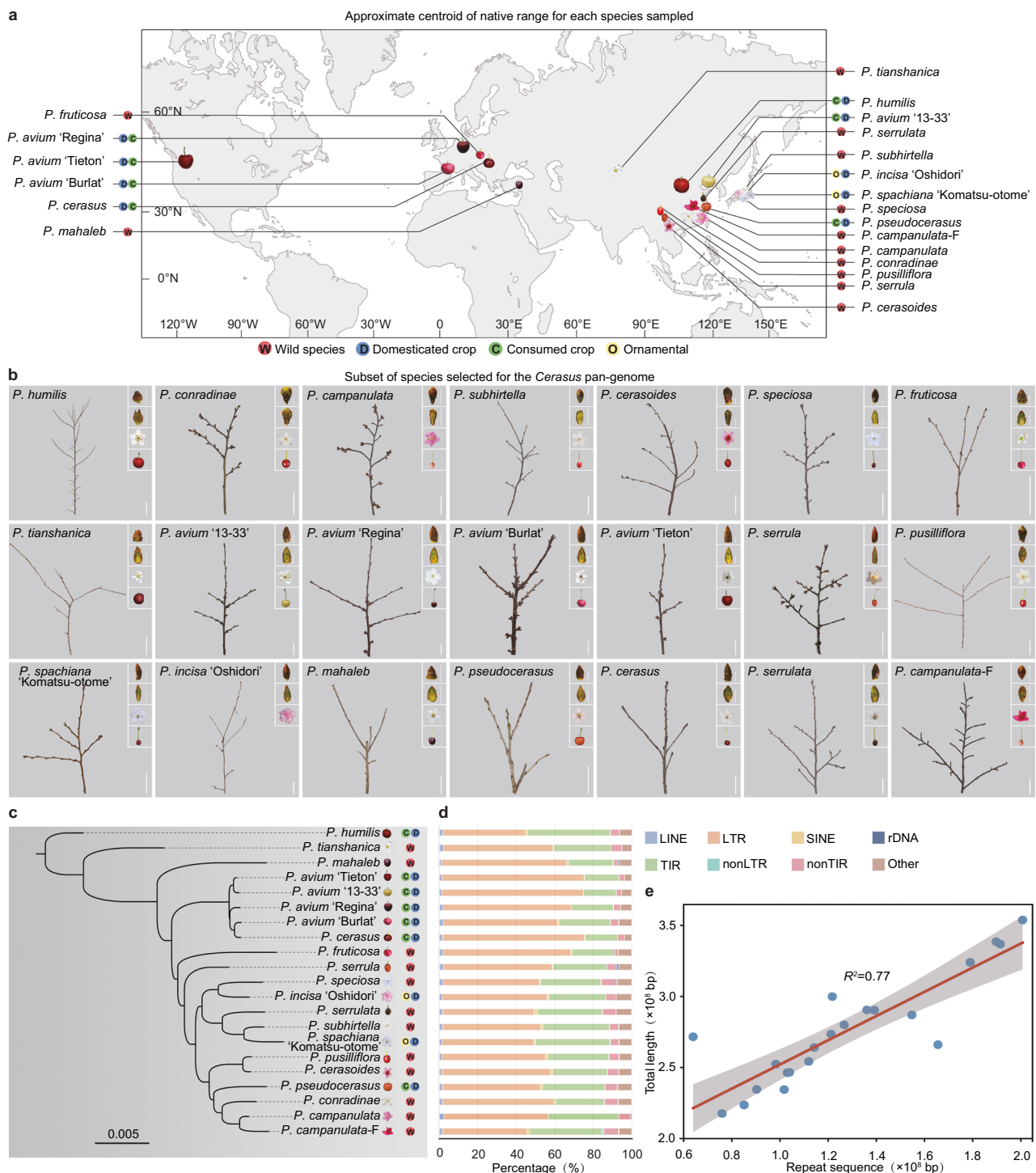


Fig. 1 | The *Cerasus* pan-genome captures the phenotypic, ecological, agricultural, and genomic diversity across the genus. **a** Approximate geographical centroid of the origin for the 21 selected *Cerasus* accessions, categorized by agricultural use: wild (W), consumed (C), ornamental (O), and domesticated crop (D). For some modern cultivated accessions, the indicated locations represent the breeding origin of the cultivar, rather than the native range of the corresponding species. **b** Representative phenotypic diversity of branches, buds, flowers, and

fruits from a subset of *Cerasus* species included in the pan-genome. Scale bars: 5 cm (shoots) and 2 mm (fruits, flower buds, and flowers). **c** Phylogenetic tree of the *Cerasus* pan-genome illustrating genetic relationships within the genus. Branch lengths represent coalescent units. **d** Ratio of different transposon elements (TEs) for each accession within the *Cerasus* pan-genome. **e** Pearson correlation coefficients ($R^2 = 0.77$) between genome size and total length of TEs. The gray band indicates the 95% confidence interval.

Plants adapt to climatic changes by evolving distinct germination and flowering periods. To evaluate flowering phenology, 21 *Cerasus* accessions were grown under uniform conditions in Shanghai, China. Substantial phenotypic diversity in blooming time was observed among these accessions (Fig. 3a and Supplementary Fig. 7), with *P.*

cerasoides, *P. campanulata*, and *P. pseudocerasus* flowering earlier than other species.

To clarify the phylogenetic relationships within subgenus *Cerasus* and between it and closely related subgenera (*Prunophora* and *Amygdalus*), we constructed a phylogenetic tree from single-copy orthologous

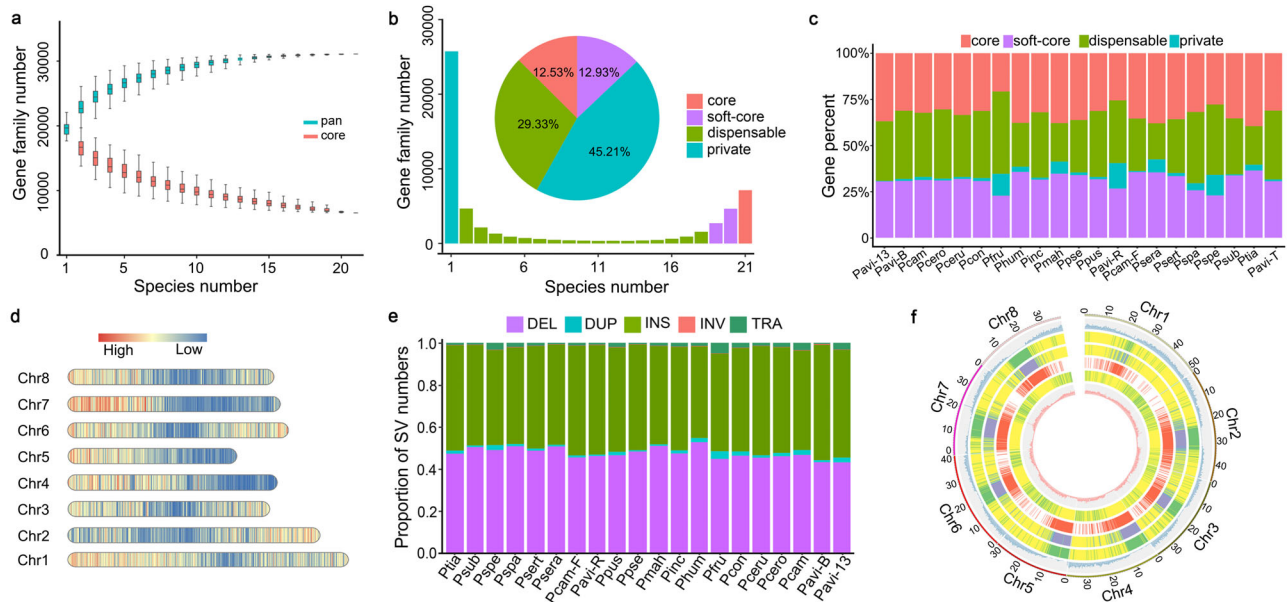


Fig. 2 | Pangenome characteristics of the cherry. **a** Number of pangenes and core gene families across different combinations of *Cerasus* genomes. The number of pan gene families and core gene families was independently stated 18 times when using a single genome and stated 21 times when using more than one genome. Each box represents the median and interquartile range (IQR), with whiskers extending to the maximum and minimum values. **b** Composition of the pangenome and individual genomes. The histogram displays the number of gene families across varying genomes, while the pie chart illustrates the proportion of gene families in each category. Gene families are classified as core (present in all genomes), softcore (over 90% of all genomes), private (over 90% of two genomes), and dispensable (the remaining) gene families. Orphan genes (that is, species-specific gene families with only one gene) were excluded. **c** Proportion of core, softcore, dispensable, and private genes within the *Cerasus* pangenome. **d** Structural variation (SV) density

across each chromosome. **e** Proportions of SV types in the *Cerasus* pangenome, including deletion (DEL), duplication (DUP), insertion (INS), inversion (INV), and translocation (TRA). **f** Distribution of genomic characteristic elements across each chromosome. The circle layers from outermost to innermost represent: chromosome length, deletion density, insertion density, transposon elements, LTR density, GC content, and gene density. Species abbreviations: *P. tianshanica* (Ptia), *P. subhirtella* (Psub), *P. speciosa* (Pspe), *P. spachiana* “Komatsu-otome” (Pspa), *P. serrulata* (Pserr), *P. serrula* (Psera), *P. campanulata-F* (Pcam-F), *P. avium* “Regina” (Pavi-R), *P. pusilliflora* (Ppus), *P. pseudocerasus* (Ppse), *P. mahaleb* (Pmah), *P. incisa* “Oshidori” (Pinc), *P. humilis* (Phum), *P. fruticosa* (Pfru), *P. conradinae* (Pcon), *P. cerasus* (Pceru), *P. cerasoides* (Pceero), *P. campanulata* (Pcam), *P. avium* “Burlat” (Pavi-B), *P. avium* “13-33” (Pavi-13), and *P. avium* “Tieton” (Pavi-T).

genes (Supplementary Data 10). *P. humilis*, *P. tianshanica*, and *P. tomentosa* (*Microcerasus*) clustered closer to *Prunophora* than to subg. *Cerasus*. *Microcerasus* and *Prunophora* diverged approximately 23.8 million years ago (15.1–33.4 Mya) during the Miocene (Fig. 3b). Subg. *Cerasus* was divided into five clades: Clade I (*P. mahaleb*) formed the basal lineage; Clade II included *P. avium*; Clade III contained *P. fruticosa*; and Clade IV comprised *P. serrula*; and Clade V encompassed *P. incisa* and *P. speciosa* (more distant from *P. serrulata*, *P. yedoensis*, *P. subhirtella*, and *P. spachiana*), together with *P. campanulata*, *P. conradinae* (Pcon), *P. pseudocerasus* (Ppse), *P. cerasoides*, and *P. pusilliflora* (Ppus). The high-altitude species *P. cerasoides* and *P. pusilliflora* formed sister taxa, diverging approximately 2.3 Mya (0.5–4.8 Mya).

We identified 411 flowering-related orthologous gene families across *Cerasus* genomes using PlantCFG⁴². PAV analysis showed that some species, such as *P. pseudocerasus* and *P. subhirtella*, lacked numerous flowering-related genes. However, Pearson correlation analysis revealed no significant association between PAV and flowering time, suggesting it may not be the primary determinant of flowering time variation (Fig. 3c and Supplementary Fig. 8).

Among 97 flowering genes characterized in *Arabidopsis thaliana* (Supplementary Data 11), 86, 74, and 103 homologs were identified in Pavi-T, *P. mahaleb* (Pmah), and Ppse, respectively (Supplementary Data 12). These genes were widely distributed across chromosomes (Supplementary Data 12). We sequenced 27 RNA-Seq datasets (Supplementary Data 13) and performed Weighted Gene Co-expression Network Analysis (WGCNA), identifying three modules in Pavi-T, two in Pmah, and two in Ppse with correlation coefficients >0.9 (Fig. 3d and Supplementary Figs. 9 and 10). In Pavi-T, the pink, turquoise, and yellow modules contained 4783, 7296, and 2552 genes, including 18, 15,

and 11 flowering-related genes, respectively (Fig. 3d). In Pmah, the purple and green modules comprised 5607 and 1135 genes, harboring 16 and 4 target genes. In Ppse, the brown and blue modules included 2387 and 3943 genes with 11 and 15 target genes (Fig. 3d).

Given the different flowering phenotypes—Pavi-T and Pmah bloom late, whereas Ppse blooms early—target genes aligned at comparable sampling stages in Pavi-T and Pmah but differed in Ppse. These targets were converted into *A. thaliana* homologs for cross-species comparison (Fig. 3e), yielding five genes from Pavi-T, four from Pmah, and seven from Ppse. Expression analysis based on FPKM values revealed that *AGAMOUS-LIKE 9* (*AGL9*) exhibited the strongest differential expression across the three species (Fig. 3f).

Genetic diversity and population structure analysis of *Cerasus*

We compiled a comprehensive genomic variation dataset for 219 cherry accessions using previously published sequencing data⁶. After selecting accessions (Supplementary Data 14) and removing SNPs with a max-missing rate ≥ 0.1 and minor allele frequency (MAF) ≤ 0.05 , we identified 853,236 SNPs and 217,080 Indels (Supplementary Data 15). SVs were characterized using the pangenome, yielding 4283 SVs (max-missing rate ≤ 0.3 and MAF ≥ 0.01), thus generating a high-resolution variation map for *Cerasus* (Supplementary Data 15 and Supplementary Fig. 11 and 12). To validate the reliability of SVs, 15 randomly selected SVs were experimentally confirmed by PCR, all showing genotype-consistent amplification patterns (Supplementary Fig. 13), supporting the reliability of the SV dataset for downstream analyses.

Phylogenetic reconstruction based on SVs delineated five major clusters (Clusters 1–5) (Fig. 4a). However, flowering phenology showed no strong correspondence with these clusters. A similar lack of

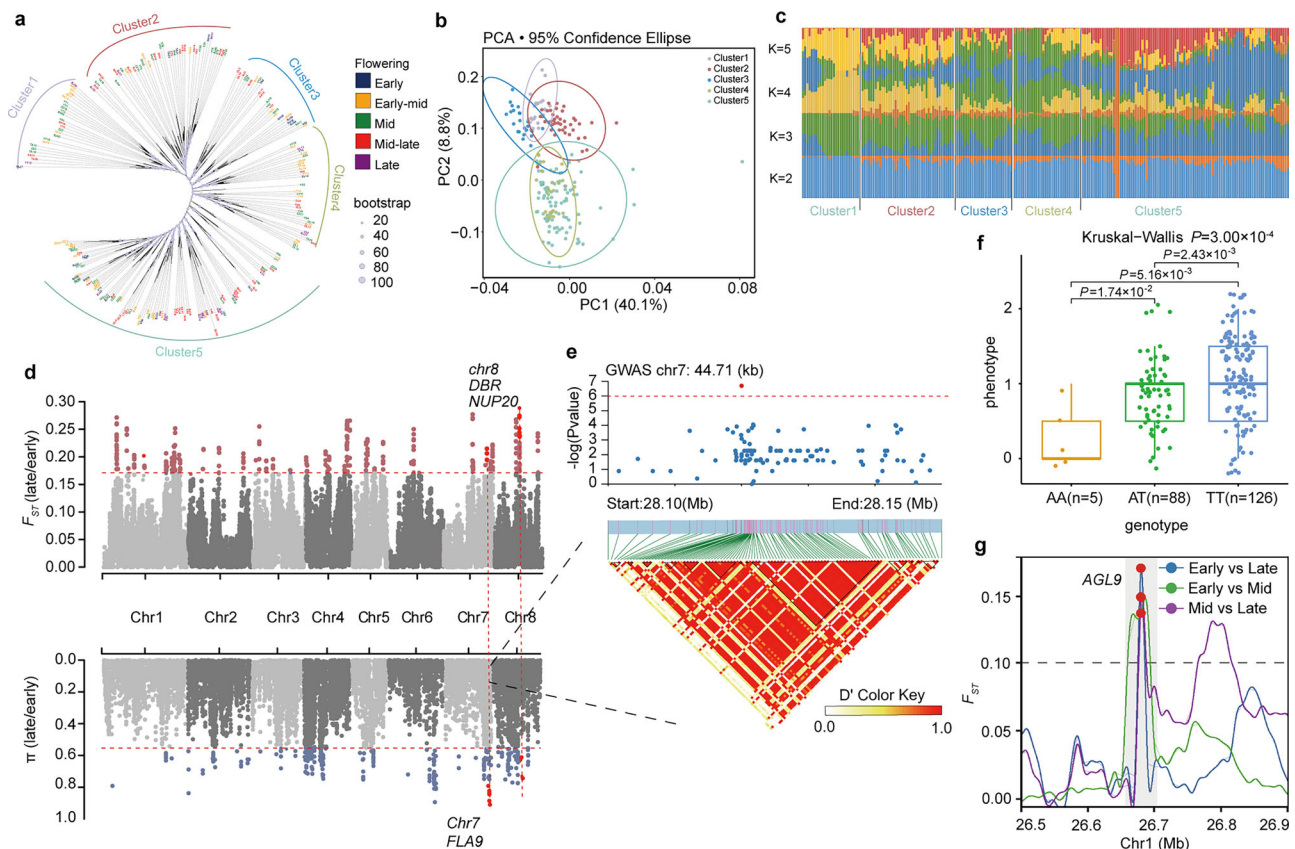


Fig. 4 | Population structure of cherry accessions and selection signatures for blooming phenophase. The neighbor-joining (NJ) phylogenetic tree (a), principal component analysis (PCA) (b), and admixture analysis (c) of 219 cherry accessions based on structural variation (SVs). d Manhattan plots showing the genomic regions under selection between early- and late-flowering groups based on the fixation index (F_{ST}) and nucleotide diversity (π) of SNPs. The dashed line represents the significant thresholds of π and F_{ST} values. e Local Manhattan plot (top) of the flowering-time association signals identified by genome-wide association analysis (GWAS) and the corresponding linkage disequilibrium (LD) heatmap (bottom) for the 44.71-kb interval on chromosome 7 (chr7: 28.10–28.15 Mb). Each dot represents an SNP, and the y-axis indicates the association significance as $-\log_{10} P$. Association P values were obtained using a GLM model implemented in TASSEL. The horizontal red dashed line indicates the significance threshold ($-\log_{10} P = 6$). Pairwise LD

across this interval is shown in the lower panel and is measured as D' . Red dots indicate non-synonymous SNPs within chr7: 28.10–28.15 Mb. f Box plots showing haplotypes at the upstream locus of the candidate genes *PAVO7G053290* (+) and *PAVO7G053280* (-). Phenotypic differences among genotype classes at the upstream SNPs of these candidate genes are shown. Each point represents a biologically independent sample. Sample sizes were AA, $n = 5$; AT, $n = 88$; and TT, $n = 126$. Phenotypic values were obtained from single measurements per sample. Box plots show the median (center line), upper and lower quartiles (box limits), and whiskers extending to $1.5 \times$ the interquartile range. Overall differences among genotype groups were assessed using a Kruskal–Wallis test ($*P < 0.05$, $**P < 0.01$, ns, $P > 0.05$), and pairwise comparisons were performed using two-sided Wilcoxon rank-sum tests. No conventional control group was included. g Zoom in on the selected region of the *AGL9* gene.

correspondence between flowering phenotypes and genetic relationships was observed in SNP- and Indel-based phylogenies (Supplementary Fig. 14a, b). This is likely attributable to the relatively narrow genetic background of cultivated sweet cherry, compounded by the limited genetic differentiation between populations.

To further investigate population structure, we performed ADMIXTURE (Fig. 4b) and principal component analyses (PCA, Fig. 4c). PCA revealed weak differentiation across clusters, consistent with limited diversity in selected sweet cherry. ADMIXTURE analysis further showed that most accessions harbored mixed ancestry components rather than forming discrete subpopulations (Supplementary Fig. 14c–f), further supporting the impact of long-term clonal propagation. Although the neutral genetic structure is shallow, this accession panel provides a robust framework for dissecting variation in flowering phenology in subsequent analyses.

Selection signature analysis of the blooming phenophase

Understanding the genetic basis of the blooming phenophase is essential for elucidating flowering regulation, evolutionary dynamics, and its potential applications in breeding. To investigate the genomic variations underlying flowering time in cherries, flowering-time

phenotypes of 219 cherry accessions were collected and then initially classified into five categories. For selection signature analyses, only the three extreme groups—early, mid, and late flowering (20, 70, and 21 accessions, respectively)—were retained, while the intermediate groups were excluded to maximize phenotypic contrast and statistical power (Supplementary Data 14).

F_{ST} analysis revealed varying genetic differentiation among groups, with the strongest divergence observed in the early–late comparison (SVs: top 1% threshold = 0.1921; SNPs = 0.1710; Indels = 0.1445), followed by mid–late (SVs = 0.1399; SNPs = 0.1173; Indels = 0.0971), and early–mid (SVs = 0.1354; SNPs = 0.0983; Indels = 0.0845) comparisons (Fig. 4d, Supplementary Figs. 15–17 and Supplementary Data 16–18). GO enrichment analysis linked these candidate genes of high F_{ST} to DNA helicase activity, vacuolar membrane, glutathione transferase activity, and ubiquitin ligase complex (Supplementary Fig. 18). These functional categories are known to participate in chromatin regulation, redox homeostasis, protein turnover, and stress-responsive signaling, processes that have been implicated in the regulation of flowering time and floral transition in multiple plant species^{43–45}.

To further identify genomic regions under selection associated with flowering time variation, the high π differences and F_{ST} values (top

1%) across the three comparisons were examined, thereby revealing 36 SNPs, 101 Indels, and seven SVs candidate regions in the early–late comparison, 87 SNPs, 68 Indels, and eight SVs regions in the early–mid comparison, and 17 Indels and four SVs in the mid–late comparison. Correspondingly, the candidate genes included 219 SNP-, 196 Indel-, and 35 SV-associated genes (early–mid comparison), 50 Indel- and 47 SV-associated genes (mid–late comparison), and 125 SNP-, 302 Indel-, and 55 SV-associated genes (early–late comparison) (Supplementary Data 16–18). The greater number of candidate genes in early-flowering comparisons suggests stronger selection pressure at this phenotypic extreme, driving divergence in flowering regulation.

Within SNP candidate regions distinguishing early- and late-flowering groups, the region with the highest π ratio value (chr7:28.1–28.15 Mb) contained a SNP (chr7:28,123,036; A/T) located upstream of two adjacent genes, *PAVO7G053280* (-) and *PAVO7G053290* (+), both of which were associated with flowering time in GWAS analyses (Fig. 4d–f). *PAVO7G053290* encodes NUP205, a nuclear pore complex protein previously shown to influence flowering via the circadian clock regulation in *Arabidopsis*^{46,47}. Additionally, *PAVO1G05220* (*PavAGL9*) exhibited strong selection signals across all SNP-based comparisons based on F_{ST} values (Fig. 4g). In addition, homology-based analyses identified four flowering-related genes in early–mid comparisons of SNPs and Indels: *PAVO1G067490* (*PavGA2ox8*), *PAVO1G068630* (*PavCSTF77*), *PAVO1G074490* (*PavDCL3*), and *PAVO7G053770* (*PavARP6*) (Supplementary Figs. 15 and 17 and Supplementary Data 19). Two non-synonymous SNPs were detected in *PavCSTF77*, while *PavDCL3* harbored six non-synonymous variants, potentially affecting protein function (Supplementary Fig. 19). In *A. thaliana*, *cstf77* mutants exhibit delayed flowering, and *dcl1* mutants display reduced microRNA expression^{48,49}. Haplotype analysis of *PAVO1G074490* (*PavDCL3*) identified a SNP (chr1:39,103,856; A/C) associated with flowering time, which resulted in a non-synonymous substitution, suggesting a potential role in flowering (Supplementary Fig. 19). Meanwhile, Indels located in the upstream regions of *PavARP6* and *PavDCL3* indicate possible *cis*-regulatory effects (Supplementary Fig. 20).

SV analysis during the mid-to-late comparison further identified *PAVO1G065550* (*PavGAI/RGA*), which encodes a DELLA protein involved in gibberellin signaling⁵⁰, within a candidate SV region (Supplementary Fig. 16). Similarly, in the early–late comparison, Indel variation highlighted *PAVO4G001540* (*PavCOPI*), a gene previously implicated in photoperiodic flowering regulation⁵¹ (Supplementary Fig. 17). These results highlight the multilayered genetic architecture of flowering time, with selection acting on both coding and regulatory regions. To further dissect flowering time variation, we performed a GWAS using phenotypic data (Supplementary Figs. 21–23 and Supplementary Data 20). Using GEMMA's linear model, we identified two SNPs, 15 SVs, and two Indels, suggesting potential functional roles in flowering phenotypes. A generalized linear model (GLM) in TASSEL detected significant associations for each variant type, further underscoring that both sequence and structural variation contribute to flowering-time regulation in sweet cherry (Supplementary Data 20).

Functional analysis of *PavAGL9* associated with flowering progression

Based on the integration of selection signature analyses (Fig. 4g) and RNA-seq expression clustering (Fig. 3e, f), *PavAGL9* was identified as a candidate gene associated with flowering phenology. To further characterize its expression pattern, we examined *PavAGL9* transcript abundance across developmental stages. Both semi-quantitative and quantitative PCR showed that *PavAGL9* expression increased markedly during flowering, supporting its involvement in this process (Fig. 5a–c). Similarly, *AGL9* orthologs exhibited comparable expression trends across multiple *Cerasus* species, indicating conserved transcriptional dynamics (Supplementary Fig. 24). Subcellular localization

assays revealed the nuclear localization of *PavAGL9* (Fig. 5d), consistent with its predicted role as a transcription factor. To explore its biological role, we generated transgenic *A. thaliana* lines overexpressing *PavAGL9* to assess its functional impact on plant development. These lines flowered earlier than the wild-type Col-0 plants (Fig. 5e). Expression analysis revealed increased transcript levels of *AtFT* in *PavAGL9*-overexpressing plants (Fig. 5f, g), consistent with known flowering-time pathways in *Arabidopsis*⁵². To evaluate whether *PavAGL9* exerts a comparable effect in sweet cherry, we transiently overexpressed *PavAGL9* in sweet cherry flower buds that had fulfilled their chilling requirement. *PavAGL9*-overexpressing buds flowered earlier than empty-vector controls, accompanied by a marked increase in *PavAGL9* transcript abundance (Fig. 5h, i). Moreover, expression of downstream flowering regulators *PavFT* and *PavAPI* was also significantly upregulated (Fig. 5j, k). Taken together, these results indicate that *PavAGL9* functions as a transcription factor associated with flowering progression in *Cerasus*.

PavBPC6 binds the *PavAGL9* promoter and represses its transcription to modulate flowering progression

To identify upstream regulators of *PavAGL9*, we first predicted 74 candidate transcription factors using PlantRegMap (https://plantregmap.gao-lab.org/regulation_prediction.php) (Supplementary Data 21). These candidates were further prioritized by integrating co-expression network analysis, which revealed that *PavBPC6* exhibited the strongest expression correlation with *PavAGL9* (Supplementary Data 22–27). Based on this combined bioinformatic evidence, the potential *PavBPC6*-*PavAGL9* regulatory module was selected for experimental validation. Yeast one-hybrid (Y1H) assays showed that yeast cells co-transformed with pB42AD-*PavBPC6* and pLacZi-pro-*PavAGL9* activated reporter gene expression, indicating a direct interaction between *PavBPC6* and the *PavAGL9* promoter (Fig. 6a). Motif analysis using PlantRegMap and FIMO (MEME Suite) identified putative *PavBPC6*-binding sites corresponding to an atypical GAGA motif within the *PavAGL9* promoter. Electrophoretic mobility shift assay (EMSA) further demonstrated that recombinant *PavBPC6*-GST protein bound biotin-labeled probes containing this motif *in vitro*, while unlabeled or mutated probes abolished binding (Fig. 6b). These results indicate that *PavBPC6* specifically binds this motif within the *PavAGL9* promoter. To determine the transcriptional effect of *PavBPC6*, we performed a dual-luciferase (Dual-LUC) assay in *Nicotiana benthamiana*. Co-expression of 35S:*PavBPC6* (effector) with pro-*PavAGL9*:LUC (reporter) significantly reduced the LUC/REN ratio compared with the reporter alone (Fig. 6c–e), indicating that *PavBPC6* functions as a transcriptional repressor of *PavAGL9*. To validate this regulatory relationship in cherries, we transiently overexpressed *PavBPC6* in sweet cherry flower buds that had satisfied chilling requirements. *PavBPC6*-overexpressing buds displayed delayed flowering relative to empty-vector controls, along with high-intensity GFP fluorescence (Fig. 6f, g). Consistently, *PavAGL9* transcript levels were significantly reduced in *PavBPC6*-overexpressing buds (Fig. 6h). Collectively, these results demonstrate that *PavBPC6* directly binds the *PavAGL9* promoter and represses its transcription, thereby contributing to the regulation of flowering progression in sweet cherry.

To explore the evolutionary conservation of this regulatory relationship, we compared upstream promoter sequences of *AGL9* orthologs across 17 *Cerasus* species represented in the pangenome dataset. A conserved motif corresponding to the *PavBPC6*-binding site was identified in 15 species (19 accessions), indicating strong conservation within subg. *Cerasus* (Supplementary Fig. 25). In contrast, the motif was absent in *P. tianshanica* and *P. humilis*, both belonging to *Microcerasus*, suggesting lineage-specific divergence. We further retrieved *BPC6* ortholog coding sequences from all 21 accessions and conducted protein–DNA binding affinity predictions using motif-scanning approaches (FIMO; <https://meme-suite.org/meme/tools/>

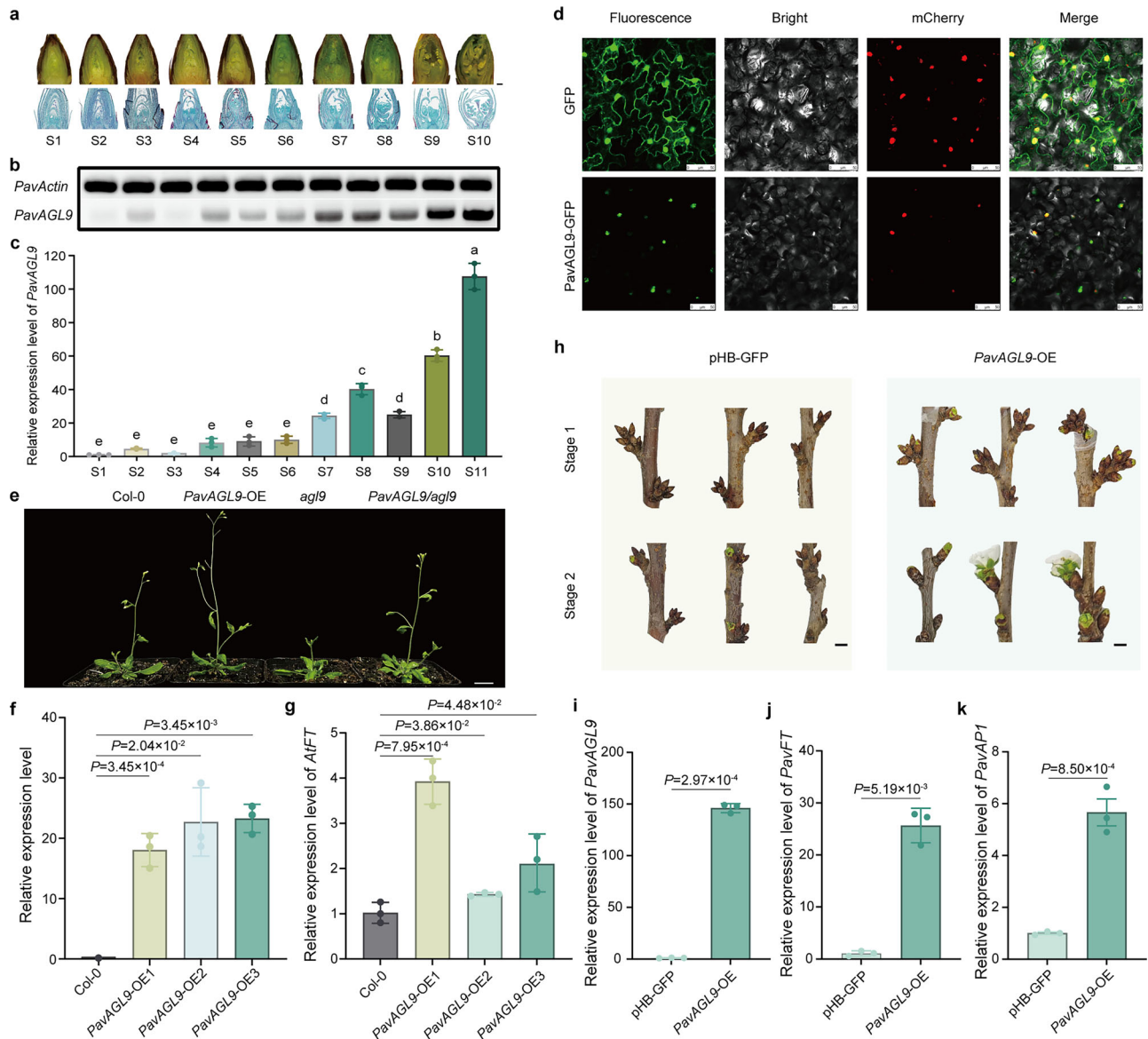


Fig. 5 | Overexpression of *PavAGL9* accelerates the blooming phenophase in cherries. **a** Phenotypic changes observed during cherry flower bud formation and dormancy release. Scale bar = 500 μ m. Both semi-quantitative (**b**) and quantitative PCR (**c**) analyses demonstrated a substantial upregulation of *PavAGL9* expression during the stages of flower bud formation and dormancy release. Data are presented as mean \pm SD ($n = 3$ biological replicates). Different lowercase letters indicate significant differences at $P < 0.05$ (one-way ANOVA followed by Tukey's multiple comparison test). S1-S11 correspond to the 11 sampling time points from 24 June 2024 to 6 April 2025. **d** Subcellular localization of the *PavAGL9* protein in tobacco leaf epidermal cells, with 35S:GFP as the positive control. Green fluorescence represents the GFP fusion protein signal, while red fluorescence marks the nucleus using p2300-mCherry. Two independent experiments with similar results

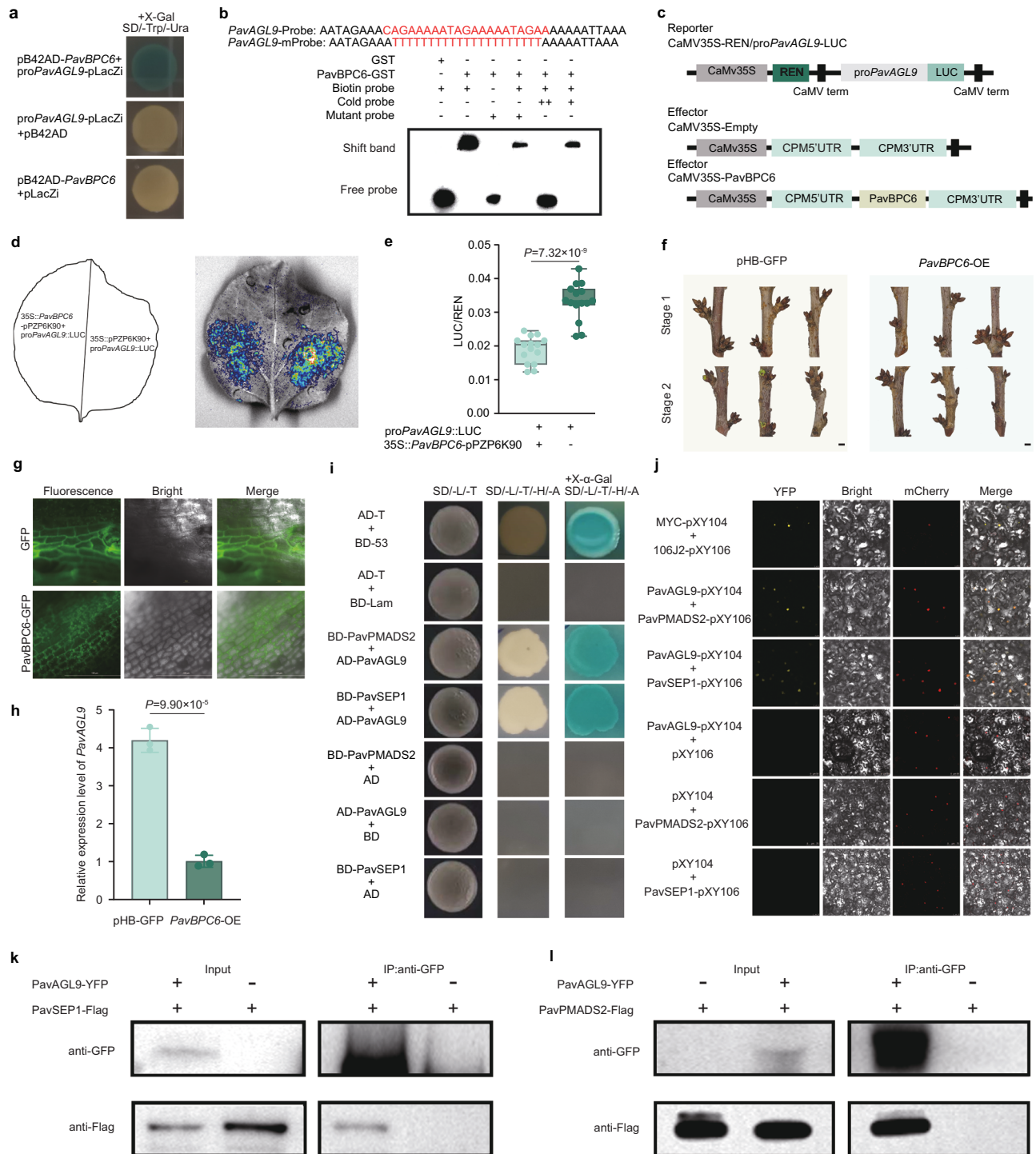
were performed. Scale bar = 50 μ m. **e** Phenotypes of Col-0, *PavAGL9*-OE, *agl9*, and *PavAGL9/agl9* plants. Scale bar = 2 cm. Expression profiles of *PavAGL9* (**f**) and *AtFT* (**g**) in Col-0 and transgenic lines are also depicted. Data are presented as mean \pm SD ($n = 3$ biological replicates). Statistical significance between groups was assessed using a two-sided Student's *t*-test. **h** Flowering phenotypes following the transient overexpression of *PavAGL9* in sweet cherry flower buds. Scale bar = 1 cm. Expression profiles of *PavAGL9* (**i**), *PavFT* (**j**), and *PavAPI* (**k**) in Col-0 and transient transgenic buds. Data are presented as mean \pm SD ($n = 3$ biological replicates). Statistical significance between groups was assessed using a two-sided Student's *t*-test. The *Actin* gene was used as the internal control. Source data are provided as a Source data file.

fimo). *In silico* analyses indicated that *BPC6* orthologs from 15 species (19 accessions) retain predicted binding capacity to the conserved motif in their respective *AGL9* promoters, suggesting that the BPC6-*AGL9* regulatory module is evolutionarily conserved within subg. *Cerasus* (Supplementary Data 28).

PavAGL9 physically interacts with PavSEPI and PavPMADS2

To investigate the involvement of *PavAGL9* in MADS-box protein complex assembly, we analyzed the protein-protein interactions among MADS-box proteins using yeast two-hybrid (Y2H) assays. *PavAGL9* was fused into pGADT7 (AD), whereas *PavSEPI* and

PavPMADS2 were fused into pGBKT7 (BD). Yeast cells harboring BD-*PavPMADS2* or BD-*PavSEPI* with AD-*PavAGL9* showed strong blue colony coloration, indicating direct interactions between *PavAGL9* and both *PavSEPI* and *PavPMADS2* (Fig. 6i). These interactions were further validated *in planta* using bimolecular fluorescence complementation (BiFC) assays in *N. benthamiana* leaves. Strong YFP fluorescence signals were observed in the nucleus when *PavAGL9*-cYFP was co-expressed with nYFP-*PavSEPI* or nYFP-*PavPMADS2*, whereas no fluorescence was observed in negative controls lacking one interaction partner (Fig. 6j). The MYC-cYFP and nYFP-106J2 pair were used as a positive control. To further confirm these interactions, we performed



co-immunoprecipitation (co-IP) assays in tobacco cells co-expressing PavAGL9-YFP with PavSEP1-Flag or PavPMADS2-Flag. Immunoprecipitation of PavSEP1-Flag or PavPMADS2-Flag pulled down PavAGL9-YFP, confirming their physical association in vivo (Fig. 6k, l). Taken together, these results suggest that PavAGL9 interacts with PavSEP1 and PavPMADS2, supporting its involvement in MADS-box protein complexes potentially associated with floral organ development.

Discussion

Plant pangenomes are rapidly emerging as foundational assets to guide genome-editing strategies and improve our understanding of genotype–phenotype relationships. While integrating forward genetics with large-scale sequencing of high-quality plant reference

genomes—including complete telomere-to-telomere assemblies—enables the discovery of SVs, SNPs, Indels, and CNVs, tracing their evolutionary trajectories remains challenging due to the complex evolutionary history of angiosperms, marked by lineage-specific fragmentation and recent duplications. Compared with species-level pangenomes, genus-level pangenomes capture broader sequence variation and paralog diversification, providing a genomic framework for trait dissection and evolutionary relationships within *Cerasus*.

Within *Prunus sensu lato*, long-standing debates regarding the phylogenetic relationship of *Cerasus* and its closely related lineages have persisted, largely due to reliance on morphological traits or limited molecular markers. By leveraging a pangenome-scale comparative framework across 21 genomes, this study resolved several

Fig. 6 | PavBPC6 binds to the promoter of *PavAGL9*, inhibiting its transcription and participating in the regulation of the blooming phenophase. **a** Yeast one-hybrid (Y1H) analysis demonstrating the binding of PavBPC6 to the *PavAGL9* promoter on selective medium lacking Ura and Trp, supplemented with 100 mg/mL X-Gal. **b** Electrophoretic mobility shift assay (EMSA) indicating the interaction of PavBPC6 with *PavAGL9* promoter fragments containing the motif. Unlabeled probes were utilized for the competition assay; “–” and “+” denote the absence and presence of the binding, respectively. **c** Schematic overview of the LUC reporter vector containing the *PavAGL9* promoter, alongside effector vectors containing PavBPC6. **d** Transient LUC imaging assays reveal that PavBPC6 represses the transcription of the pro*PavAGL9*:LUC reporter. Representative images show LUC activity in *N. benthamiana* leaves 48 h post-infiltration. Three independent experiments with similar results were performed. **e** Dual-Luciferase (Dual-LUC) assay depicting promoter activity expressed as the ratio of LUC to 35S:Renilla (REN). Data are presented as means \pm SD of fifteen independent measurements ($n = 15$ biological replicates), with significant differences evaluated using two-sided Student's *t*-test (centerline, median; box limits, first and third quartiles; whiskers, minimum and maximum values). **f** The phenotypes resulting from the transient overexpression of *PavBPC6* in flower buds of sweet cherry. Scale bar = 1 cm. **g** GFP

signal in the transiently transformed flower buds was detected using a fluorescence microscope. Two independent experiments with similar results were performed. **h** Expression profiles of *PavAGL9* in transient transgenic buds of sweet cherry. Data are presented as mean \pm SD ($n = 3$ biological replicates). Statistical significance between groups was assessed using a two-sided Student's *t*-test. **i** Yeast two-hybrid (Y2H) analysis demonstrating the interaction between *PavAGL9* and *PavSEPI1*, as well as *PavPMADS2*. Protein interactions were screened on a selective medium lacking Trp, His, Leu, and Ade, and supplemented with 20 mM X- α -Gal. AD-T and BD-lam were co-transferred as negative controls, while AD-T and BD-53 served as positive controls. **j** Bimolecular fluorescence complementation (BiFC) assay illustrating the interaction between *PavAGL9* and *PavSEPI1* as well as *PavPMADS2* in tobacco cells. MYC-pXY104 and 106j2-pXY106 pairs were used as a positive control. Yellow fluorescence represents the YFP fusion protein signal, while red fluorescence marks the nucleus using p2300-mCherry. Images show overlays of fluorescence and light views. Two independent experiments with similar results were performed. **k** Co-Immunoprecipitation (Co-IP) assay showing interaction of *PavAGL9* with *PavSEPI1*. **l** Co-IP assay demonstrating the interaction of *PavAGL9* with *PavPMADS2*. *PavSEPI1*-Flag and *PavPMADS2*-Flag were each co-expressed with *PavAGL9*-YFP in tobacco leaves. Source data are provided as a Source data file.

stable clades within *Cerasus* that are broadly consistent with recent chloroplast-based phylogenies^{9,33}, while revealing discordance with traditional morphology-based classifications⁴. In particular, several species historically assigned to *Microcerasus* (*P. tianshanica*, *P. humilis*, and *P. tomentosa*) did not form a monophyletic lineage within *Cerasus*, but instead cluster more closely with other *Prunus* sections, thereby corroborating earlier molecular evidence. Although taxonomic revision was beyond the scope of this study, these findings demonstrate how genus-level pangenomes can bridge traditional classifications with genome-scale evolutionary evidence and provide a foundation for downstream analyses.

By integrating 13 published genomes^{20,35,39} with eight newly assembled accessions, we constructed a 21-accession *Cerasus* pangenome. Although this dataset failed to capture the full sequence diversity present within the *Cerasus* population, particularly among wild species in China, the broad representation it provides offers key insights into diversity, despite the smaller number of accessions compared to some plant pangenomes^{28,31}. Numerous SVs were identified, including variants associated with fruit traits and, to a lesser extent, flowering time. These resources provide a valuable foundation for molecular breeding and genome editing.

Although flowering phenotypes were recorded under long-term uniform conditions to minimize environmental effects, residual phenotypic plasticity may persist in accessions originating from highly distinct native habitats. In the absence of explicit quantification of phenotypic plasticity or incorporation of environmental covariates, the GWAS and selection signals may partially reflect genotype-by-environment interactions and should therefore be interpreted with caution. Future multi-location trials and multi-year phenological monitoring will help disentangle the relative contributions of genetic and environmental factors to flowering-time variation.

Through combined selective sweep and transcriptomic analyses, *PavAGL9* was identified as a key gene associated with flowering progression. Expression of *AGL9* orthologs increased during post-dormancy stages and peaked at full bloom (Supplementary Fig. 24), and functional assays in *Arabidopsis* and sweet cherry buds that had fulfilled their chilling requirements demonstrated that elevated *PavAGL9* expression accelerates floral opening. In contrast, ectopic expression in buds that had not satisfied chilling requirements failed to promote bud break (Supplementary Fig. 26), indicating that *PavAGL9* functions downstream of dormancy release and is more closely associated with post-dormancy flowering progression rather than as a primary regulator of dormancy release.

Moreover, definitive functional validation of *PavAGL9* in sweet cherry will require stable genetic manipulation, which remains

technically challenging due to long juvenile phases, self-incompatibility, low fruit set, and the limited efficiency of transformation systems in *Cerasus* species^{33,34,54}. Although PavBPC6 was validated as a transcriptional repressor of *PavAGL9* in sweet cherry by EMSA and Dual-LUC assays, extension of this regulatory relationship across *Cerasus* species is currently supported primarily by sequence conservation and computational predictions. Accordingly, the inference of a conserved regulatory module remains provisional and requires further cross-species experimental validation. Substantial interspecific variation in flowering phenology suggests that additional regulatory layers, including variation in upstream regulators, downstream targets, or epigenetic modulation, may also contribute to species-specific flowering patterns. Elucidating these mechanisms represents an important direction for future studies.

Our selective sweep and GWAS analysis identified a set of flowering-related genes within the selective regions. Although several genes were functionally verified, future studies should refine candidate identification using more robust strategies. Considerable progress has been made in linking cherry horticultural traits to candidate genes^{55–58}, yet compared with other perennial fruit crops such as apple, citrus, and grapevine, functional genomics in cherries is lagging. In addition, although this study integrates population genomics, the regulatory mechanism of SVs remains incomplete. In particular, establishing causal links between SVs in *cis*-regulatory regions and flowering phenotypes will ultimately require cross-species expression profiling and targeted functional perturbations, which remain challenging in cherry. Overall, this work expands genomic resources for *Cerasus*, providing a pangenome dataset that deepens understanding of flowering biology and supports future molecular breeding efforts.

Methods

Sample collection and sequencing

The *Cerasus* species samples used in this study for DNA and RNA sequencing were collected from six locations across China. *P. cerasoides* (Pcero) was collected from Yunnan Agricultural University, Kunming, China (102°45'36"E, 25°8'24"N), *P. cerasus* (Pceru) from the Cherry Experimental Station of Northwest A&F University, Tongchuan, China (108°51'0"E, 34°52'12"N), *P. incisa* “Oshidori” (Pinc), *P. mahaleb* (Pmah), and *P. subhirtella* (Psub) from the Shanghai Botanical Garden, Shanghai, China (121°27'4"E, 31°9'14"N), *P. campanulata* (Pcam) from the Botanical Garden of Shanghai Jiao Tong University, Shanghai, China (121°26'24"E, 31°1'48"N), *P. serrula* (Psera) from Shangri-la, Yunnan, China (99°49'12"E, 27°49'12"N), *P. tianshanica* (Ptia) from the Ili Kazakh Autonomous Prefecture, Xinjiang, China (81°57'0"E, 43°7'12"N). The flower buds of sweet cherry “Tieton” were collected at

11 sampling time points: 24 June, 14 July, 2 August, 22 August, 10 September, 21 October, 23 November 2024, and 8 January, 23 February, 25 March, and 6 April 2025, for anatomical and expression analysis. Additionally, branches bearing flower buds that had fulfilled chilling requirements were collected from sweet cherries for transient transformation assays.

High-purity genomic DNA was extracted using a DNeasy Plant Kit (Tiangen Biotech Co., Ltd., Beijing, China). After assessing DNA purity, concentration, and integrity, both long- and short-read sequencing libraries were constructed. Short-read libraries were sequenced on an Illumina HiSeq X Ten platform (Illumina Inc., San Diego, CA, USA). Hi-C libraries were constructed through chromatin extraction, digestion, ligation, purification, and fragmentation⁵⁹, followed by sequencing on the Illumina HiSeq X Ten platform. ONT libraries were sequenced on the Oxford Nanopore PromethION 48 platform. For PacBio HiFi sequencing of Ptia, Psera, and Pmah, a standard SMRTbell library was prepared from 50 µg of DNA using the SMRTbell Express Template Prep Kit 2.0 and sequenced on a PacBio Sequel II (Pacific Biosciences, Menlo Park, CA, USA).

Total RNA for both RNA-seq and qRT-PCR was extracted using the RNAPrep Pure Plant Kit (TIANGEN Biotech Co., Ltd., Beijing, China). Reverse transcription was performed using the PrimeScript RT Master Mix Kit (Takara, Dalian), and cDNA was synthesized from 1 µg of total RNA for subsequent experiments. The cDNA libraries were sequenced on an Illumina NovaSeq platform (Bluescape Hebei Biotech Co., Ltd., China). RNA-seq data were generated from floral bud samples of Pavi-T, Pmah, and Ppse at three time points (T1–T3) during the 2024 growing season, corresponding to 20 February, 10 March, and 25 March, respectively.

Phylogenetic analysis of chloroplast genomes

Chloroplast genomes of *P. mahaleb* (NCBI accession number: PX991928), *P. spachiana* (PX991929), and *P. tianshanica* (PX991930) were assembled using the GetOrganelle v1.7.7.1⁶⁰. A total of 66 chloroplast genomes of *Prunus* species were retrieved from NCBI (Supplementary Data 29). Sequences were aligned using MAFFT v7.525⁶¹, then the maximum-likelihood (ML) tree was constructed using IQ-TREE v1.6.12⁶² with 1000 bootstrap replicates. *P. padus* was used as the outgroup.

Evaluation of flowering phenophase

We introduced several wild species from Yunnan and Xinjiang to the Cherry Experimental Farm at Shanghai Jiao Tong University by transplanting whole plants, collecting seedlings from root sprouts, or applying grafting and cutting, thereby minimizing variations in phenological periods due to regional differences. Flowering phenophases for all species were recorded under uniform conditions at botanical gardens and experimental farms in Shanghai⁶³. Flowering stages, including first swelling, green tip, and full bloom (defined as 50% flower opening), were recorded for 21 *Cerasus* accessions. *P. cerasoides*, an extremely early flowering species⁶³, served as the control. Species blooming within 30 days after *P. cerasoides* were classified as “early,” those flowering 30–41 days later as “middle,” and those blooming more than 41 days later as “late.” Specifically, the flowering phenophase was evaluated for *P. spachiana* “Komatsu-otome” (Pspa), *P. serrulata* (Psert), and *P. speciosa* (Pspe) at the Shanghai Chenshan Botanical Garden (121° 11' 24" E, 31° 4' 48" N); *P. incisa* “Oshidori”, *P. subhirtella*, *P. conradinae*, *P. cerasoides*, and *P. mahaleb* at the Shanghai Botanical Garden; *P. pusilliflora*, *P. serula*, *P. tianshanica*, *P. avium* [“Burlat,” “Tieton” (Pavi-T), “13-33” (Pavi-13) and “Regina” (Pavi-R)], *P. pseudocerasus*, *Prunus humilis* (Phum), and *Prunus cerasus* at the Cherry Experimental Farm at Shanghai Jiao Tong University. In addition, we assessed early-flowering (*P. campanulata*, Pcam) and late-flowering (*P. campanulata*-F, Pcam-F) types at the Botanical Garden of Shanghai Jiao Tong University. Flowering phenophase data from 219 cherry accessions were evaluated based on the date of full

bloom (defined as 50% flower opening). The well-known sweet cherry cultivar “Van,” characterized by a middle flowering time⁶⁴, was used as a reference. Accessions blooming within ±3 days of Van’s full bloom were classified as “middle flowering.” Those blooming 4–10 days earlier and 11–17 days earlier than “Van” were categorized as “early-middle” and “early,” respectively. Similarly, accessions blooming 4–10 days and 11–17 days later than “Van” were classified as “middle-late” and “late.”

Genome assembly and assessment

Before assembly, the genome sizes of the eight *Cerasus* accessions were estimated using flow cytometry. Preparation of the cell suspensions involved chopping fresh leaf tissue in nuclei isolation buffer, followed by filtration through a nylon mesh and staining with propidium iodide^{65,66}. Stained nuclear suspensions were analyzed with a BD FACSCalibur flow cytometer, where propidium iodide fluorescence was detected under 488 nm blue-light excitation. Genome sizes were independently estimated using *Solanum lycopersicum* or *Ricinus communis* as internal reference standards. For quality control, Fastp v0.20.2⁶⁷ was used to process NGS data, including Hi-C, RNA-Seq, and whole-genome paired-end reads, with default parameters. Nanopore reads were assembled de novo using NECAT v0.0.1⁶⁸. Redundant sequences were removed using purge_dups v1.2.5⁶⁹, and the contigs were polished through three iterative rounds using Racon with NGS data⁷⁰.

Chromosome anchoring was performed with Hi-C data. Clean Hi-C reads were mapped to the draft assembly using Juicer v1.7.6⁷¹, and chromosome-scale scaffolds were generated with the 3D-DNA pipeline v180922⁷². Hi-C heatmaps were manually inspected and corrected with Juicebox Assembly v1.9.1⁷³, followed by re-scaffolding using 3D-DNA. NGS reads were aligned to the assembly with Burrows-Wheeler Aligner v0.7.12⁷⁴. For PacBio Sequel II HiFi reads, genomes were assembled using hifiasm v0.19.9-r616⁷⁵ with default parameters, followed by the identical scaffolding and correction procedures. Genome completeness and accuracy of these genomes were evaluated using BUSCO v5.3.1⁷⁶. Centromeres were predicted with quar-TeT v1.2.5⁷⁷. The LAI was calculated using LTR retriever v3.0.4 with default parameters to evaluate assembly quality (https://github.com/oushujun/LTR_retriever)⁷⁸.

Genome annotation, GO, and KEGG enrichment analysis

TEs were annotated using EDTA v2.2.2⁷⁹ with parameters (--sensitive 1 --anno 1 --to evaluate 1 --force 1). Protein-coding genes were predicted using a combined strategy integrating ab initio, homology-based, and transcriptome-based approaches. Ab initio predictions were performed with Augustus v3.0.3⁸⁰, SNAP v2006-07-28⁸¹, and GlimmHMM v3.0.1⁸². Homology-based predictions were conducted using Exonerate v2.2.0⁸³, with sequences from *Prunus cerasus* (PRJNA295439 and PRJNA327561), *P. avium* (PRJNA73727, PRJNA419491, PRJNA595502, and PRJNA550274), and *P. subhirtella* (PRJNA596558). Transcriptome-based gene models were generated using StringTie v1.3.4⁸⁴ and PASA⁸⁵, based on Illumina sequencing data (PRJNA260424 and PRJNA1041553). All gene models were integrated using Evidence Modeler v1.1.1⁸⁶.

Functional annotation was performed by assessing sequence similarity and domain conservation. BLAST searches were conducted against NR, KEGG, and Swiss-Prot databases. HMMER v3.0⁸⁷ was used for Pfam domain identification, and InterProScan v5.53-87.0⁸⁸ was used to assign GO terms. Non-coding RNAs were predicted with tRNAscan-SE v1.3.1 (tRNA)⁸⁹, RNAmmer v1.2 (rRNA)⁹⁰, and INFERNAL v1.1.2 (miRNA and snRNA)⁹¹. Additional ncRNAs were annotated with Rfam v1.0.4⁹². GO and KEGG enrichment analyses were performed by using clusterProfiler⁹³.

Pangenome construction

Initially, Pavi-T was selected as the reference genome, and the remaining 20 genomes were aligned against it using minimap2 v2.1 (-x asm5)⁹⁴. Subsequently, SVs were identified using SyRi v1.6.3 (--nosnp)⁹⁵

and Assemblytics v1.2.1⁹⁶. The SV-calling results were merged using SURVIVOR v1.0.7 (merge 200,1,1,1,0,50)⁹⁷. A graph-based pangenome was constructed based on the SV-calling results using VG v1.53.0⁹⁸, which enables the integration of structural variations into a variation graph, allowing accurate read mapping and genotyping across diverse genomes beyond the constraints of a single linear reference. For gene analysis, GffRead v0.12.7⁹⁹ was used to extract the longest CDS and protein sequences from the 21 accessions. Gene families were identified using OrthoFinder v3.0.1b1¹⁰⁰, and subsequently classified as core, softcore, dispensable, and private according to the number of sample accessions represented in each orthologous group.

Phylogenetic analysis of cherry genomes

To investigate the phylogenetic relationships within *Cerasus*, we analyzed 21 accessions, including 13 publicly available genomes and eight newly assembled in this study (Supplementary Data 1). Seven additional *Prunus* species were included: *P. cerasus* (haplotypes A and F), *P. dulcis*, *P. mume*, *P. persica*, *P. tomentosa*, and *P. yedoensis*. *Rosa chinensis*, *Vitis vinifera*, and *A. thaliana* served as outgroups. Gene families across all genomes were identified using OrthoFinder v3.0.1b1¹⁰⁰. Single-copy orthologs were aligned using MUSCLE v3.8.31¹⁰¹, and well-aligned regions were retained using Gblocks v0.91b¹⁰². The best-fit substitution model was determined by ModelTest-NG v0.1.7¹⁰³, and a maximum likelihood (ML) phylogeny was inferred using RAXML-NG v1.2.2¹⁰⁴ with 1000 bootstrap replicates. Divergence times were estimated with the MCMCTree program in the PAML package v4.9j¹⁰⁵. To ground divergence-time estimation in reliable paleobotanical evidence, we constrained the *Prunus* stem node using the early Eocene *P. wutuensis* fossil (-55 Mya)¹⁰⁶, implemented as a soft calibration bound in line with established Rosaceae dating practices. We further incorporated two widely supported secondary calibration intervals (*Arabidopsis-Vitis*: 110–124 Mya; *Vitis-Rosa*: 102–113 Mya) to stabilize deeper nodes. An initial substitution-rate estimate obtained with *base1* was then used to parameterize the rate prior in MCMCTree. Gene family expansion and contraction across the phylogeny were assessed with CAFE v5¹⁰⁷.

Analysis of flowering genes

Flowering genes were retrieved from the *A. thaliana* database (<https://www.arabidopsis.org/>) and used as targets for scanning the Pavi-T, Pmah, and Ppse genomes. These genes were aligned against the protein sequences of the three genomes using BLASTP v2.16.0 (E value = $1e-5$). Genes with identities >70% were classified as homologs, while those below this threshold or without hits were treated as candidates for re-annotation. New genes were added only when they did not overlap with previously annotated ones, and the annotations were updated accordingly.

Clean reads of RNA-seq data were aligned to the genome using HISAT2 v2.2.1¹⁰⁸. Expression levels were quantified as fragments per kilobase of transcript per million mapped reads (FPKM) using StringTie2 v3.0.0 (default parameters). Co-expression networks were constructed using WGCNA¹⁰⁹ in R v4.2.3 with parameters (minModuleSize = 30, mergeCutHeight = 0.25, and power = 16). Modules showing >90% correlation at any developmental stage were considered target modules. Genes from target modules were compared across Pavi-T, Pmah, and Ppse to identify those with consistent expression in Pavi-T and Pmah but divergent patterns in Ppse.

SNPs, Indels, and SV calling of sweet cherry accessions

The SNPs and Indels of 219 cherry accessions were obtained from a previous study⁶. Short reads from these accessions were used to genotype variants in the pangenome with the VG v1.53.0⁹⁸. Only SVs ≥ 50 bp were retained for downstream analyses. To assess the reliability of SV detection, a subset of SVs was randomly selected for PCR validation (Supplementary Data 30).

Population structure analysis

To assess the impact of SV sets on population genetic analyses, we compared them with SNP and Indel data from 219 cherry accessions collected across diverse regions. To ensure representative variation and reduce computational load, SNPs, Indels, and SVs were filtered by minor allele frequency and missing rate. Variant annotation was performed using ANNOVAR v2015-12-14¹¹⁰. Shared allele distance matrices were calculated in PLINK v1.90b7¹¹¹. PCA was conducted using the PLINK PCA module. To infer phylogenetic relationships from genome-wide structural variations (SVs), filtered biallelic SV genotypes were converted into a binary presence/absence matrix and formatted into PHYLIP v3.698¹¹². Maximum-likelihood phylogeny was reconstructed using IQ-TREE v1.6.12⁶² under the MK + ASC model with 1000 ultrafast bootstrap replicates. Phylogenetic trees were visualized with the iTOL web server (<https://itol.embl.de>). For population structure analysis, SNPs in linkage disequilibrium (LD) were pruned in PLINK v1.90b7 using a 50-SNP sliding window (step size = 10 SNPs) with an r^2 threshold of 0.2. ADMIXTURE v1.3¹¹³ was then used to infer population structure, with K values set from 2 to 10, and cross-validation error estimates were used to determine the best fit.

Genomic selection signature identification

To detect a selective sweep, we performed a genome-wide scan of selection signals using the cross of F_{ST} and nucleotide diversity (π). A sliding window of 100 kb with 10 kb steps was applied, and F_{ST} and π values were calculated with VCFtools v0.1.16¹¹⁴. Genes located within the top 1% of regions showing significant selection signals were identified as candidate loci. Haplotype block regions were visualized using LDblockshow v1.4¹¹⁵.

GWAS analysis

To identify genetic loci associated with flowering time in cherries, GWAS was conducted using SNP, Indel, and SV genotypes from 219 accessions. Analyses were performed with both the linear mixed model (LM) implemented in GEMMA v0.98.5¹¹⁶ and the GLM in TASSEL v5.2.89¹¹⁷. GWAS results, including Manhattan and Q-Q plots, were visualized with the qqman package in R software v4.2.3.

Quantitative real-time reverse transcription PCR (qRT-PCR)

Real-time PCR was performed on a BIOER LineGene 9600 Plus Fluorescent Quantitative Detection System (FQD-96A; BIOER, Hangzhou, China) using the TB GreenTM Premix Ex TaqTM II kit (TaKaRa, Tokyo, Japan). Amplification was performed under the following conditions: an initial denaturation at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 30 s, and a final denaturation step at 95 °C for 10 s¹¹⁸. Relative expression of the target genes was calculated using the $2^{-\Delta\Delta CT}$ method¹¹⁹, with *PavActin* serving as the reference gene for normalization. Primer sequences used for qRT-PCR are listed in Supplementary Data 31.

Microstructure assessment of cherry flower buds

The phenotypes of cross sections of sweet cherry flower buds at ten developmental stages were examined using a stereomicroscope (SZ810, Cnoptec). Paraffin sections were prepared by fixing samples in FAA solution, followed by dehydration through an ethanol series, paraffin embedding, and sectioning into 8–10 μ m slices using a microtome¹²⁰. Slides were analyzed and imaged using a BX61 microscope (Olympus, Tokyo, Japan). GFP signals in transiently transformed flower buds were detected with a Leica fluorescence microscope equipped with a Leica DFC450 C camera (Leica Microsystems Inc.).

Subcellular localization

The CDS of *PavAGL9* without the stop codon was amplified using the primers listed in Supplementary Data 31. The CDS was cloned into the binary vector pHB, which carries two cauliflower mosaic virus 35S

promoters, a translation enhancer, and a GFP fluorescent protein tag. This generated a fusion construct designated p35S-*PavAGL9*-GFP. The fusion and control vectors (pHB) were introduced into *Agrobacterium tumefaciens* strain GV3101 and subsequently used to infiltrate the leaves of 3–5-week-old tobacco plants. Three days post-infiltration, when GFP fluorescence reached its peak, the localization of the fluorescent proteins was observed using a confocal laser scanning microscope (Leica TCS SP8 STED 3X; Wetzlar, Germany).

Stable plant transformation

The full-length CDS of *PavAGL9* was cloned into the binary pHB-GFP vector to generate an overexpression construct (*PavAGL9*-OE), which was subsequently transformed into *A. tumefaciens* strain GV3101. Transgenic *A. thaliana* plants were generated by immersing flowering plants in *Agrobacterium tumefaciens* GV3101 suspensions carrying the construct, followed by seed collection and selection of transformants¹²¹, targeting the Col-0 and *agl9* mutants. Plants were grown in a controlled growth chamber under a 16-h light/8-h dark photoperiod at 23 °C. After screening for herbicide resistance and confirming transgenic integration by DNA detection, transgenic lines with high *PavAGL9* expression were selected for phenotypic analyses.

Transient transformation

Transient overexpression was performed using an *Agrobacterium*-mediated transformation approach with minor modifications¹²². Two recombinant plasmids (*PavAGL9*-OE and *PavBPC6*-OE) for overexpression were introduced into the *A. tumefaciens* strain GV3101. An empty pHB-GFP vector served as a negative control. *Agrobacterium* cultures were grown to an optical density (OD₆₀₀) of approximately 0.6–0.8, collected by centrifugation, and resuspended in infiltration buffer. Subsequently, whole sweet cherry flower buds were immersed in the bacterial suspension and subjected to vacuum infiltration in a vacuum chamber connected to a vacuum pump. The infiltrated flower buds were then cultured under controlled conditions for subsequent analyses.

Selection of candidate transcription factors binding the *PavAGL9* promoter

Candidate transcription factors were predicted using PlantRegMap (https://plantregmap.gao-lab.org/regulation_prediction.php). Co-expression modules between *PavAGL9* and other genes were constructed using the WGCNA shiny plugin implemented in TBtools v2.056 software¹²³. Network analysis was performed using default parameters, except that the soft-thresholding power was set to 30 and the minimum module size to 30 genes. The clusters were merged based on eigengenes, and eigengenes values of each module were calculated to examine associations with *PavAGL9*. Candidate transcription factors were prioritized by integrating promoter binding predictions with co-expression network analysis.

Yeast one-hybrid assay

The full-length CDS of *PavBPC6* was cloned into the pB42AD vector at the *EcoRI* site. A 1900-bp promoter fragment of *PavAGL9* was amplified and inserted into the pLacZi vector at the *EcoRI* site. The reconstructed pLacZi and pB42AD vectors were co-transfected into yeast one-hybrid EGY48 cells. Protein–DNA interactions were assessed on a selective medium that lacked Ura and Trp, and supplemented with 100 mg/mL X-Gal. The primers used to construct the Y1H assay are listed in Supplementary Data 31.

Dual-LUC assay

The promoter fragment of *PavAGL9* was cloned into the pGreenII 0800-LUC vector using *Bam*HI and *Pst*I sites to generate the reporter construct pro*PavAGL9*: LUC. The full-length CDS of *PavBPC6* was inserted into the pPZP6K90 vector at the *Xba*I endonuclease site. Both

constructs were introduced into *A. tumefaciens* strain GV3101. *Agrobacterium* cultures carrying the effector (35S:*PavBPC6*-pPZP6K90) and the reporter construct pro*PavAGL9*: LUC were co-infiltrated into *N. benthamiana* leaves, with pPZP6K90 and pro*PavAGL9*: LUC serving as controls. After 2–3 days, luciferase reporter assays were performed using commercial reagents (YEASEN, Shanghai, China). Luciferase signals were visualized and quantified using a Tanon-5200 imaging system (Tanon, Shanghai, China). Primer sequences used for this assay are listed in Supplementary Data 31.

Electrophoretic mobility shift assay (EMSA)

The full-length CDS of *PavBPC6* was cloned into the pGEX4T-2 vector in-frame with a GST tag using the *Bam*HI and *Eco*RI sites. The constructs were transformed into Rosetta (DE3) competent cells for recombinant protein expression. Potential *PavBPC6*-binding sites within the *PavAGL9* promoter were predicted using PlantRegMap and FIMO (MEME Suite, <https://meme-suite.org/meme/tools/fimo>). Biotin-labeled double-stranded DNA probes containing the predicted motifs were synthesized for EMSA analysis. Recombinant proteins were purified using glutathione (GSH) beads, and EMSA was performed with the LightShift Chemiluminescent EMSA Kit (Thermo Fisher Scientific, Waltham, MA, USA). Protein purity was verified by SDS–PAGE. Gels were imaged using a Tanon 5200 Chemiluminescent Imaging System and scanned using a ChemiDoc™ XRS⁺ system (Bio-Rad, Hercules, CA, USA). All probes were synthesized by Sangon Biotech Co., Ltd. (Shanghai, China). Primer sequences used in these assays are provided in Supplementary Data 31.

Yeast two-hybrid assay

Full-length CDS of *PavPMADS2* and *PavSEPI* were cloned into the pGBKT7 vector using *Nde*I and *Eco*RI sites, whereas *PavAGL9* was cloned into the pGADT7 vector using *Bam*HI sites. All reconstructed vectors were confirmed by Sanger sequencing. The reconstructed and empty vectors were then introduced into yeast Y2HGold cells. Protein interactions were screened on a selective medium lacking Trp, His, Leu, and Ade, and supplemented with 20 mM X- α -Gal. AD-T and BD-lam were co-transferred as negative controls, while AD-T and BD-53 served as positive controls. Primer sequences used for the Y2H assay are listed in Supplementary Data 31 and were synthesized by Sangon Biotech Co., Ltd. (Shanghai, China).

Bimolecular fluorescence complementation (BiFC)

Full-length CDS of *PavPMADS2* and *PavSEPI* were inserted into the pXY106 vector (using *Bam*HI and *Sal*I sites, respectively), whereas *PavAGL9* was cloned into the pXY104 vector using the same restriction sites. The resulting constructs were introduced into *A. tumefaciens* strain GV3101. *Agrobacterium* cultures were grown to an OD₆₀₀ of 0.6–0.8 and then resuspended in infiltration buffer. The suspensions were co-infiltrated into leaves of *N. benthamiana* using a needleless syringe. Fluorescence signals were observed 2–3 days post-infiltration using a confocal laser scanning microscope (Leica TCS SP8 STED 3X; Wetzlar, Germany). Primer sequences used in this assay are listed in Supplementary Data 31.

Co-immunoprecipitation (Co-IP)

The constructs *PavAGL9*-YFP, *PavPMADS2*-Flag, and *PavSEPI*-Flag were generated by inserting the full-length coding sequences of *PavAGL9*, *PavPMADS2*, and *PavSEPI*, respectively, into the pHB-X-YFP and pHB-X-Flag vectors using *Bam*HI and *Spe*I sites. The recombinant plasmids were transformed into *A. tumefaciens* strain GV3101. For transient expression, *N. benthamiana* leaves were infiltrated with *Agrobacterium* suspensions harboring the indicated constructs and incubated for 2–3 days. Total proteins were extracted from infiltrated leaf tissues using protein extraction buffer containing protease inhibitors, and the supernatant was obtained by centrifugation. For immunoprecipitation,

protein extracts were incubated with anti-GFP protein G magnetic beads (Thermo Fisher Scientific, Waltham, MA, USA) at 4 °C for 2–4 h with gentle rotation. After extensive washing, the bound proteins were eluted and analyzed by immunoblotting using anti-GFP and anti-Flag antibodies, respectively, to detect protein–protein interactions. All antibodies used in this study are purchased from Beijing TransGen Biotech Co., Ltd. (Beijing, China). Primer sequences used for plasmid construction are listed in Supplementary Data 3L.

Statistical analysis

Statistical analysis for multiple group comparisons was performed using one-way analysis of variance (ANOVA) by two-sided Tukey's post-hoc test. For comparisons between two groups, two-sided Student's *t*-test was applied. All data are presented as mean ± standard deviation (SD), and statistical significance was defined as $P < 0.05$. Bar charts and box plots were generated using GraphPad Prism v10.5.0.

Reporting summary

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

Data availability

The raw genome sequencing and RNA-Seq data have been deposited in the National Genomics Data Center under BioProject [PRJCA045945](https://www.ncbi.nlm.nih.gov/bioproject/PRJCA045945). Genome annotation files generated in this study are publicly available at Zenodo [<https://doi.org/10.5281/zenodo.19640357>]²⁴. Source data are provided with this paper.

Code availability

All custom scripts used in this study are available on GitHub [<https://github.com/hanlu211/Cerasus-pangenome-v1>] and Zenodo [<https://doi.org/10.5281/zenodo.19571818>]²⁵.

References

- Chin, S.-W., Shaw, J., Haberle, R., Wen, J. & Potter, D. Diversification of almonds, peaches, plums and cherries—molecular systematics and biogeographic history of *Prunus* (Rosaceae). *Mol. Phylogenet. Evol.* **76**, 34–48 (2014).
- Zhang, J. et al. Evolution of Rosaceae plastomes highlights unique *Cerasus* diversification and independent origins of fruiting cherry. *Front. Plant Sci.* **12**, 736053 (2021).
- Wang, T. et al. Pan-genome analysis of 13 *Malus* accessions reveals structural and sequence variations associated with fruit traits. *Nat. Commun.* **14**, 7377 (2023).
- Yü, D. J., Lu, L. T., Ku, T. C., Li, C. L. & Chen, S. X. *Flora of China* (Science Press, 1986).
- Bi, W. *Lü Shi Chun Qiu* (Shanghai Classics Publishing House, 1995) (in Chinese).
- Lei, Y. et al. Population sequencing of cherry accessions unravels the evolution of *Cerasus* species and the selection of genetic characteristics in edible cherries. *Mol. Hortic.* **5**, 25 (2025).
- Webster, A. D. & Looney, N. E. *Cherries: Crop Physiology, Production and Uses* (Cambridge University Press, 1996).
- Chen, T. et al. The complete chloroplast genome of Tomentosa cherry *Prunus tomentosa* (Prunoideae, Rosaceae). *Mitochondrial DNA Part B* **3**, 672–673 (2018).
- Song, Y.-F. et al. Molecular phylogenetics and biogeography reveal the origin of cherries (*Prunus* subg. *Cerasus*, Rosaceae). *Bot. J. Linn. Soc.* **204**, 304–315 (2024).
- Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**, 840–852 (2013).
- Olsen, K. M. & Wendel, J. F. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* **64**, 47–70 (2013).
- De Coster, W. & Van Broeckhoven, C. Newest methods for detecting structural variations. *Trends Biotechnol.* **37**, 973–982 (2019).
- Lye, Z. N. & Purugganan, M. D. Copy number variation in domestication. *Trends Plant Sci.* **24**, 352–365 (2019).
- Ma, Z. et al. High-quality genome assembly and resequencing of modern cotton cultivars provide resources for crop improvement. *Nat. Genet.* **53**, 1385–1391 (2021).
- Alonge, M. et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**, 145–161.e23 (2020).
- Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
- Ho, S. S., Urban, A. E. & Mills, R. E. Structural variation in the sequencing era. *Nat. Rev. Genet.* **21**, 171–189 (2020).
- Li, X. et al. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in Brassica oleracea. *Nat. Genet.* **56**, 517–529 (2024).
- Zhang, C. et al. High-quality genome of a modern soybean cultivar and resequencing of 547 accessions provide insights into the role of structural variation. *Nat. Genet.* **56**, 2247–2258 (2024).
- Nie, C. et al. Genome assembly, resequencing and genome-wide association analyses provide novel insights into the origin, evolution and flower colour variations of flowering cherry. *Plant J.* **114**, 519–533 (2023).
- Jiu, S. et al. Chromosome-scale genome assembly of *Prunus pusilliflora* provides novel insights into genome evolution, disease resistance, and dormancy release in *Cerasus* L. *Hortic. Res.* **10**, uhad062 (2023).
- Liu, Y. et al. Pan-genome of wild and cultivated soybeans. *Cell* **182**, 162–176.e13 (2020).
- Jiao, C. et al. Pan-genome bridges wheat structural variations with habitat and breeding. *Nature* **637**, 384–393 (2025).
- Feng, J.-W. et al. A haplotype-resolved pangenome of the barley wild relative *Hordeum bulbosum*. *Nature* **645**, 429–438 (2025).
- Li, N. et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* **55**, 852–860 (2023).
- Qiao, Q. et al. Evolutionary history and pan-genome dynamics of strawberry (*Fragaria* spp.). *Proc. Natl. Acad. Sci. USA* **118**, e2105431118 (2021).
- Zhang, Y. et al. Telomere-to-telomere *Citrullus* super-pangenome provides direction for watermelon breeding. *Nat. Genet.* **56**, 1750–1761 (2024).
- Li, W. et al. Pan-genome analysis reveals the evolution and diversity of *Malus*. *Nat. Genet.* **57**, 1274–1286 (2025).
- Huang, Y. et al. Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in citrus fruits. *Nat. Genet.* **55**, 1964–1975 (2023).
- Liu, Z. et al. Grapevine pangenome facilitates trait genetics and genomic breeding. *Nat. Genet.* **56**, 2804–2814 (2024).
- Benoit, M. et al. *Solanum* pan-genetics reveals paralogues as contingencies in crop engineering. *Nature* **640**, 135–145 (2025).
- Goeckeritz, C. Z. et al. Genome of tetraploid sour cherry (*Prunus cerasus* L.) ‘Montmorency’ identifies three distinct ancestral *Prunus* genomes. *Hortic. Res.* **10**, uhad097 (2023).
- Jiu, S. et al. Haplotype-resolved genome assembly for tetraploid Chinese cherry (*Prunus pseudocerasus*) offers insights into fruit firmness. *Hortic. Res.* **11**, uhae142 (2024).
- Jiu, S. et al. Chromosome-level genome assembly provides insights into the genetic diversity, evolution, and flower development of *Prunus conradinae*. *Mol. Hortic.* **4**, 25 (2024).

35. Wöhner, T. W. et al. The draft chromosome-level genome assembly of tetraploid ground cherry (*Prunus fruticosa* Pall.) from long reads. *Genomics* **113**, 4173–4183 (2021).
36. Zhao, Y. et al. Chromosome-scale *Cerasus humilis* genome assembly reveals gene family evolution and possible genomic basis of calcium accumulation in fruits. *Sci. Hortic.* **299**, 111012 (2022).
37. Shirasawa, K. et al. Phased genome sequence of an interspecific hybrid flowering cherry, ‘Somei-Yoshino’ (*Cerasus* × *yedoensis*). *DNA Res.* **26**, 379–389 (2019).
38. Shirasawa, K., Itai, A. & Isobe, S. Genome sequencing and analysis of two early-flowering cherry (*Cerasus* × *kanzakura*) varieties, ‘Kawazu-zakura’ and ‘Atami-zakura’. *DNA Res.* **28**, dsab026 (2021).
39. Yi, X.-G. et al. The genome of Chinese flowering cherry (*Cerasus serrulata*) provides new insights into *Cerasus* species. *Hortic. Res.* **7**, 165 (2020).
40. Hufford, M. B. et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
41. Bozan, I. et al. Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. *Proc. Natl. Acad. Sci. USA* **120**, e221117120 (2023).
42. Liu, D. et al. PlantCFG: a comprehensive database with web tools for analyzing candidate flowering genes in multiple plants. *Plant Commun.* **5**, 100733 (2024).
43. Gallé, Á et al. Plant glutathione transferases and light. *Front. Plant Sci.* **9**, 1944 (2019).
44. Schippers, J. H., Foyer, C. H. & Van Dongen, J. T. Redox regulation in shoot growth, SAM maintenance and flowering. *Curr. Opin. Plant Biol.* **29**, 121–128 (2016).
45. Shi, H. et al. Genome-wide identification and expression analysis of the glutathione transferase gene family and its response to abiotic stress in rye (*Secale cereale*). *BMC Genom.* **25**, 1142 (2024).
46. Cheng, Z. et al. Nup96 and HOS1 are mutually stabilized and gate CONSTANS protein level, conferring long-day photoperiodic flowering regulation in *Arabidopsis*. *Plant Cell* **32**, 374–391 (2020).
47. Huang, P. et al. The nuclear pore Y-complex functions as a platform for transcriptional regulation of *FLOWERING LOCUS C* in *Arabidopsis*. *Plant Cell* **36**, 346–366 (2024).
48. Gascioli, V., Mallory, A. C., Bartel, D. P. & Vaucheret, H. Partially redundant functions of *Arabidopsis* dicer-like enzymes and a role for DCL4 in producing *trans*-acting siRNAs. *Curr. Biol.* **15**, 1494–1500 (2005).
49. Liu, F., Marquardt, S., Lister, C., Swiezewski, S. & Dean, C. Targeted 3' processing of antisense transcripts triggers *Arabidopsis FLC* chromatin silencing. *Science* **327**, 94–97 (2010).
50. Park, J., Nguyen, K. T., Park, E., Jeon, J.-S. & Choi, G. DELLA proteins and their interacting RING finger proteins repress gibberellin responses by binding to the promoters of a subset of gibberellin-responsive genes in *Arabidopsis*. *Plant Cell* **25**, 927–943 (2013).
51. Jang, S. et al. *Arabidopsis* COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. *EMBO J.* **27**, 1277–1288 (2008).
52. Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S. & Yanofsky, M. F. Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell* **11**, 1007–1018 (1999).
53. Shen, X. et al. Evolution of cherries (*Prunus* Subgenus *Cerasus*) based on chloroplast genomes. *Int. J. Mol. Sci.* **24**, 15612 (2023).
54. Liu, Z., Bernard, A., Wang, Y., Dirlwanger, E. & Wang, X. Genomes and integrative genomic insights into the genetic architecture of main agronomic traits in the edible cherries. *Hortic. Res.* **12**, uhae269 (2025).
55. Branchereau, C. et al. New insights into flowering date in *Prunus*: fine mapping of a major QTL in sweet cherry. *Hortic. Res.* **9**, uhac042 (2022).
56. Sooriyapathirana, S. S. et al. QTL analysis and candidate gene mapping for skin and flesh color in sweet cherry fruit (*Prunus avium* L.). *Tree Genet. Genomes* **6**, 821–832 (2010).
57. Cachi, A. M. & Wünsch, A. Characterization and mapping of non-S gametophytic self-compatibility in sweet cherry (*Prunus avium* L.). *J. Exp. Bot.* **62**, 1847–1856 (2011).
58. Castède, S. et al. Mapping of candidate genes involved in bud dormancy and flowering time in sweet cherry (*Prunus avium*). *PLoS ONE* **10**, e0143250 (2015).
59. Belton, J.-M. et al. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
60. Jin, J.-J. et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).
61. Katoh, K. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).
62. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
63. Wang, Q., Liu, X. & Xu, L. *Illustrated Guide to Major Cultivated Flowering Cherry Varieties in China* (Zhejiang Science and Technology Publishing House, 2015).
64. Wenden, B., Barreneche, T., Meland, M. & Blanke, M. M. Harmonisation of phenology stages and selected cherry cultivars as bioindicators for climate change. *Acta Hortic.* **9**, 12 (2017).
65. Doležel, J. Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* **95**, 99–110 (2005).
66. Doležel, J., Greilhuber, J. & Suda, J. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**, 2233–2244 (2007).
67. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
68. Chen, Y. et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* **12**, 60 (2021).
69. Guan, D. et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
70. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
71. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
72. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
73. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
74. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
75. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
76. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
77. Lin, Y. et al. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Hortic. Res.* **10**, uhad127 (2023).
78. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).

79. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
80. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).
81. Korf, I. Gene finding in novel genomes. *BMC Bioinform.* **5**, 59 (2004).
82. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
83. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **6**, 31 (2005).
84. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
85. Haas, B. J. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
86. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
87. Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
88. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
89. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
90. Lagesen, K. et al. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
91. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
92. Griffiths-Jones, S. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2004).
93. Yu, G. C., Wang, L. G., Han, Y. Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
94. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).
95. 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).
96. Nattestad, M. & Schatz, M. C. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32**, 3021–3023 (2016).
97. Jeffares, D. C. et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **8**, 14061 (2017).
98. Garrison, E. et al. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875–879 (2018).
99. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9**, 304 (2020).
100. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
101. Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* **13**, 6968 (2022).
102. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
103. Darriba, D. et al. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).
104. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
105. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13**, 555–556 (1997).
106. Li, Y. et al. Endocarps of *Prunus* (Rosaceae: Prunoideae) from the early Eocene of Wutu, Shandong Province, China. *Taxon* **60**, 555–564 (2011).
107. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
108. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
109. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
110. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
111. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015-0047-8 (2015).
112. Felsenstein, J. PHYLIP: Phylogeny Inference Package. Version 3.2. *Q. Rev. Biol.* **64**, 539–541 (1989).
113. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**, 246 (2011).
114. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
115. Dong, S.-S. et al. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* **22**, bbaa227 (2021).
116. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
117. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
118. Jiu, S. et al. The role of *VvMYBA2r* and *VvMYBA2w* alleles of the *MYBA2* locus in the regulation of anthocyanin biosynthesis for molecular breeding of grape (*Vitis* spp.) skin coloration. *Plant Biotechnol. J.* **19**, 1216–1239 (2021).
119. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
120. Nii, N., Kawahara, T. & Nakao, Y. The development of stone cells in Japanese pear fruit. *J. Hortic. Sci. Biotechnol.* **83**, 148–153 (2008).
121. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
122. Zhai, Z. et al. Abscisic acid-responsive transcription factors PavDof2/6/15 mediate fruit softening in sweet cherry. *Plant Physiol.* **190**, 2501–2518 (2022).
123. Chen, C. et al. TBtools: an integrative toolkit developed for interactive analyses of big Biological Data. *Mol. Plant* **13**, 1194–1202 (2020).
124. Jiu, S., & Lei, Y. The pan-genome annotations for *Cerasus*. Zenodo, <https://doi.org/10.5281/zenodo.19640357> (2026).
125. Jiu, S., & Lei, Y. Codes for *Cerasus* pangenome v1. Zenodo, <https://doi.org/10.5281/zenodo.19571818> (2026).

Acknowledgements

The authors thank Dr. Wei Wu from the School of Agriculture and Biology, Shanghai Jiao Tong University, for providing valuable discussions

and suggestions. The authors also thank Prof. Junwei Huo from Northeast Agricultural University, China, and Prof. Wenjuan Geng from Xinjiang Agricultural University, China, for providing plant materials. This study was funded by the National Natural Science Foundation of China (Grant No. 32541114 to S.J. and 32260734 to F.Y.), the China Agriculture Research System (Grant No. CARS-30 to C.Z.), the Natural Science Foundation of Shanghai (Grant No. 23ZR1430600 to S.J.), the project of Qianjiang Distinguished Expert of Hangzhou with Hangzhou Academy of Agricultural Science (C.Z.), and the Key Research and Development Program of Shaanxi Province, China (Grant No. 2023-ZDLNY-28 to C.Z.).

Author contributions

C.X.Z., S.J., S.D., and Z.S. conceived and designed the experiments; S.J. drafted the manuscript; S.J., Y.L., L.F., Y.H., C.Z., and Y.X. performed sequencing data analysis; S.J., L.F., L.H., and Y.X. performed experiments; S.J. and Y.L. performed the assembly and annotations; S.J., Lei Wang, J.W., X.Z., Y.C., W.Z., F.L., C.L., H.L., F.Y., M.L., J.W.W., L.C., C.Y., L.P., J.Y.W., Y.W., Y.Z., D.E., A.B., and Z.Z. collected the samples; S.J., Z.L., X.L., J.Z., X.A., L.Y., T.Z., Li Wang, and X.L. worked on the phenotyping; S.J. and Y.L. performed the statistical analysis; W.C., X.D., R.L., M.N., D.L., Q.Z., H.X., Y.D., B.L., and S.W. participated in discussions and provided valuable advice. All authors have approved the final manuscript for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-72832-8>.

Correspondence and requests for materials should be addressed to Songtao Jiu, Zongyi Sun, Shengchang Duan or Caixi Zhang.

Peer review information *Nature Communications* thanks Meng Li, Daniel Potter, Steve van Nocker, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

¹Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, PR China. ²Key Laboratory of National Forestry and Grassland Administration on Cherry, Shanghai, PR China. ³College of Science, Yunnan Agricultural University, Kunming, PR China. ⁴College of Horticulture, Anhui Agricultural University, Hefei, PR China. ⁵College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, PR China. ⁶GrandOmics Biosciences, Beijing, PR China. ⁷Shandong Institute of Pomology, Taian, PR China. ⁸Yantai Academy of Agricultural Sciences, Yantai, PR China. ⁹College of Horticulture, Northwest A&F University, Yangling, PR China. ¹⁰Dalian Academy of Agricultural Sciences, Dalian, PR China. ¹¹Zhengzhou Fruit Tree Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, PR China. ¹²Horticulture Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu, PR China. ¹³Shanghai Botanical Garden, Shanghai, PR China. ¹⁴Horticultural Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, PR China. ¹⁵Forestry and Fruit Research Institute, Beijing Academy of Agriculture and Forestry Sciences, Beijing, PR China. ¹⁶Changli Institute of Pomology, Hebei Academy of Agriculture and Forestry Sciences, Qinhuangdao, PR China. ¹⁷Dali Bai Autonomous Prefecture Academy of Agricultural Sciences and Extension, Dali, PR China. ¹⁸College of Landscape and Horticulture, Yunnan Agricultural University, Kunming, PR China. ¹⁹Department of Horticulture, School of Bioengineering and Biotechnology, Tianshui Normal University, Tianshui, PR China. ²⁰Shanghai Chenshan Botanical Garden, Shanghai, PR China. ²¹INRAE, Univ. Bordeaux, UMR BFP, Bordeaux, France. ²²The United Graduate School of Agricultural Science, Gifu University, Gifu, Japan. ²³The School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, PR China. ²⁴These authors contributed equally: Songtao Jiu, Yahui Lei. ✉ e-mail: jiusongtao@sjtu.edu.cn; sunzongyi@163.com; duanshengchang@163.com; acaizh@sjtu.edu.cn