Genome assembly, resequencing and genome-wide association analyses provide novel insights into the origin, evolution and flower colour variations of flowering cherry

Article *in* The Plant Journal · February 2023

DOI: 10.1111/tpj.16151

10 authors, including:

Chaoren Nie
Beijing Forestry University
5 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Nian Wang
Huazhong Agricultural University
119 PUBLICATIONS **2,164** CITATIONS
SEE PROFILE

Nie Chaoren (Orcid ID: 0000-0002-4631-3841)
Wang Nian (Orcid ID: 0000-0001-9185-4199)

Genome assembly, resequencing and genome-wide association analyses provide novel insights into the origin, evolution and flower colour variations of flowering cherry

Chaoren Nie[1, 2], Yingjie Zhang[4], Xiaoqin Zhang[2], Wensheng Xia[2], Hongbing Sun[2], Sisi Zhang[2], Na Li[2],Zhaoquan Ding[2], Yingmin Lv[1*], Nian Wang[3*]

[1] School of Landscape architecture, Beijing Forestry of University, Beijing, 100083, China; [2] Wuhan Institute of Landscape Architecture, Wuhan, 430081, China; [3] College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, 430070, China; [4] Yantai Academy of agricultural Sciences, Shandong province, Yantai, 265500, China.

[*] Corresponding author

Email Address:

Chaoren Nie: 281560660@qq.com

Yingjie Zhang: 447477668@qq.com

Xiaoqin Zhang: 1044239123@qq.com

Wensheng Xia: 1870978491@qq.com

Hongbing Sun: 420384652@qq.com

Sisi Zhang: 597002017@qq.com

Na Li: 543633019@qq.com

Zhaoquan Ding: 522539700@qq.com

Yingmin LV: yinminlv@bjfu.edu.cn

Nian Wang: wangn@mail.hzau.edu.cn

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.16151

Abstract

Flowering cherry is a very popular species around the world. High-quality genome resources for different elite cultivars are needed, and the understanding of their origins and the regulation of key ornamental traits are limited for this tree. Here, a high-quality chromosome-scale genome of *Prunus campanulata* 'Plena' (PCP), which is a native and elite flowering cherry cultivar in China, was generated. The contig N50 of the genome was 18.31 Mb, and 99.98% of its contigs were anchored to eight chromosomes. Furthermore, a total of 306 accessions of flowering cherry germplasm and 6 lines of outgroups were collected. Resequencing of these 312 lines was performed, and 761267 high-quality genomic variants were obtained. The origins of flowering cherry were predicted, and these 306 accessions could be classified into three clades, A, B and C. According to phylogenetic analysis, we predicted two origins of flowering cherry. Flowering cherry in clade A originated in southern China, such as in the Himalayan Mountains, while clades B and C originated in northeastern China. Finally, a genome-wide association study (GWAS) of flower colour was performed for all 312 accessions of flowering cherry germplasm. A total of seven quantitative trait loci (QTLs) were identified. One gene encoding glycosylate transferase was predicted as the candidate gene for one QTL. Taken together, our results provide a valuable genomic resource and novel insights into the origin, evolution and flower colour variations of flowering cherry.

Keywords:

Flowering cherry; Subgenus *Cerasus;* High-quality genome; Origins of flowering cherry; Genome-wide association study (GWAS); Flower colour variations

Introduction

Subgenus *Cerasus* belongs to the genus *Prunus* of the Rosaceae family (Chin *et al.*, 2014). This subgenus has approximately 100 species and is widely cultivated as fruit crops, ornamental plants, and medicinal and industrial materials because of its diverse uses (Chin *et al.*, 2014, Wang *et al.*, 2022). It is mainly distributed in temperate and subtropical regions of the Northern Hemisphere. Most species in this subgenus are distributed in eastern Asia, including China, Japan and Korea. Flowering cherry, which is one of the most famous species in the *subgenus Cerasus*, is a popular ornamental woody plant owing to its beautiful flowers, attractive colour, and blossoms in spring. It is widely cultivated worldwide, especially in Japan, China and other countries (Kato *et al.*, 2014).

There were historical records of cherry cultivation in China as early as 2,000 years ago (Kato *et al.*, 2012). In the Tang and Song dynasties, flowering cherries were widely planted in palace gardens, appearing in a number of poems for their beautiful flowers (Zhao *et al.*, 2016). The cultivation of flowering cherries in Japan can be traced back to 1,000 years ago, and flowering cherry is regarded as the national flower (Niwa, 1936, Wybe, 1999). At present, approximately 300 flowering cherry cultivars have been developed worldwide (Hideaki Ohba, 2007). Although flowering cherry has been cultivated for a long time, the understanding of its origin and evolution is limited. Some studies based on morphological observations have indicated that cultivars in Japan originated from native taxa or hybrids, while the other cultivars were believed to have close relations with the wild lines in China (Hideaki Ohba, 2007, Kato *et al.*, 2014, Zhao *et al.*, 2016). Later, this was partially confirmed by molecular marker assays (Katori *et al.*, 2002, Ogawa *et al.*, 2012, Kato *et al.*, 2014). However, there is still no clear evidence for the origin and evolution of flowering cherries.

Flower colour is one of the most important traits for ornamental plants. The development of bright flower colours is an effective strategy for plants to attract

pollinators, and these different colours also make them useful as ornamental plants (Zhao and Tao, 2015, Narbona *et al.*, 2021). Flower colour is mainly affected by petal pigment components, epidermal cell structure, the pH value of cell vacuoles, complexation of metal ions and so on (Zhao and Tao, 2015). Plant pigments mainly include flavonoids, carotenoids and alkaloids. Anthocyanins are one type of flavonoid compound that also belongs to the group of plant secondary metabolites (Mattioli *et al.*, 2020). Seven anthocyanin glycosides have been identified, including pelargonidin, cyanidin, delphinidin, peonidin, malvidin, petunidin and hirsutidin (Mattioli *et al.*, 2020). The biosynthesis of anthocyanins has been comprehensively investigated, and a number of transcription factors (TFs) and key genes encoding enzymes have been identified in plants (Hichri *et al.*, 2011). For flowering cherry, different flower colours can be observed in different lines. Generally, different lines/cultivars of flowering cherry can produce flowers with white, red and pink colours. In recent decades, some studies have focused on describing the biosynthesis of pigments in flowering cherry. For example, different pathways and metabolites for flower colour among different *Prunus serrulata* cultivars were analysed (Yang, 2006). However, the mechanisms underlying different flower colours in flowering cherry remain elusive.

With the development of sequence technology, especially the appearance of third-generation DNA sequencing technology, it has become possible to obtain a high-quality chromosome-level genome for nonmodel plants. Recently, chromosome-scale genomic sequences of many ornamental plants have been released. The genomes of many fruit trees that belong to the *Prunus* genus have also been released, such as apple (Velasco *et al.*, 2010, Li *et al.*, 2016), pear (Wu *et al.*, 2013), peach (International Peach Genome *et al.*, 2013), sweet cherry (Wang *et al.*, 2020) and apricot (Jiang *et al.*, 2019). Genomic resequencing of a large panel of germplasms was also conducted for a number of plants. With these high-quality genomic resources, a number of biological questions have been addressed. For example, the molecular mechanism of breaking dormancy and flowering under low temperature in early spring was revealed by using genomic information in *Prunus mume* (Zhang *et al.*, 2012). Genome-wide resequencing of sweet cherry (*Prunus avium*) also revealed a

modifier gene mutation conferring pollen-part self-compatibility (Ono *et al.*, 2018). For flowering cherry, the draft genome sequence of wild *P. yedoensis* was assembled by using long-read sequencing and sequence phasing (Baek *et al.*, 2018). Cross-species hybridization was observed between *P. yedoensis* and its related taxa by comparative genomic analysis. The phased genome sequence of an interspecific hybrid flowering cherry, 'Somei-Yoshino' (*Cerasus* × *yedoensis*), was also generated (Shirasawa *et al.*, 2019). The genome of Chinese flowering cherry (*Cerasus serrulata*) was also generated by combining Nanopore and Hi-C sequencing technologies (Yi *et al.*, 2020). This genomic resource largely facilitated our understanding of the evolution, origin and genomic selections of flowering cherry. However, considering the complex hybridization among different species/subspecies in the *Cerasus* subgenus and controversial views of the origin of this ornamental tree, more high-quality chromosome-scale genomes for different lines/subspecies in the *Cerasus* subgenus are still needed. Additionally, a genome-wide investigation of the genetic basis of key ornamental traits, such as flower colour, petal patterns and plant architecture, is also necessary.

Here, we provide a high-quality chromosome-scale genome assembly of *P. campanulata* 'Plena' (PCP), which is a flowering cherry native to the southeastern and Taiwanese regions of China. Furthermore, we resequenced a large panel of flowering cherry germplasms that included 312 flowering cherry accessions (160 accessions of cultivars, 77 F1 hybrids and 75 wild individual lines). The population structure and genetic relationships of these 312 accessions were analysed. The origin, evolution, and migration of flowering cherry were also predicted according to the large dataset. Finally, a genome-wide association study (GWAS) of flower colour was performed, and some candidate genes for the regulation of flower colour in flowering cherry were predicted. With this study, we were able to provide a valuable genomic resource and novel insights into the origin, evolution and flower colour variations of flowering cherry.

Results

A high-quality chromosome-scale genome sequence for PCP

The PCP line with red flowers was used as the plant material in this study (Fig. 1a). Its native habitat is in mountains with altitudes between 500 and 2000 metres in the northern and central regions of Taiwan, China. The genome size of PCP was estimated to be 293.85 Mb by K-mer analysis, and its heterozygosity rate was 0.6% (Fig. S1). After filtering low-quality reads and adaptor sequences, a total of 76.42 Gb of Oxford Nanopore long reads with N50 and average lengths of 30.32 and 21.55 Kb, respectively, were obtained. Meanwhile, a total of 29.72 Gb of Illumina short paired-end (PE) reads were also generated (Table S1). The long reads were used for genome assembly by using Next Denovo software (https://github.com/Nextomics/NextDenovo). The preliminary genome size and N50 length were 278.78 Mb and 18.20 Mb, respectively. After genome correction with Nextpolish, the final PCP genome assembly was 280.20 Mb, consisting of 41 contigs, with an N50 size of 18.31 Mb. The statistics for the PCP genome assembly are provided in Table 1 and Table S2. The PCP genome showed much greater contig N50 values than the other two genomes in the subgenus *Cerasus* (Table S2).

Hi-C sequencing generated 42.87 Gb of clean reads. After mapping the Hi-C reads onto the assembly of the PCP contigs, 65653085 valid reads, accounting for 79.94% of the unique mapped read pairs, were used for further Hi-C analysis (Table S1). A total of 279.39 Mb of contigs were clustered into 8 pseudochromosomes (hereafter chromosomes), and the result showed that the rate of 99.98% was higher than that of *Cerasus serrulata* (99.16%) (Yi *et al.*, 2020). According to the Hi-C interaction heatmap for the chromosome-scale genome of PCP (Fig. 1b), we can see a high intensity of interaction within each of the 8 pseudochromosomes, and this result suggests that the genome assembly is of high quality.

The PCP genome was also assessed with 3 different strategies. First, Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis showed that the

complete BUSCO rate of the assembled genome was 98.70% (Fig. 1c). This result is higher than that of *C. serrulata,* at 94.67%. Second, the short PE reads were also mapped onto the PCP genome assembly, and the average mapping rate and coverage were 98.5% and 100×, respectively. Third, the nanopore long reads mapped back onto the PCP genome suggested that the average depth was 257.03×. A total of 99.99% of the PCP genome could be covered by at least one long read (Table S3). Taken together, the above findings suggest that we have obtained a high-quality chromosome-scale genome for PCP and that this genome can be used in further studies.

Genome annotation

The PCP genome included 8 chromosomes ranging in size from 51.55 Mb to 26.67 Mb (Table S4). According to the size of these 8 chromosomes, they were assigned the names Chr01 to Chr08 (from the longest to the shortest). The GC content for the PCP genome was 39.48%, which is highly consistent with the K-mer 17 survey (Table S2), while that of the other released chromosome-scale genome of flowering cherry was 38.51% (*C. serrulata*) (Table S2). Therefore, these two genomes in the subgenus *Cerasus* showed very similar GC contents. The repeat sequences of the PCP genome were also identified, and a total of 140.63 Mb of sequences accounting for 50.33% of the PCP genome were annotated as repeat elements. The proportion of repeat sequences was also very similar to that of *C. serrulata* (48.99%). Among these repetitive elements, LTR retrotransposons (32.61%) accounted for the largest proportion, followed by Gypsy (9.47%) and Copia (4.45%) (Table S5). In the PCP genome, the total LTR retrotransposon size was 91.13 Mb, accounting for 32.61% of its genome. This was higher than in the other 3 released genomes (*P. avium* 'tieton' 19.71%, *C. serrulata* 23.81% and *P. yedoensis* 22.75%) in the subgenus *Cerasus* (Table S2), indicating that LTR retrotransposons vary greatly among different species in the same subgenus.

Three strategies, including ab initio prediction, homology searching and RNA-Seq guidance, were applied for gene prediction in the PCP genome. A total of 27181 protein-coding genes were predicted, and it showed a similar number of

protein-coding genes with *C. serrulata* (27181 *vs*. 26820) (Table S6). In gene function annotation with the NR, GO, KEGG, KOG and Swiss-Prot databases, 26820 genes were annotated in at least one database, accounting for 98.67% of all predicted genes (Table S7). Additionally, a total of 2191 noncoding RNAs, including 2 cis-regulatory elements and 758 ribosomal, 699 transfer and 732 small RNAs, were annotated in the PCP genome (Table S8). The distributions of gene density, repeat sequences, Gypsy and Copia transposable elements and GC contents among different chromosomes are illustrated in Fig. 1d.

Genome comparison and evolutionary analyses

High-quality chromosome-scale genomes are available for two other species/lines, *C. serrulata* and *P. yedoensis var. Nudiflora*; thus, collinearity analyses between the PCP genome and these two genomes were performed (Baek *et al.*, 2018, Yi *et al.*, 2020). In total, there were 1507 collinear blocks with a similarity >95% and size above 15 kb between the PCP and *Cerasus serrulata* genomes (Table S9). These collinear blocks only accounted for 11.5% (33.5 Mb) of the 280.2 Mb PCP genome. Similarly, there were 2283 collinear blocks with a similarity >95% and size above 15 kb between PCP and *P. yedoensis var. nudiflora* (Table S10). These collinear blocks only accounted for 17.3% (48.5 Mb) of the 280.2 Mb PCP genome. The genome of *P. yedoensis var. nudiflora* harbours two sets of phased chromosomes; thus, this 48.5 Mb sequences should be redundant, and the unique sequence in the collinear blocks could be half of these sequences. When considering collinearity by all sizes of blocks, both of these comparisons showed good collinearity (Fig. 2a, Fig. S2 and Fig. S3). Considering that the large collinear blocks accounted for a small amount of the whole genome and that there was good collinearity between the PCP genome and the other two genomes in the subgenus *Cerasus*, these results suggested that there is extensive genomic variation between PCP and the two other flowering cherries.

The genome duplications were also analysed by calculating synonymous substitution (Ks) and fourfold degenerative third-codon transversion values (4DTv) for five plants, including *Arabidopsis thaliana*, *M. × domestica. P. persica and C. serrulata.* Clearly, two peaks for both values could be observed for plants within the

Rosaceae family (Fig. 2b and 2c). These data suggested that two whole-genome duplications (WGDs) occurred in the Rosaceae family. The first WGD generated peaks of 4DTv at ca. 0.38 and Ks at 1.5, and this pattern could be observed for a number of plants. This result was reported in a number of previous studies. The second WGD in the Rosaceae family seemed to have occurred recently. According to the peak value (5.27×10^{-2}) of Ks corresponding to the recent WGD for PCP, we could predict that this WGD occurred at 1.75 MYA (t= Ks/2r, r represents the substitution rate per year and equals 1.5×10^{-8}).

To gain insight into PCP genome evolution, an orthologue cluster analysis was performed for PCP and seven other plants. These seven genomes include those of *Populus trichocarpa*, *Vitis vinifera*, *A. thaliana*, *P. persica*, *Malus* × *domestica, C. serrulata* and haplotype A of *C.*× *yedoensis 'Somei Yoshino'*. The orthologue analysis revealed that a total of 36895 gene families were clustered in these eight plant species. A total of 9437 gene families were common among all plants, while 89 gene families, including 2268 genes, were unique to PCP (Fig. 3a). Of the 9437 gene families common to all eight plants, there were 1067 single-copy orthologous genes. A phylogenetic tree was constructed based on these 1067 single-copy genes (Fig. 3b). Obviously, the three plants in the subgenus *Cerasus* are clustered together. Divergence analysis revealed that PCP diverged from the common ancestor of *C. serrulata* and *P.* × *yedoensis* 'Somei-Yoshino' at 8.9 MYA, and their ancestor diverged from *P. persica* at 12.4 MYA. Specifically, PCP showed earlier divergence than the other two flowering cherries. These data will help us to elucidate the origin and evolution of flowering cherry in subsequent analyses.

According to orthologue cluster analysis, there were 89 and 179 specific gene families in PCP and the three other flowering cherries, respectively. The 89 PCP-specific gene families included 375 genes, while the 179 flowering cherry-specific gene families included 574 genes. GO functional enrichment analysis showed that PCP-specific genes were enriched in protein binding and protein dimerization activity (Fig. 3c), while flowering cherry-specific genes were enriched in 12 GO terms, such as anion and DNA binding (Fig. 3c). Expansion analyses revealed

234 and 656 gene families that expanded in flowering cherry and PCP ($P<0.05$), respectively. These 234 and 656 gene families correspond to 1379 and 4813 genes, respectively. GO functional enrichment analysis showed that PCP was mainly enriched in functions related to DNA repair (Fig. 3c), while flowering cherry expanded genes were mainly enriched in defence responses (Fig. 3c). These data suggested that the expanded gene in all three flowering cherry accessions enabled this plant to evolve enhanced disease resistance. Moreover, the expansion of DNA repair-related genes in PCP might be attributed to this species originally growing in the high mountains of the southeastern regions of China (Huang, 2003； Zhao *et al.*, 2016, Jia *et al.*, 2017).

Genomic variation map for 312 flowering cherry lines

The evolutionary relationships of different species and cultivars of flowering cherries are still controversial. Thus, a total of 312 tree lines, including flowering cherries in *subgenus Cerasus* and several lines of their close relatives, were collected as the core germplasm panel in this study (Table S11). Of these 312 tree lines, 306 lines can be considered flowering cherry germplasms, while the other six lines could be considered relatives of flowering cherry or members of *subgenus Cerasus*. The taxonomic classification of these six lines could be performed in a subsequent study. Of the 306 lines, 71, 158 and 77 lines could be classified as wild species, cultivars and breeding lines (Table S11), respectively. For these 71 lines of wild species, 69 were native to China, and 2 were native to Japan. Of the 158 cultivars, 131, 9 and 18 were developed in Japan, China and other countries, respectively. The 77 breeding lines were collected as part of our research programs.

To obtain the genotypes of this germplasm panel, whole genome resequencing for each line was performed by the Illumina Hi-Seq platform. A total of 1830 GB of high-quality clean paired-end (PE) reads were generated with an average of 5.83 GB of data per plant, which is 20× coverage of the PCP genome. These clean reads were mapped onto our assembled PCP genome, and variants were called with standardized pipelines. After quality filtering of these variants, a total of 761267 high-quality genomic variants, including 621010 SNPs and 140257 insertions and deletions

(InDels), were identified. According to the locations of these genomic variants, the top three types of locations were downstream of gene variants (28.21%), intergenic regions (27.28%), and upstream gene variants (27.04%) (Table S12). Specifically, there were 120,587 (7.04%) genomic variants located in exon regions (Table S12). Of these 761267 genomic variants, 53850, 601 and 61898 showed missense, nonsense and silent mutations in genes; thus, these missense and nonsense types of mutations play major roles in generating trait variations in this germplasm (Table S12).

Population structure, evolution and differentiation

Based on the genotypes of all 312 accessions in our germplasm, the phylogenetic tree was first constructed with a maximum-likelihood (ML) method. According to the taxonomic classification, Mei flower (*P. mume*) is the most distant plant from all other lines in the phylogeny; thus, it was used as the outgroup to root our phylogenetic tree. Clearly, all 306 accessions could be classified into 3 clades, A, B and C (Fig. 4a). According to the original locations of these 3 clades, A and B were from China, while C was from Japan (Table S11). In total, 75, 82 and 149 accessions were in clades A, B and C, respectively (Table S13). With further insight into this phylogenetic tree, several results could be summarized. First, Mei flower seemed to be the common ancestor of the other five plants in outgroups, including a66 (*P. tomentosa*), a24 (*P. glandulosa*), a23 (*P. virginiana 'Schubert'*), a221 (*P. mahaleb*) and a43 (*P. maackii*) (Fig. 4a). Second, a66 and a24 formed one evolutionary branch, while a23, a221, and a43 formed the other branch. Third, the first branch formed by a66 and a24 seemed to be the ancestor of clade A, while the other branch formed by a23, a221, and a43 seemed to be the ancestor of clades B and C. Fourth, clades B and C seemed to challenge parallel evolution. A principal component analysis (PCA) also yielded a similar result. The six accessions in the outgroup were placed in the middle of all plant materials when using the first three components, while all the other 306 flowering cherry accessions were placed in three distinct positions (Fig. 4b). According to these results, we speculate that there were two origins of flowering cherry: one formed clade A, and the other formed clades B and C.

Population structures for all 306 flowering cherry accessions were also predicted.

The optional subpopulation numbers were determined by the delta K method (Evanno *et al.*, 2005). Clearly, the cross-validation (CV) error showed the lowest number when K was set as 10 (Fig. S4); thus, the optional subpopulation number was determined to be 10. The proportion of ancestors for each accession is illustrated in Fig. 4c. According to these data, there were 48, 94, 37, 25, 19, 13, 23, 10, 33, and 4 accessions in subpopulations 1 (Pop1) to 10 (Pop10) (Fig. 4d and Table S13), respectively. Accessions in clade A mainly formed Pop1, Pop8 and Pop10, accessions in clade B mainly formed Pop3 and Pop7, and accessions in clade C mainly formed Pop2, Pop5, Pop6 and Pop9. Pop4 showed a mixture of classes A, B and C. According to these data, we could summarize that all three clades of flowering cherry formed more than one subpopulation, and the pedigrees of some accessions in our germplasm set were extensively mixed (Pop4).

The nucleotide diversity (Pi) for the three clades was also calculated, and the Pi values for clades A, B and C were 0.31, 0.11 and 0.09, respectively. According to these data, nucleotide diversity was much higher in A than in B and C ($P<0.05$) (Fig. 4e). The fixation index (Fst) values were calculated to evaluate the genetic differentiation among the three clades. Clades A/B, A/C and B/C showed Fst values of 0.20, 0.21 and 0.12, respectively. These data indicated a larger degree of differentiation of clade A from B/C than from clades B and C, although both clades A and B were from China. This result was highly consistent with our prediction of the origin of flowering cherry; clade A clearly originated from one branch, while clades B and C originated from the other branch. The higher nucleotide diversity of clade A indicates that this branch evolved earlier or that more pedigrees from other plants in the *Prunus* genus were incorporated into flowering cherry breeding. The linkage disequilibrium (LD) decay analysis revealed higher LD in clade A than in clades B and C (Fig. 4f). This phenomenon might be attributed to a large number of breeding lines being collected for clades A and C, whereas clade B basically consisted of wild species individuals.

To investigate the genomic differentiation among the three clades of our germplasms, the cross-population composite likelihood ratio (XP-CLR) selective

sweep parameter was calculated for A/C, A/B and B/C. Genomic regions with the top 5% XP-CLR values for each comparison were regarded as highly differentiated genomic regions. In total, there were 6980, 6991 and 6556 genes located in the highly differentiated genomic regions between clades A/B, A/C and B/C, respectively (Fig. 5a. 5c and 5d, Tables S14-16), respectively. When we investigated the functions of genes located in the top 30 differentiated genomic regions for each comparison, we found genes related to the regulation of plant height (*GA2OX8*, *DWARF* and auxin-related genes), anthocyanin biosynthesis (*WD40*), flower development (*AGL60*, *TCP* and *AGAMOUS-like 62*), nitrogen and phosphate uptake (*Pht1;4* and *Nitrate transporter 1.1*), ultraviolet ray response (*FAR*) and temperature response (*HSF20* and *CBF*) located in the peak regions (Fig. 5a. 5c and 5d). These patterns are highly consistent with some different traits among these three clades. For example, accessions in clade A usually grow in the southeastern and southern regions of China, and they grow faster, have darker flowers, and have less resistance to cold but greater tolerance to heat and ultraviolet rays than those in B and C. Moreover, functional enrichment analysis of genes in the highly differentiated genomic regions between A and B revealed that they are associated with a number of biological GO terms related to floral fragrance, such as cinnamyl-alcohol dehydrogenase activity, sinapyl alcohol dehydrogenase activity and the isoprenoid metabolic pathway (Fig. 5b). These data are highly consistent with the different floral fragrances between clades A and B. Generally, accessions in clade A showed no floral fragrance, while most of the accessions in clade B showed slight floral fragrance. Similarly, GO terms related to the oxylipin biosynthetic process and auxin response were enriched in comparisons of A with C and B with C (Fig. 5d and 5f). This suggested different disease resistance and growth speeds between the two comparisons.

Genome-wide association study (GWAS) of flower colour variations and identification of causal genes

Flower colour is a major trait for flowering cherry, and uncovering the regulation of this trait would facilitate the breeding of more valuable cultivars for this ornamental tree. Petals with red, yellow and white colours from the fresh flowers of different

flowering cherry lines were used to investigate the compositions of pigments. Nontargeted metabolite examinations revealed that three previously described pigments, cyanidin 3-O-glucoside (Cy3G), delphinidin-3-O-glucoside (Dp3G) and pelargonidin3-O-glucoside (Pg3G), showed clearly different contents among red, yellow and white flowers. Setting the intensity of white flowers as 1, the red and yellow flowers are *ca*. 1186- and 17-fold for Cy3G, respectively (Table S17). The intensities of Dp3G are *ca*. 10- and 0.5-fold greater than the white flowers for red and yellow flowers, respectively. The intensities of Pg3G are *ca*. 0.4- and 49-fold to the white flowers for red and yellow flowers, respectively. Considering the colour of these three types of pigments, it could be summarized that Cy3G is the major anthocyanin in red flowers, while Cy3G and Pg3G are the major anthocyanins in yellow flowers.

To investigate QTLs for flower colour of flowering cherry, this trait was observed for the 312 accessions in two supersessive years. A colorimeter was used to measure variations in flower colour for all 312 accessions. Combining the genotype and phenotype data, GWASs of flower colour with different statistical models and software were performed for our germplasm panel. One QTL located on Chr02 could be detected in most models, and this QTL identified in the linear mixed model (LMM) implemented in genome-wide efficient mixed model association (GEMMA) software showed the most significant marker–trait association (MTA). Thus, the MTAs in this model were selected for further analyses. A total of 57 genomic variants showed a significant association with flower colour variations in this model, and these variants could be merged into seven QTLs according to their locations (Fig. 6a and Table S18). The Q-Q plot of all the adjusted P values revealed high confidence in our GWAS results (Fig. 6b). Of these seven QTLs, QTL5 and QTL7 harboured only one significant SNP and were designated Chr02: 16721554 and Chr05: 9502164 (Table S18). Comprehensive investigation of genes located in these QTLs showed that QTL1 to QTL7 harboured 5, 32, 1, 0, 1, 1 and 0 genes, respectively (Table S18). The effects of all 57 genomic variants were analysed with SNPEff pipelines, and one SNP, Chr02:9710510, showed a moderate effect on its gene, *evm.model.LG02.1464*. Two

alleles of this SNP, G and T, formed three genotypes, GG, GT and TT. The SNP from G to T in the gene *evm.model.LG02.1464* changed its encoded amino acid from leucine to isoleucine. The parameters of flower colour for these three genotypes were significantly different from each other. The exact values of these parameters for GG, GT and TT were 74.0±6.4, 66.0 ±7.7 and 60.6 ±4.0, respectively (Fig. 6c). These results suggested that this SNP was highly associated with flower colour variations in flowering cherry. Further investigation of the gene function of *evm.model.LG02.1464* revealed that it encodes glycosyl transferase. Considering that glycosyl transferase is one of the most important enzymes for the biosynthesis of glucoside of anthocyanins, *evm.model.LG02.1464* is a candidate gene for flower colour variations in flowering cherry.

Discussion

PCP is native to China and is an excellent flowering cherry cultivar due to its attractive flower traits, including dark red colour, multiple petals and blossoms in early spring. Recently, PCP has been widely cultivated in central and southern China (Zhou *et al.*, 2019). In this study, the high-quality chromosome-scale genome of PCP was generated with a combination of long- and short-read sequencing. The Hi-C scaffolding strategy enabled us to anchor 99.98% of the assembled sequences to eight chromosomes. Compared with the other two published genomes of flowering cherry (Baek *et al.*, 2018, Yi *et al.*, 2020), the PCP genome showed higher completeness and continuity. The contig N50 of PCP is 18.31 MB, which is longer than that of *C. serrulata* and *P. yedoensis* (Table S2). BUSCO analysis revealed that the PCP genome covered 96.18% of core genes, which is also higher than that of *C. serrulata* (Chin *et al.*, 2014, Baek *et al.*, 2018). In the collinearity analyses between the PCP genome and the other two genomes in flowering cherry, high chromosome-scale collinearity was clear among these three genomes (Fig. 2a, Figs. S2 and S3); however, there were still extensive variations in small fragments. Only 11.5% of PCP genomes showed high identity to the genome of *C. serrulata*. These data suggested that one genome for

flowering cherry is not enough. In the germplasm resequencing analysis, the PCP relatives belonged to clade A (a78 and a120, Table S13), while the relatives of *C. serrulata* (a37 and a59, Table S13) and *P. yedoensis* (a101, Table S13) belonged to clade C. Clade A originated from one branch in Fig. 4a, and clades B and C originated from the other branch. Thus, all three genomes are valuable resources representing different origins of flowering cherry. Taken together, these findings indicate that the PCP genome is high-quality and can be used as a reference for flowering cherry in future studies.

In previous reports, there were debates about the origins of flowering cherries. According to morphological studies on Japanese flowering cherry cultivars, some studies indicated that most cultivars originated from native Japanese taxa and hybrids between them (Kato *et al.*, 2014), while other studies indicated that these Japanese flowering cherry cultivars showed close relations with wild lines in China (Katori *et al.*, 2002, Ogawa *et al.*, 2012). Some reports have speculated that Japanese flowering cherries spread eastwards from the Himalayan Mountains to Japan and constantly diverged into many species (Ma *et al.*, 2009). It is speculated that there must be flowering cherry species in the eastern Himalayas corresponding to Japanese flowering cherries (Ma *et al.*, 2009). In this study, our results provide novel insights into the origin of flowering cherries. We found two origins of flowering cherries, and Mei flower is the common ancestor of these two origins. In our germplasm panel, five accessions, a24 (*P. glandulosa*), a66 (*P. tomentosaa*), a23 (*P. virginiana* 'Schubert'), a221 (*P. mahaleb*) and a43 (*P. maackii*), were identified as outgroups or close relatives to flowering cherries according to their trait values. In the phylogenetic analysis, these five accessions were clearly assigned to two branches (Fig. 4a). Lines a24 and a66 on branch A could be considered the ancestors of clade A, and lines a23, a221 and a43 on branch B could be considered the ancestors of clades B and C (Fig. 4a). The PCA of all 312 accessions also supported the phylogenetic analysis (Fig. 4b). According to the distributions of accessions in these three clades, accessions in clade A were mainly distributed in central, southern and eastern China (Table S11), accessions in clade B were mainly distributed in central and northern China, and

accessions in clade C were mainly distributed in Japan (Table S11). Thus, accessions a24 and a66 in clade A were distributed in a vast area in China, from south to north, while accessions a23, a221 and a43 were mainly distributed in northern and northeastern China. Therefore, these distributions are highly consistent with the distributions of accessions in clades A, B and C. Additionally, some lines, such as a85, a47 and a57, in clade A close to the outgroups a24 and a66 were wild lines, and they were originally collected in Fujian or Yunnan Province, which is close to the Himalayan region (Table S11). The locations of these two wild lines would be close to the origin site of clade A. In other words, flowering cherries in clade A originated from southern China, such as the Himalayan Mountains. For clades B and C, the ancestors of a21 and a43 were originally collected in Ha'erbin, Heilongjiang Province, northeast China (Table S11). Thus, these data suggested that clades B and C originated from northeastern China. After origination, some accessions spread to the south (west and central China) to form clade B, while some accessions spread to Japan to form clade C. Population differentiation also supported our speculations. The Fst values of A/C and A/B are much higher than those of B/C (Fig. 4e). Additionally, nucleotide diversity was higher in A than in B and C, which might suggest that clade A originated earlier than clades B and C (Fig. 4e). The phenotypes of accessions in these three clades also supported our speculation. Accessions in A usually show red flowers, while accessions in B and C usually show white or pink flowers. The previous controversial views for the origin of cherry flowers might be because some studies collected plant materials from clades A and C, while some other studies collected plant material from clades B and C, or these studies did not collect enough genotypic data.

Petal colour is a very important ornamental feature of flowering cherries. Metabolite examination revealed three different types of anthocyanins from different colours of flowers in flowering cherries. Furthermore, we speculated that Cy3G would be the major anthocyanin for red flowers, while a mixture of Cy3G and Pg3G would be the major anthocyanin for yellow flowers. According to the GWAS, we were able to identify candidate genes for the biosynthesis of different contents of

anthocyanins in different types of flowers. The candidate gene encodes a glycosylation transferase, which is one of the key enzymes in the biosynthesis of anthocyanins. A number of previous studies also noted that glycosylation transferase catalyses the glycosylation of anthocyanins (Cheng *et al.*, 2014, Zhao *et al.*, 2020). These results suggested that this candidate gene has a high possibility of regulating different flower colours in flowering cherry and can be used as a target for the molecular breeding of more cultivars of this ornamental tree. However, we should mention that we identified seven QTLs for flower colour variations, and only one candidate gene was predicted for one QTL. Additionally, a number of other ornamental traits, such as the number of petals, tree architecture, plant height, and tolerance to heat, cold and high salinity, are also important in this tree. Thus, further comprehensive studies can be performed on our flowering cherry germplasms.

Materials and methods

Plant materials

The PCP cultivar growing at the Institute of Landscape Architecture in Wuhan (Wuhan, China) was used as plant material for genome sequencing and RNA-Seq. Three lines, PCP (deep red flowers), *P. serrulata 'Grandiflora'* (yellow flowers) *and P. conradinae 'Chuyuanfenghou'* (white flowers), were used for pigment analysis. Three biological replicates were set per sample. A total of 312 accessions, including 306 flowering cherry and 6 outgroup lines, were collected for resequencing. All detailed information for these 312 lines is listed in Table S11.

Genome sequencing, resequencing and RNA-Seq Samples of young leaves were used to extract genomic DNA with a QIAGEN® Genomic kit according to the standard operating procedure. Genome sequencing was performed on the Nanopore PromethION (Oxford Nanopore Technologies, Oxford Science Park, UK) and Illumina NovaSeq 6000 platforms (Illumina, Inc., San Diego, USA). RNA was isolated from five PCP tissues (roots, stems, leaves, petioles and flowers) by using RNAiso Plus (Cat. No. 9109, TAKARA Biotech (Beijing) according to the

manufacturer's instructions for gene prediction. Genome sequencing, resequencing and RNA-Seq were completed by Wuhan Hope Group Biology Co., Ltd. (Software New Town, Wuhan, China).

Genome assembly

For de novo genome assembly, Nanopore PromethION and a high-throughput chromatin conformation capture (Hi-C)-based scaffolding method were used to obtain the chromosome-level assembly. After filtering with Guppy V3.2.2 (Wick *et al.*, 2019), the preliminary genome was assembled by NextDenovo v2.3.0 (https://github.com/Nextomics/NextDenovo.git) with the correction parameters (reads_cutoff: 1k, seed_cutoff: 44k) and default assembly parameters. To improve the accuracy of the assembly, the contigs were refined with Racon v1.3.1 (https://github.com/isovic/racon.git) using ONT long reads and Nextpolish using Illumina short reads with default parameters (Hu *et al.*, 2020). To discard possibly redundant contigs and generate a final assembly, similarity searches were performed with the same parameters.

The clean reads were mapped to the draft assembled genome using Bowtie2 v2.3.2 with default parameters to obtain the unique mapped reads (Langmead and Salzberg, 2012). Valid interaction reads were identified and retained in HiC-Pro v2.8.1 from unique mapped paired-end reads for further analysis (Servant *et al.*, 2015). The scaffolds were further clustered, ordered, and oriented onto chromosomes by LACHESIS (Burton *et al.*, 2013). Finally, placement and orientation errors exhibiting obvious discrete chromatin interaction patterns were manually adjusted.

Repeat annotation

We first used the software GMATA v2.2 (Wang and Wang, 2016) to identify simple sequence repeats (SSRs) with the default settings and Tandem Repeats Finder (TRF) v 4.07b (Gary, 1999, Xuewen *et al.*, 2016) with the default parameters to determine all tandem repeat elements (TRFs) throughout the genome. Two methods were combined to search for transposable elements (TEs) in our genome: de novo and

homology-based methods. Briefly, a de novo repeat library was obtained using MITE-hunter (Han and Wessler, 2010) with default parameters and Repeat Modeller v1.0.11 (Joseph *et al.*, 2000, Yujun *et al.*, 2010). Then, the library was aligned to TE classes in Repbase to classify them into different repeat families (György *et al.*, Jurka *et al.*, 2005b). Furthermore, Repeat Masker v1.331 was employed to search for known and novel TEs by aligning sequences against the de novo repeat library and Repbase TE library (Jurka *et al.*, 2005a). Overlapping TEs of the same repeat class were collated and combined.

Gene prediction

Three independent methods, namely, de novo prediction, homology searching, and RNA-Seq-based prediction, were used for gene prediction in the repeat-masked genome. In detail, GeMoMa v1.6.1 was used to align the homologous peptides from the five representative species, namely, *A. thaliana*, *P. persica*, *M.* × *domestica*, *P. dulcis*, and *Rubus occidentalis*, to the assembled genome. The predicted results of all homologous species were then integrated to obtain the structural information of the genome (Jens *et al.*, 2019). For gene prediction based on RNA-Seq, RNA isolated from five PCP tissues (roots, stems, leaves, petioles and flowers) was sequenced and generated approximately 43.3 Gb clean Illumina PE mRNA-Seq reads. These clean mRNA-Seq reads were aligned to the reference genome using STAR v2.7.3a (Dobin *et al.*, 2012). The transcripts were then assembled using StringTie v1.3.4d, and open reading frames (ORFs) were predicted using PASAv2.3.3 (Haas *et al.*, 2008, Kovaka *et al.*, 2019). After that, a training set was obtained by the above two steps, and Augustus v3.3.1 was applied for ab initio gene prediction with this training set (https://github.com/Gaius-Augustus/Augustus). Finally, Evidence Modeller (EVM) v1.1.1 was used to obtain an integrated gene set in which genes with TEs were removed using the Transposon PSI package, and the miscoded genes were further filtered out (Haas *et al.*, 2008, Kovaka *et al.*, 2019). The longest transcripts for each locus were kept for further analyses.

Annotation of noncoding RNAs (ncRNAs)

Two methods, searching against the database and based on model prediction, were used to annotate the noncoding RNA (ncRNA). Transfer RNAs (tRNAs) were predicted using tRNAscan-SE v1.1.2 with default parameters. MicroRNA, rRNA, and snRNA were identified using Infernal software to search the Rfam database (Griffiths-Jones *et al.*, 2005, Nakatsuka *et al.*, 2008). The rRNAs and their subunits were predicted using RNAmmer v1.2 software (Griffiths-Jones *et al.*, 2005).

Functional annotation of genes

Gene function information, motifs and domains of their proteins were annotated by aligning against public databases, including NR, KOG, Gene Ontology, KEGG and SwissProt. The putative domains and GO terms of genes were identified using the InterProScan program with default parameters (Jones *et al.*, 2014). For the other four databases, blastp was employed to compare the Evidence Modeller-integrated protein sequences against the four well-known public protein databases with an E value cut-off of $1E^{-05,}$ and the hits with the lowest E value were retained. Ultimately, the results from the five databases were merged.

Genome assembly evaluation

The evaluation of genome assembly completeness was carried out with BUSCO v4.0.5 with default parameters (Simao *et al.*, 2015). The sequence conformity assessment was completed by BWA (Heng *et al.*, 2010) and SAMtools v1.4 (Li, 2009).

Genome comparison and evolution analyses

Gene family and single-copy orthologous genes between PCP and 7 other plants were identified using OrthoMCL v2.0.9 and MAFFT v7.313 (Li *et al.*, 2003, Kazutaka *et al.*, 2013). Phylogenetic analysis was performed using RAXML v8.2.10 with the parameters "-m PROTGAMMAAUTO -p 12345 -T 8 -f b" (Yang, 2007, Kazutaka *et al.*, 2013). The phylogenetic tree was constructed by MEGA (Yang, 2007) and Figtree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). Gene expansions and contractions

were detected using CAFÉ v4.2.1 with the parameters "-p 0.05 -t 10 -r 10000", which simulate a random birth and death process to predict the gene family evolution of different species on each evolutionary branch (Bie *et al.*, 2006). Furthermore, to identify positively selected genes in PCP, we calculated average Ka/Ks values and calculated likelihood values of the branch-site models using Ka/Ks_calculator v2.0 (Zhang, 2022). Genes with a p value < 0.05 were considered positively selected genes. Genome collinearity analysis among PCP, *C. serrulata* and *P. yedoensis* was conducted with wgd (https://github.com/arzwa/wgd).

Genomic variation identification

Flowering cherry germplasm lines were resequenced by the Illumina 2000 Platform. After filtering out the duplicated reads, the adaptor and low-quality reads were filtered using Trimmomatic v0.40 (Bolger *et al.*, 2014). Clean reads were mapped onto the reference PCP genome with BWA (Li, 2009). A standardized GATK pipeline was employed for genomic variant calling (do Valle *et al.*, 2016, Brouard *et al.*, 2019). The final genotyping of the population was performed using Genotype GVCFs under default settings. The SNPs were filtered for quality to apply the following criteria: quality/depth<2.0 || FS>60.0 || MQ (quality of the mapped reads of one site) <40.0 || MQRanksun <-12.5 || ReadposRankSum <-8.0. The SNPs resulting from joint genotyping were further filtered to remove SNP sites with an MAF<0.05, SNPs with a sequencing depth<4, and those that had samples with missing data.

Population structure analysis

FastTree software was used to build a maximum-likelihood (ML) phylogenetic tree for the 312 germplasm lines, and the tree was visualized using iTOL (http://itol.embl.de) (Price *et al.*, 2009, Liu *et al.*, 2011, Letunic and Bork, 2019). The GCTA v1.94.1 program was employed to perform principal component analysis (PCA) with the default parameters (Yang *et al.*, 2011). ADMIXTURE v1.3.0 was used to investigate population structure, specifying *K* values ranging from 2 to 20 (Alexander *et al.*, 2009). The most suitable number of ancestral populations was determined by the *K* value with the lowest cross-validation (CV) error.

Fst between pairs of subpopulations was calculated using VCFtools v0.1.13 with a sliding window size of 100 kb and step size of 10 kb (Danecek *et al.*, 2011). The XP-CLR values were calculated for predicting the genomic regions with selective sweeps by using XP-CLR packages implemented in Python (Chen *et al.*, 2010). Genomic regions with values above the top XP-CLR values were considered to show selective sweeps.

Genome-wide association study (GWAS)

GWASs were carried out using genomic association and prediction integrated tool (GAPIT) software v3.0 based on different statistical models, including the generalized linear model (GLM), mixed linear model (MLM), compressed mixed linear model (CMLM), efficient mixed model association expedited (EMMAX), and factored spectrally transformed linear mixed model (FAST-LMM) (Lipka *et al.*, 2012). GWASs were also carried out using GEMMA software (Zhou and Stephens, 2012). The first three principal components of all the flowering cherry panels were calculated with GAPIT software, and they were considered covariates in these models. The SNP association *p* value was adjusted for the false discovery rate (FDR) using the Benjamini and Hochberg method. The threshold for significant association was set as 0.05/total SNPs. SNPs with an FDR less than the threshold were considered significant SNPs. Flower colour was measured using a Precise Color Reader WF32 (ShenZhen Wave Optoelectronics Technology Co., Ltd., China). Lightness (L*) and the two chromatic components a* and b* of the CIEL*a*b* colour coordinates were measured under daylight conditions. Chroma (C*) was calculated based on the equation $C^* = (a^{*2} + b^{*2})^{1/2}$. The average value of the 5 petal colours of each material was measured at the centre of the blooming petal. The flower colour was measured in two successive years, 2021-2022, and all these germplasm lines were grown at the Institute of Landscape Architecture in Wuhan (Wuhan, China). The best linear unbiased prediction (BLUP) values for the two-year flower colour intensities were calculated for GWASs according to the formula $Y_{ij}=H_i+a_j+e_{ij}$, where Y_{ij} indicates values of traits, H_i indicates the fixed effect of the two-year environment, a_j indicates the random effect of genotypes and e_{ij} indicates the residuals.

Detecting petal pigments

A total of 0.5 g fresh petals per biological sample was collected for detecting nontarget pigment contents. Anthocyanin extraction was conducted according to the manufacturer's instructions (Adjé *et al.*, 2010). The anthocyanins were analysed using a UHPLC–MS instrument, and the parameters were as follows: mobile phase A was 0.1% formic acid in water, and mobile phase B was acetonitrile. The column temperature was set at 40 °C. The autosampler temperature was set at 4 °C, and the injection volume was 2 μL. A Sciex QTrap 6500+ (Sciex Technologies) was applied for assay development. Typical ion source parameters were as follows: IonSpray voltage: +5500/-4500 V, curtain gas: 35 psi, temperature: 400 °C, ion source gas 1:60 psi, ion source gas 2: 60 psi, and DP: ± 100 V. UHPLC–MS data analysis was performed using previously reported methods (Guy *et al.*, 2008).

Acknowledgements

Financial support for this work was provided by the Knowledge Innovation Project of Wuhan Science and Technology Bureau (No: 202202221011015010), the Science and Technology Planning Project of Wuhan Landscape and Forestry Bureau (No.: 20174901) and the Forestry Science and Technology Promotion Project of Wuhan Landscape and Forestry Bureau (No.: 20221619).

Author contributions

Chaoren Nie, Xiaoqin Zhang, Wensheng Xia, Hongbing Sun,and Na Li collected samples and performed phenotypic measurements.Chaoren Nie, Yingjie Zhang, Sisi Zhang and Na Li conducted data analyses. Chaoren Nie, Nian Wang, and Yingmin LV participated in manuscript writing and editing. Yingmin LV, Nian Wang and Zhaoquan Ding supervised this project.

Data availability

The whole-genome sequence, RNA-Seq and short read data for the 312 accessions have been deposited in the NCBI database under accession number PRJNA903270. The genome sequence of PCP can be obtained from our own website (http://tree-bio.hzau.edu.cn/download/PCP/).

Competing interests

The authors declare no conflicts of interest.

References

Adjé, F., Lozano, Y.F., Lozano, P., Adima, A., Chemat, F. and Gaydou, E.M. (2010) Optimization of anthocyanin, flavonol and phenolic acid extractions from Delonix regia tree flowers using ultrasound-assisted water extraction. *Industrial Crops & Products*, **32**, 439-444.

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, **19**, 1655-1664.

Baek, S., Choi, K., Kim, G.B., Yu, H.J., Cho, A., Jang, H., Kim, C., Kim, H.J., Chang, K.S., Kim, J.H. and Mun, J.H. (2018) Draft genome sequence of wild Prunus yedoensis reveals massive inter-specific hybridization between sympatric flowering cherries. *Genome Biol*, **19**, 127.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.

Brouard, J.S., Schenkel, F., Marete, A. and Bissonnette, N. (2019) The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *J Anim Sci Biotechnol*, **10**, 44.

Burton, J., Adey, A., Patwardhan, R., Qiu, R., Kitzman, J. and Shendure, J. (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature biotechnology*, **31**.

Chen, H., Patterson, N. and Reich, D. (2010) Population differentiation as a test for selective sweeps. *Genome Research*, **20**, 393-402.

Cheng, J., Wei, G.C., Zhou, H., Gu, C., Vimolmangkang, S., Liao, L. and Han, Y.P. (2014) Unraveling the Mechanism Underlying the Glycosylation and Methylation of Anthocyanins in Peach. *Plant Physiol*, **166**, 1044-1058.

Chin, S.W., Shaw, J., Haberle, R., Wen, J. and Potter, D. (2014) Diversification of almonds, peaches, plums and cherries - Molecular systematics and biogeographic history of Prunus (Rosaceae). *Mol Phylogenet Evol*, **76**, 34-48.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. and Genomes Project Analysis, G. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

do Valle, I.F., Giampieri, E., Simonetti, G., Padella, A., Manfrini, M., Ferrari, A., Papayannidis, C., Zironi, I., Garonzi, M., Bernardi, S., Delledonne, M., Martinelli, G., Remondini, D. and Castellani, G. (2016) Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. *Bmc Bioinformatics*, **17**, 341.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J. and Gingeras, T.R. (2012) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A. and Bateman, A. (2005) Rfam: Annotating Non-Coding RNAs in Complete Genomes. *Nucleic Acids Research*, **33**, D121-124.

Guy, P.A., Tavazzi, I., Bruce, S.J., Ramadan, Z. and Kochhar, S. (2008) Global metabolic profiling analysis on human urine by UPLC-TOFMS: issues and method validation in nutritional metabolomics. *Journal of Chromatography B Analytical Technologies in the Biomedical & Life Sciences*, **871**, 253-260.

Haas, B.J., Salzberg, S.L., Wei, Z. and Pertea..., M. (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology*, **9**, R7.

Heng, Li, Durbin and Richard (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*.

Hideaki Ohba, T.K., Hideaki Tanaka,Hiroshi Hihara (2007) *Flowering cherries of Japan,Newth edn. Yama-kei, Tokyo (in Japanese)*: YAMA-KEI Pubishiers.

Hu, J., Fan, J.P., Sun, Z.Y. and Liu, S.L. (2020) NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, **36**, 2253-2255.

Huang, T. (2003) Flora of Taiwan, Vol. 6.

International Peach Genome, I., Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J., Cattonaro, F., Zuccolo, A., Rossini, L., Jenkins, J., Vendramin, E., Meisel, L.A., Decroocq, V., Sosinski, B., Prochnik, S., Mitros, T., Policriti, A., Cipriani, G., Dondini, L., Ficklin, S., Goodstein, D.M., Xuan, P., Del Fabbro, C., Aramini, V., Copetti, D., Gonzalez, S., Horner, D.S., Falchi, R., Lucas, S., Mica, E., Maldonado, J., Lazzari, B., Bielenberg, D., Pirona, R., Miculan, M., Barakat, A., Testolin, R., Stella, A., Tartarini, S., Tonutti, P., Arus, P., Orellana, A., Wells, C., Main, D., Vizzotto, G., Silva, H., Salamini, F., Schmutz, J., Morgante, M. and Rokhsar, D.S. (2013) The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet*, **45**, 487-494.

Jens, Keilwagen, Frank, Hartung, Jan and Grau (2019) GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. *Methods in Molecular Biology*.

Jia, Y., Liu, M.-L., Yue, M., Zhao, Z., Zhao, G.-F. and Li, Z. (2017) Comparative Transcriptome Analysis Reveals Adaptive Evolution of Notopterygium incisum and Notopterygium franchetii, Two High-Alpine Herbal Species Endemic to China. *Molecules*, **22**, 1158.

Jiang, F., Zhang, J., Wang, S., Yang, L., Luo, Y., Gao, S., Zhang, M., Wu, S., Hu, S., Sun, H. and Wang, Y. (2019) The apricot (Prunus armeniaca L.) genome elucidates Rosaceae evolution and beta-carotenoid synthesis. *Hortic Res*, **6**, 128.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W.Z., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R. and Hunter, S. (2014) InterProScan 5: genome-scale protein function classification.

Bioinformatics, **30**, 1236-1240.

Kato, S., Matsumoto, A., Yoshimura, K., Katsuki, T., Iwamoto, K., Kawahara, T., Mukai, Y., Tsuda, Y., Ishio, S., Nakamura, K., Moriwaki, K., Shiroishi, T., Gojobori, T. and Yoshimaru, H. (2014) Origins of Japanese flowering cherry (Prunus subgenus Cerasus) cultivars revealed using nuclear SSR markers. *Tree Genet Genomes*, **10**, 477-487.

Kato, S., Matsumoto, A., Yoshimura, K., Katsuki, T., Iwamoto, K., Tsuda, Y., Ishio, S., Nakamura, K., Moriwaki, K., Shiroishi, T., Gojobori, T. and Yoshimaru, H. (2012) Clone identification in Japanese flowering cherry (Prunus subgenus Cerasus) cultivars using nuclear SSR markers. *Breed Sci*, **62**, 248-255.

Katori, M., Watanabe, K., Nomura, K. and Yoneda, K. (2002) Cultivar Differences in Anthocyanin and Carotenoid Pigments in the Petals of the Flowering Lotus (Nelumbo spp.). *Journal of the Japanese Society for Horticultural Science*, **71**, 473-479.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R. and Pertea, M. (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome biology*, **20**.

Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-U354.

Letunic, I. and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*, **47**, W256-W259.

Li, H. (2009) Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics (Oxford, England)*, **25**, 1754-1760.

Li, X., Kui, L., Zhang, J., Xie, Y., Wang, L., Yan, Y., Wang, N., Xu, J., Li, C., Wang, W., van Nocker, S., Dong, Y., Ma, F. and Guan, Q. (2016) Improved hybrid de novo genome assembly of domesticated apple (Malus x domestica). *Gigascience*, **5**, 35.

Lipka, A.E., Tian, F., Wang, Q.S., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S. and Zhang, Z.W. (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics*, **28**, 2397-2399.

Liu, K., Linder, C.R. and Warnow, T. (2011) RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. *PLoS One*, **6**, e27731.

Ma, H., Olsen, Richard, Pooler, Margaret, Kramer and Matthew (2009) Evaluation of Flowering Cherry Species, Hybrids, and Cultivars Using Simple Sequence Repeat Markers. *Journal of the American Society for Horticultural Science*.

Mattioli, R., Francioso, A., Mosca, L. and Silva, P. (2020) Anthocyanins: A Comprehensive Review of Their Chemical Properties and Health Effects on Cardiovascular and Neurodegenerative Diseases. *Molecules*, **25**.

Nakatsuka, T., Sato, K., Takahashi, H., Yamamura, S. and Nishihara, M. (2008) Cloning and characterization of the UDP-glucose : anthocyanin 5-O-glucosyltransferase gene from blue-flowered gentian. *Journal of experimental botany*, **59**, 1241-1252.

Narbona, E., del Valle, J.C., Arista, M., Buide, M.L. and Ortiz, P.L. (2021) Major Flower Pigments Originate Different Colour Signals to Pollinators. *Front Ecol Evol*, **9**.

Niwa, T. (1936) Flowering Cherries in Japan observed from Landscape View Point. (No. 2.). *Journal of the Japanese Institute of Landscape Architects*, **3**, 168-188c.

Ogawa, T., Kameyama, Y., Kanazawa, Y., Suzuki, K. and Somego, M. (2012) Origins of early-flowering cherry cultivars, Prunus × kanzakura cv. Atami-zakura and Prunus × kanzakura cv. Kawazu-zakura, revealed by experimental crosses and AFLP analysis. *Scientia Horticulturae*,

140, 140-148.

Ono, K., Akagi, T., Morimoto, T., Wunsch, A. and Tao, R. (2018) Genome Re-Sequencing of Diverse Sweet Cherry (Prunus avium) Individuals Reveals a Modifier Gene Mutation Conferring Pollen-Part Self-Compatibility. *Plant Cell Physiol*, **59**, 1265-1275.

Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*, **26**, 1641-1650.

Servant, Nicolas, Varoquaux, Nelle, Lajoie, Bryan, R., Viara, Eric, Chen and Chong-Jian (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing.

Shirasawa, K., Esumi, T., Hirakawa, H., Tanaka, H., Itai, A., Ghelfi, A., Nagasaki, H. and Isobe, S. (2019) Phased genome sequence of an interspecific hybrid flowering cherry, 'Somei-Yoshino' (Cerasus x yedoensis). *DNA Res*, **26**, 379-389.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210-3212.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., Goremykin, V., Komjanc, M., Longhi, S., Magnago, P., Malacarne, G., Malnoy, M., Micheletti, D., Moretto, M., Perazzolli, M., Si-Ammour, A., Vezzulli, S., Zini, E., Eldredge, G., Fitzgerald, L.M., Gutin, N., Lanchbury, J., Macalma, T., Mitchell, J.T., Reid, J., Wardell, B., Kodira, C., Chen, Z., Desany, B., Niazi, F., Palmer, M., Koepke, T., Jiwan, D., Schaeffer, S., Krishnan, V., Wu, C., Chu, V.T., King, S.T., Vick, J., Tao, Q., Mraz, A., Stormo, A., Stormo, K., Bogden, R., Ederle, D., Stella, A., Vecchietti, A., Kater, M.M., Masiero, S., Lasserre, P., Lespinasse, Y., Allan, A.C., Bus, V., Chagne, D., Crowhurst, R.N., Gleave, A.P., Lavezzo, E., Fawcett, J.A., Proost, S., Rouze, P., Sterck, L., Toppo, S., Lazzari, B., Hellens, R.P., Durel, C.E., Gutin, A., Bumgarner, R.E., Gardiner, S.E., Skolnick, M., Egholm, M., Van de Peer, Y., Salamini, F. and Viola, R. (2010) The genome of the domesticated apple (Malus x domestica Borkh.). *Nat Genet*, **42**, 833-839.

Wang, J., Liu, W., Zhu, D., Hong, P., Zhang, S., Xiao, S., Tan, Y., Chen, X., Xu, L., Zong, X., Zhang, L., Wei, H., Yuan, X. and Liu, Q. (2020) Chromosome-scale genome assembly of sweet cherry (Prunus avium L.) cv. Tieton obtained using long-read and Hi-C sequencing. *Hortic Res*, **7**, 122.

Wang, L., Wang, Y., Zhang, J., Feng, Y., Chen, Q., Liu, Z.S., Liu, C.L., He, W., Wang, H., Yang, S.F., Zhang, Y., Luo, Y., Tang, H.R. and Wang, X.R. (2022) Comparative Analysis of Transposable Elements and the Identification of Candidate Centromeric Elements in the Prunus Subgenus Cerasus and Its Relatives. *Genes-Basel*, **13**.

Wick, R.R., Judd, L.M. and Holt, K.E. (2019) Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, **20**.

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M.A., Tao, S., Korban, S.S., Wang, H., Chen, N.J., Nishio, T., Xu, X., Cong, L., Qi, K., Huang, X., Wang, Y., Zhao, X., Wu, J., Deng, C., Gou, C., Zhou, W., Yin, H., Qin, G., Sha, Y., Tao, Y., Chen, H., Yang, Y., Song, Y., Zhan, D., Wang, J., Li, L., Dai, M., Gu, C., Wang, Y., Shi, D., Wang, X., Zhang, H., Zeng, L., Zheng, D., Wang, C., Chen, M., Wang, G., Xie, L., Sovero, V., Sha, S., Huang, W., Zhang, S., Zhang, M., Sun, J., Xu, L., Li, Y., Liu, X., Li, Q., Shen, J., Wang, J., Paull, R.E., Bennetzen, J.L., Wang, J. and Zhang, S. (2013) The genome of the pear (Pyrus bretschneideri Rehd.). *Genome Res*, **23**, 396-408.

Wybe, K. (1999) *Japanese flowering cherries*: Japanese flowering cherries.

Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*, **88**, 76-82.

Yang, Q. (2006) Extraction of Flowering Cherry Pigment and Study of Its Stability. *Guizhou Chemical Industry*.

Yi, X.G., Yu, X.Q., Chen, J., Zhang, M., Liu, S.W., Zhu, H., Li, M., Duan, Y.F., Chen, L., Wu, L., Zhu, S., Sun, Z.S., Liu, X.H. and Wang, X.R. (2020) The genome of Chinese flowering cherry (Cerasus serrulata) provides new insights into Cerasus species. *Hortic Res-England*, **7**.

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., Tao, Y., Wang, J., Yuan, Z., Fan, G., Xing, Z., Han, C., Pan, H., Zhong, X., Shi, W., Liang, X., Du, D., Sun, F., Xu, Z., Hao, R., Lv, T., Lv, Y., Zheng, Z., Sun, M., Luo, L., Cai, M., Gao, Y., Wang, J., Yin, Y., Xu, X., Cheng, T. and Wang, J. (2012) The genome of Prunus mume. *Nature Communications*, **3**, 1318.

Zhang, Z. (2022) KaKs_calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinformatics*.

Zhao, S., Chen, L.-Y., Muchuku, J., Hu, G.-W. and Wang, Q.-F. (2016) Genetic Adaptation of Giant Lobelias (Lobelia aberdarica and Lobelia telekii) to Different Altitudes in East African Mountains. *Front Plant Sci*, **7**.

Zhao, D.Q. and Tao, J. (2015) Recent advances on the development and regulation of flower color in ornamental plants. *Front Plant Sci*, **6**.

Zhao, Q., Haibo, L.I., Yan, Q.U., Wang, Q., Guo, J., Liang, X.U. and Academy, Z.F. (2016) SCAR markers-based molecular identification of 20 flowering cherries(Cerasus) cultivars. *Journal of Nanjing Forestry University(Natural Sciences Edition)*.

Zhao, Y., Dong, W.Q., Zhu, Y.C., Allan, A.C., Kui, L.W. and Xu, C.J. (2020) PpGST1, an anthocyanin-related glutathione S-transferase gene, is essential for fruit coloration in peach. *Plant Biotechnol J*, **18**, 1284-1295.

Zhou, Y.Z., Zheng, Y., Chen, B., Wei, Z.L., Lin, W.J. and Zhao, K. (2019) Chloroplast characterizations and phylogenetic location of a common ornamental cherry cultivar, Prunus campanulata 'Kanhizakura-plena' (Rosaceae). *Mitochondrial DNA B*, **4**, 3938-3940.

Zhou X. and Stephens M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821–824.

Tables

Table 1. Statistics of the *Prunus campanulata* 'Plena' (PCP) genome assembly and annotation

Repeat sequences			
Type	Number	Length (bp)	Percent (%)
LINEs	24576	6725983	2.41
LTR elements	217779	91127962	32.61
DNA transposons	96646	29703125	10.63
Rolling-circles	3774	1123198	0.4
Unclassified	24770	4466166	1.6
Low complexity	150	15462	0.01
Simple repeats	943	85718	0.03
Total	461403	140633497	50.33
Gene predictions			
	Numbers	Mean size (bp)	
Gene	27181	2572.7	
Transcript	27181	1251.3	
Exon	126547	268.8	
Intron	99366	361.5	

Figure legends

Figure 1. Genome assembly of *Prunus campanulata* 'Plena' (PCP)

(a). The plant used for genome assembly. This plant is growing at the Institute of Landscape Architecture in Wuhan (Wuhan, China). The photos in this figure are originals and were taken by the first author, Mr. Chaoren Nie. (b). Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment of the completeness of the PCP genome. (c). Hi-C interaction map for eight pseudochromosomes of the PCP genome. (d). Genome features of the chromosome-scale genome of PCP. Numbers 1 to 6 represent scaffolds (pseudochromosomes), gene density, repeat content, Gypsy transposon content, Copia transposon content, and GC content, respectively.

Figure 2. Genome comparison analysis for *Prunus campanulata* 'Plena' (PCP)

(a). Genome synteny analysis among the three flowering cherries, PCP, *Cerasus* ×

yedoensis 'Somei-Yoshino' and *Cerasus serrulata*. The number above the bars indicates the chromosome index. Synteny blocks were identified by wgd software, and each grey line indicates a gene orthologue. (b) and (c). Synonymous substitution (Ks) (b) and fourfold degenerate third-codon transversion values (4DTv) (c) for five plants, including *Arabidopsis thaliana*, *M. × domestica, P. persica, C. serrulata* and *Prunus campanulata* 'Plena' (PCP).

Figure 3. Genome evolution analysis for *Prunus campanulata* 'Plena' (PCP)
(a). Venn diagram for orthologue cluster analysis of PCP and the other seven plants. These seven genomes are from *Populus trichocarpa*, *Vitis vinifera*, *A. thaliana*, *P. persica*, *Malus × domestica, C. serrulata* and haplotype A of *C.× yedoensis 'Somei Yoshino'*. (b). Phylogenetic tree for PCP and the other seven plants. The red numbers outside the square brackets indicate the average divergence time, while the red numbers inside the square brackets indicate the 95% confidence interval of divergence time. The blue numbers indicate expanded gene families, while the green numbers indicate contracted gene families. (c). Functional enrichment of specific genes in flowering cherry. (d). Functional enrichment of expanded genes in flowering cherry.

Figure 4. Population analysis of the 312 accessions of flowering cherry germplasm.
(a). Phylogenetic tree of the 312 flowering cherries (including six outgroup lines). The branch colours represent the different clades. Red, green and blue branches represent clades A, B and C, respectively. The black branch represents outgroups. The symbols at the end of branches represent the individual's category. The square, circle and star represent breeding lines, wild individuals and cultivars, respectively. The colours of the outermost rectangle represent different species of all 312 flowering cherries. (b). Principal component analysis (PCA) of the 312 flowering cherries. Red, green, blue and black dots represent clades A, B, and C and outgroups, respectively. (c). Population structure analysis of 306 flowering cherries. (d). Subpopulations (Fig. 4c) in the 3 clades (Fig. 4a). (e). Population differentiation and nucleotide diversities for clades A, B and C. (f). linkage disequilibrium (LD) decay for clades A, B and C. The Y-axis indicates r^2, while the x-axis indicates distance (kb).

Figure 5. Selective sweep analysis among clades A, B and C

(a), (c) and (e) illustrate the cross-population composite likelihood ratio (XP-CLR) of clades A/B, A/C and B/C, respectively. Genes located in the top 30 differentiated genomic regions for each comparison are labelled above these peaks. (b), (d) and (f) show functional enrichment of genes in the highly differentiated genomic regions (genomic regions with top 5% XP-CLR values).

Figure 6. Genome-wide association study analysis of flower colour variations for the 312 lines of flowering cherry germplasm

(a). Manhattan plot for flower colour variations and identification of the candidate gene within QTL Chr02: 9426524-9872586. (b). The quantile–quantile (Q-Q) plot of all the P values in (a). (c). Flower colour variation analysis of the three genotypes classified by *evm.model.LG02.1464*. Multiple comparisons of the mean colour intensities among the three genotypes GG, GT and TT at P values equal to 0.01 (differences marked by different letters A, B and C).

Supplementary data

Fig. S1. K-mer distribution of the *Prunus campanulata* 'Plena' (PCP) genome

Fig. S2. Collinearity comparisons between the PCP and *C. serrulata* genomes

Fig. S3. Collinearity comparisons between the PCP and *P. yedoensis* var. *nudiflora* genomes

Fig. S4. Determination of subpopulation number for flowering cherry germplasms

Fig. S5. Detailed information for the phylogenetic tree of the 312 flowering cherry germplasm lines

Table S1. Summary of sequence data used for genome assembly

Table S2. Comparisons of three genome assemblies in the subgenus Cerasus

Table S3. Statistics of genome coverage with Nanopore long reads

Table S4. Genome sizes of 8 chromosomes in PCP

Table S5. TE repeat sequence statistics

Table S6. Statistics of protein-coding gene annotation

Table S7. Summary of genes annotated with the Nr, TAIR, Swiss-Prot, KEGG and InterPro databases

Table S8. Statistics of noncoding RNA annotation

Table S9. Collinear blocks between the genomes of *Cerasus serrulata* and PCP

Table S10. Collinear blocks between the genomes of *P. yedoensis var. nudiflora* and PCP

Table S11. Information on the 312 accessions of flowering cherry used for resequencing

Table S12. Classification of genomic variants for 312 flowering cherry germplasms

Table S13. The population structure of the 312 accessions of flowering cherries

Table S14. Genes located in the top 5% differentiated genomic regions between clades A and B

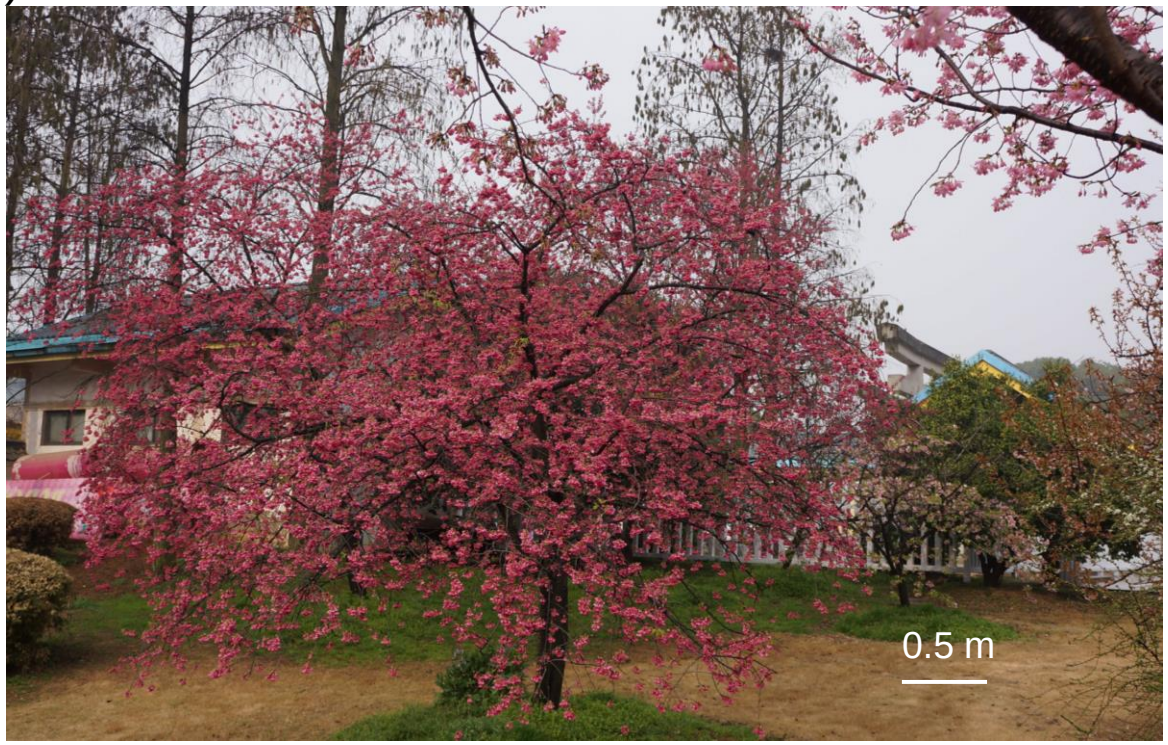
Table S15. Genes located in the top 5% differentiated genomic regions between clades A and C.

Table S16. Genes located in the top 5% differentiated genomic regions between clades A and C.

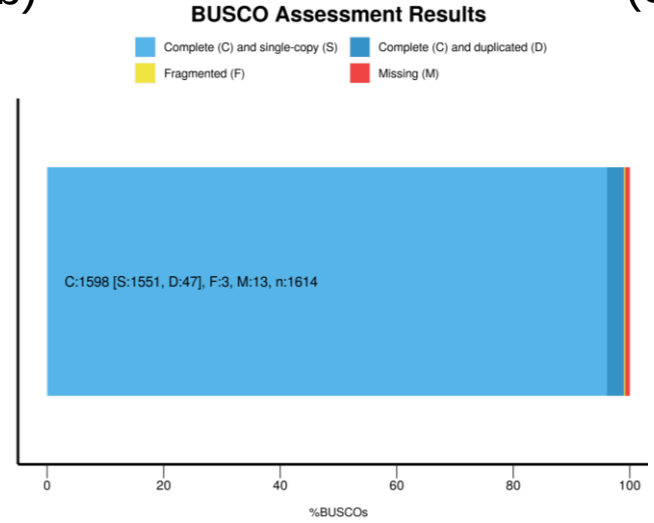
Table S17. Pigment components of 3 flowering cherry cultivars

Table S18. QTLs for flower colour variations in the 312 flowering cherry germplasm lines

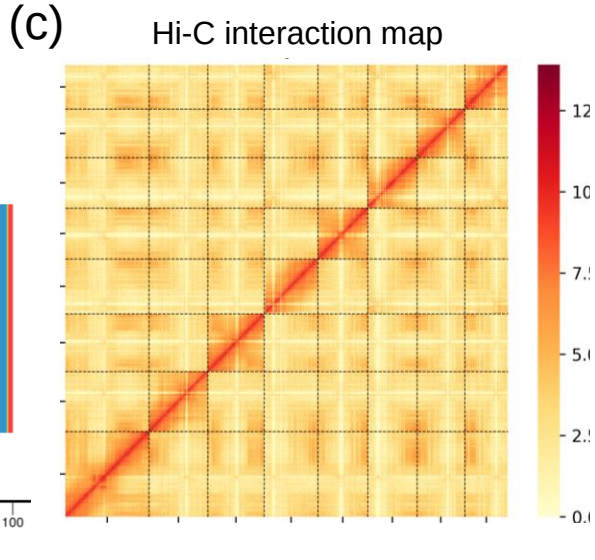
(a)



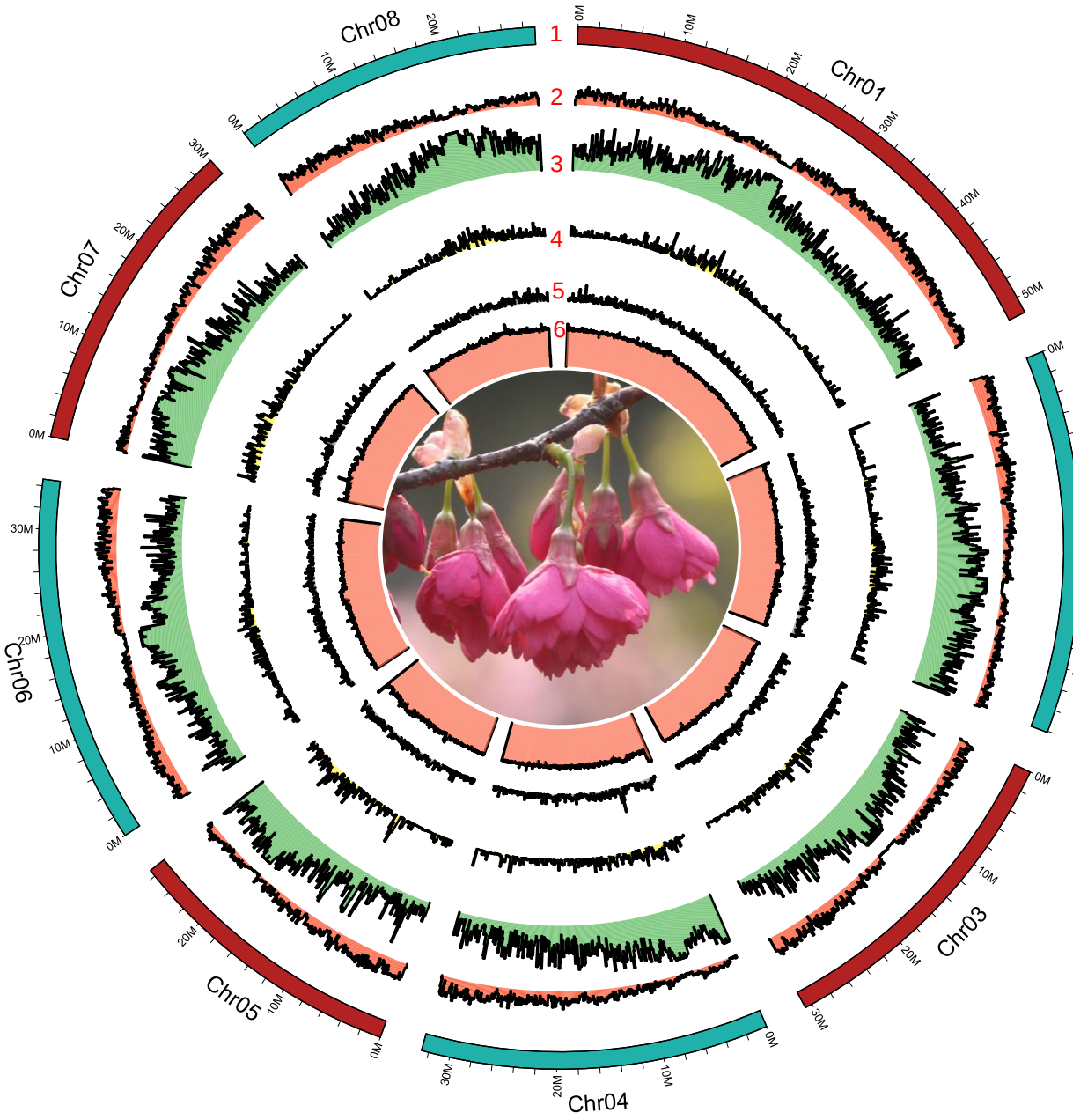
(b)



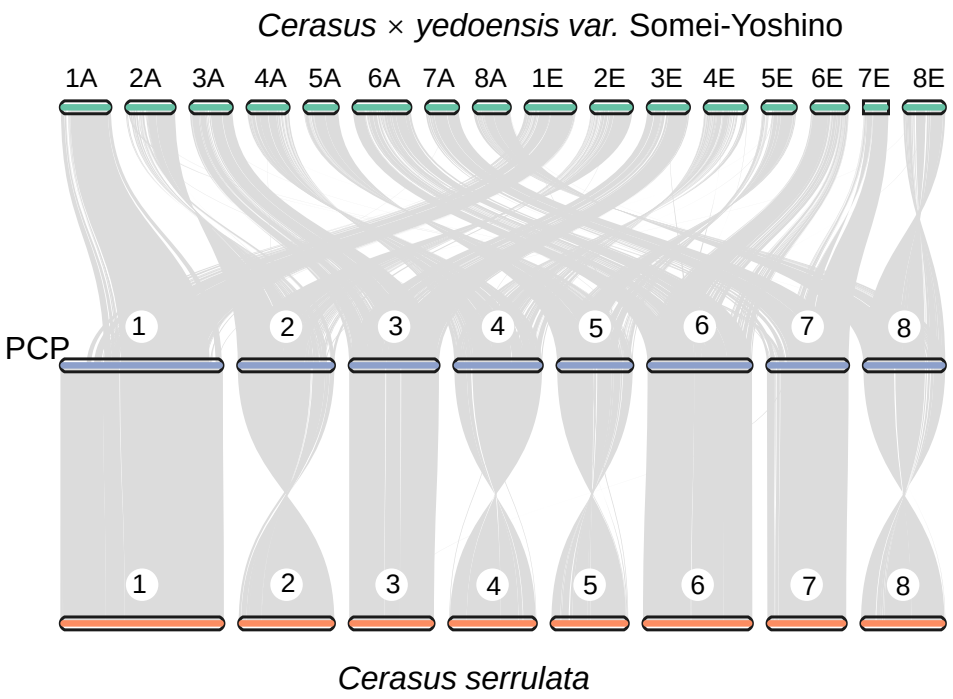
(c)



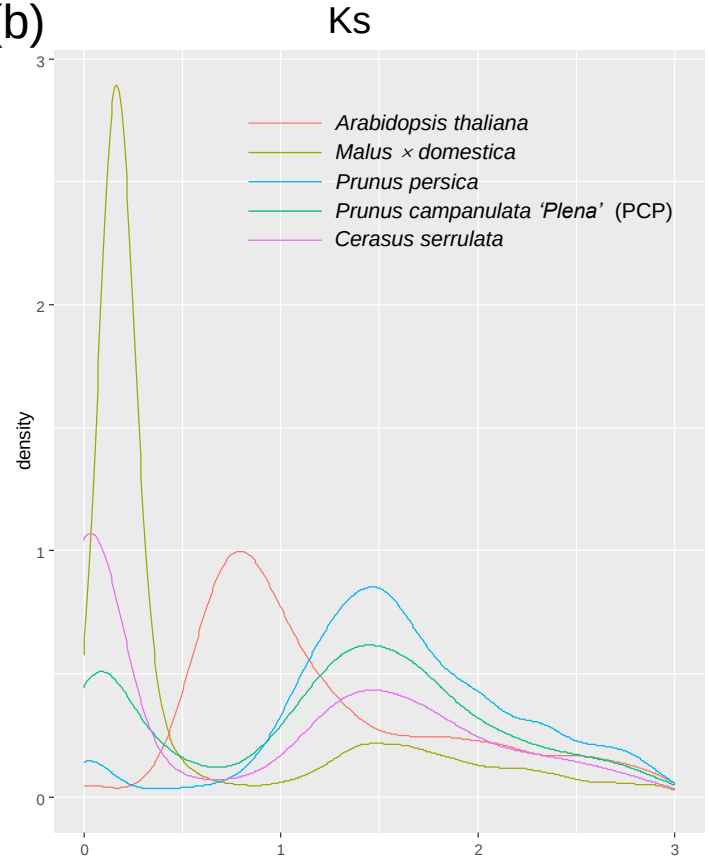
(d)



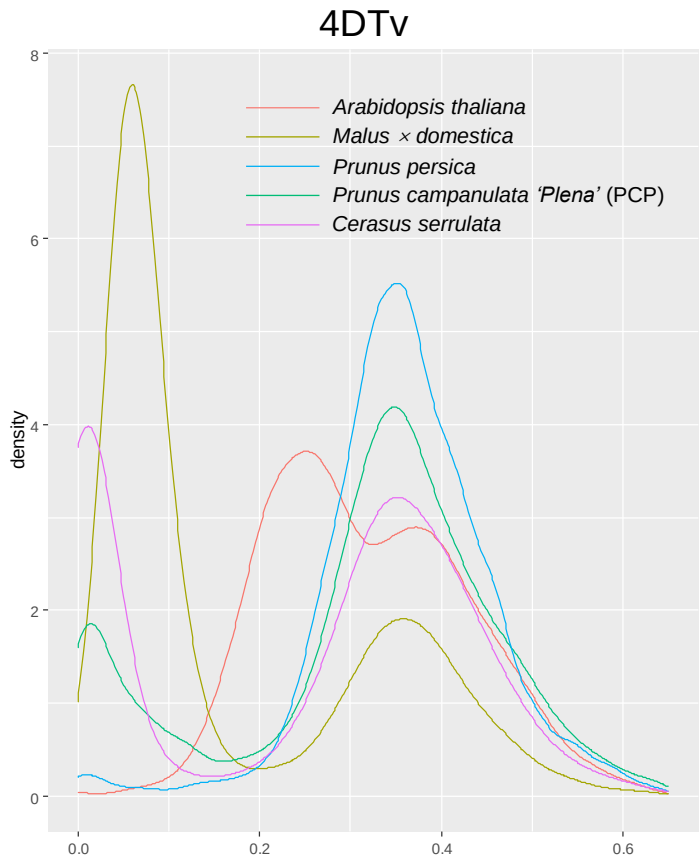
(a)



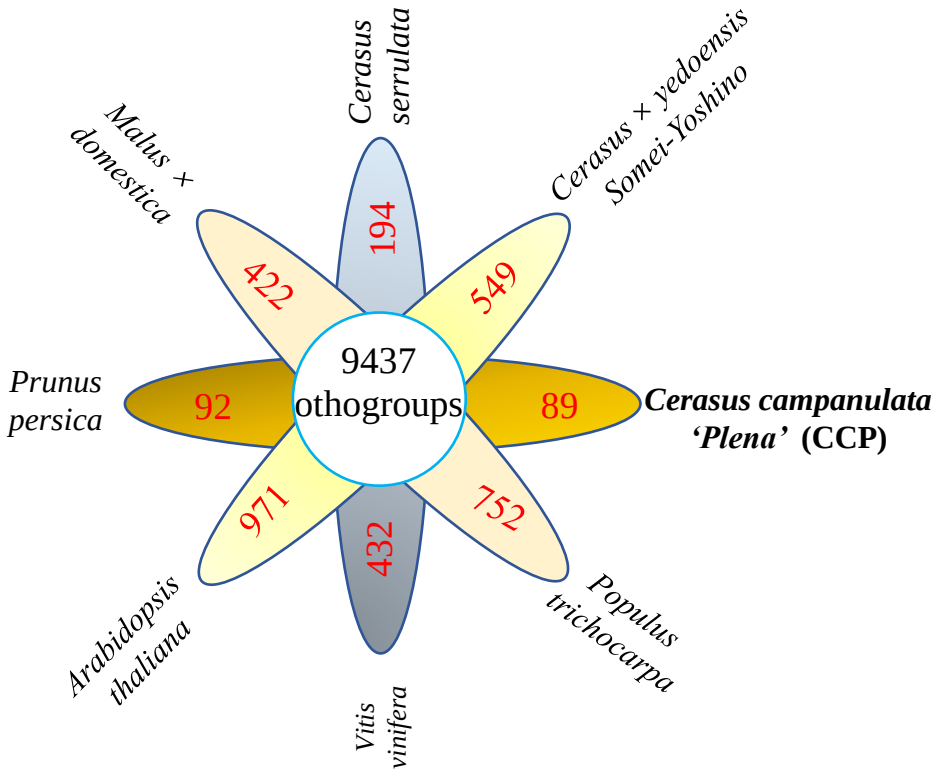
(b)



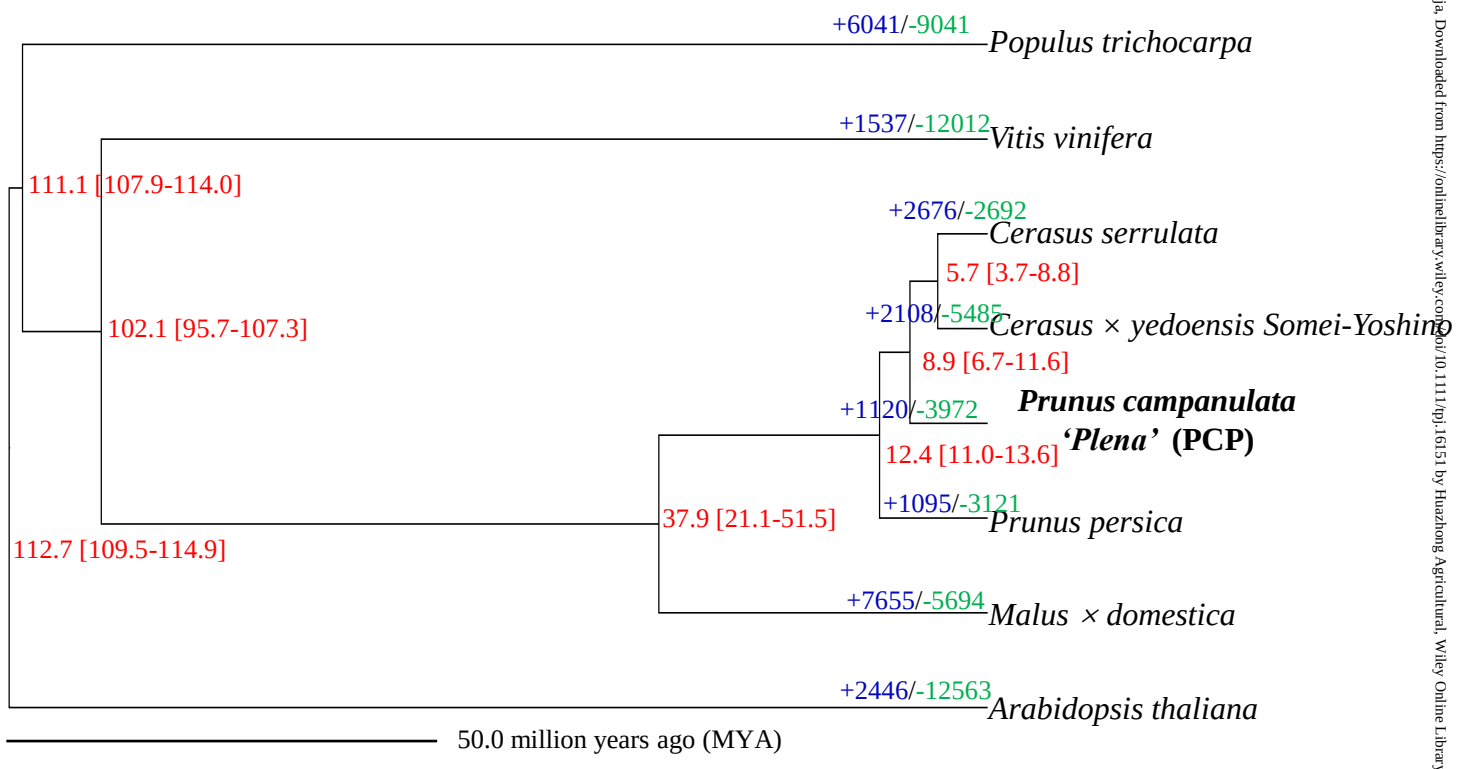
(c)



(a)

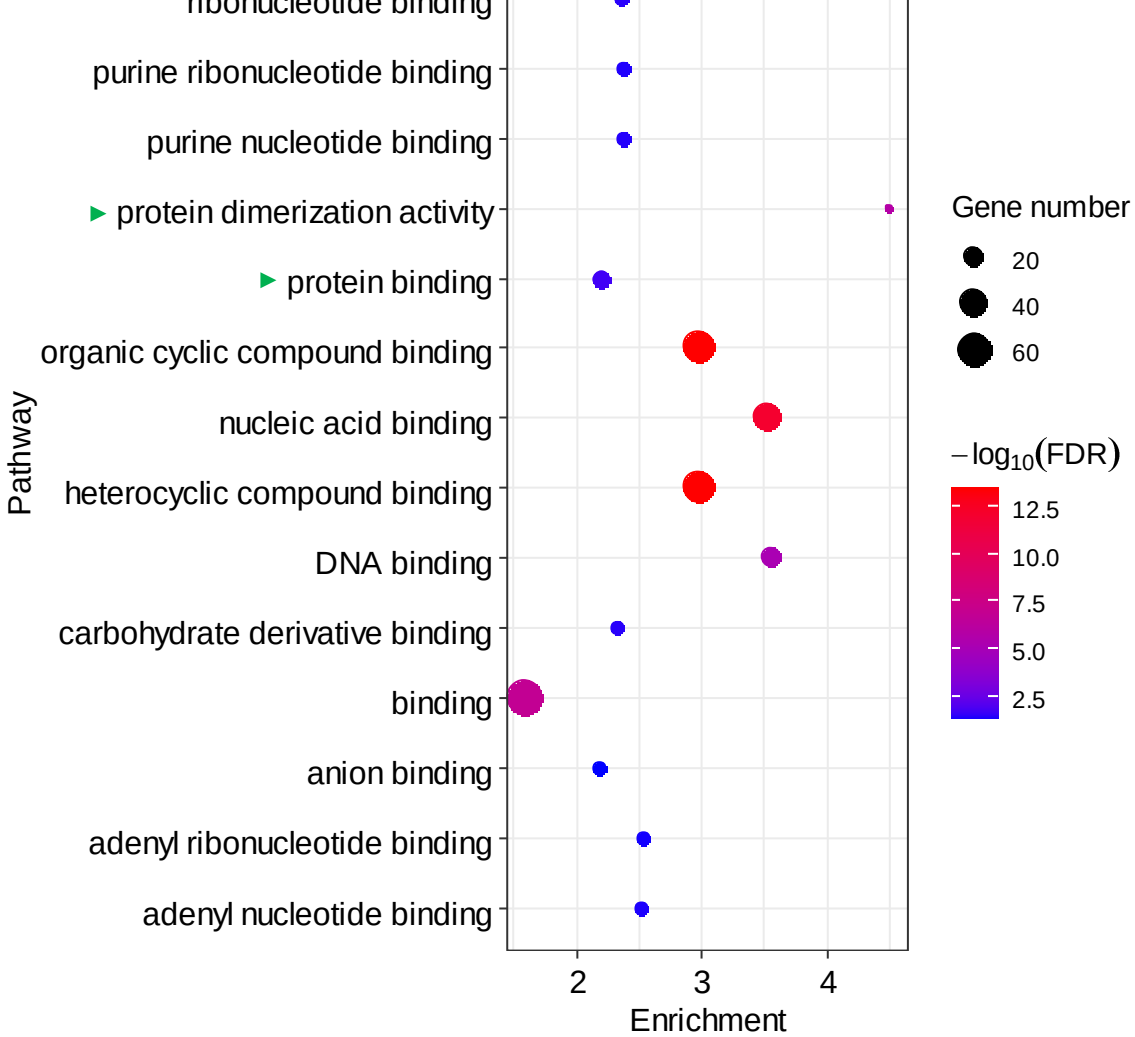


(b)



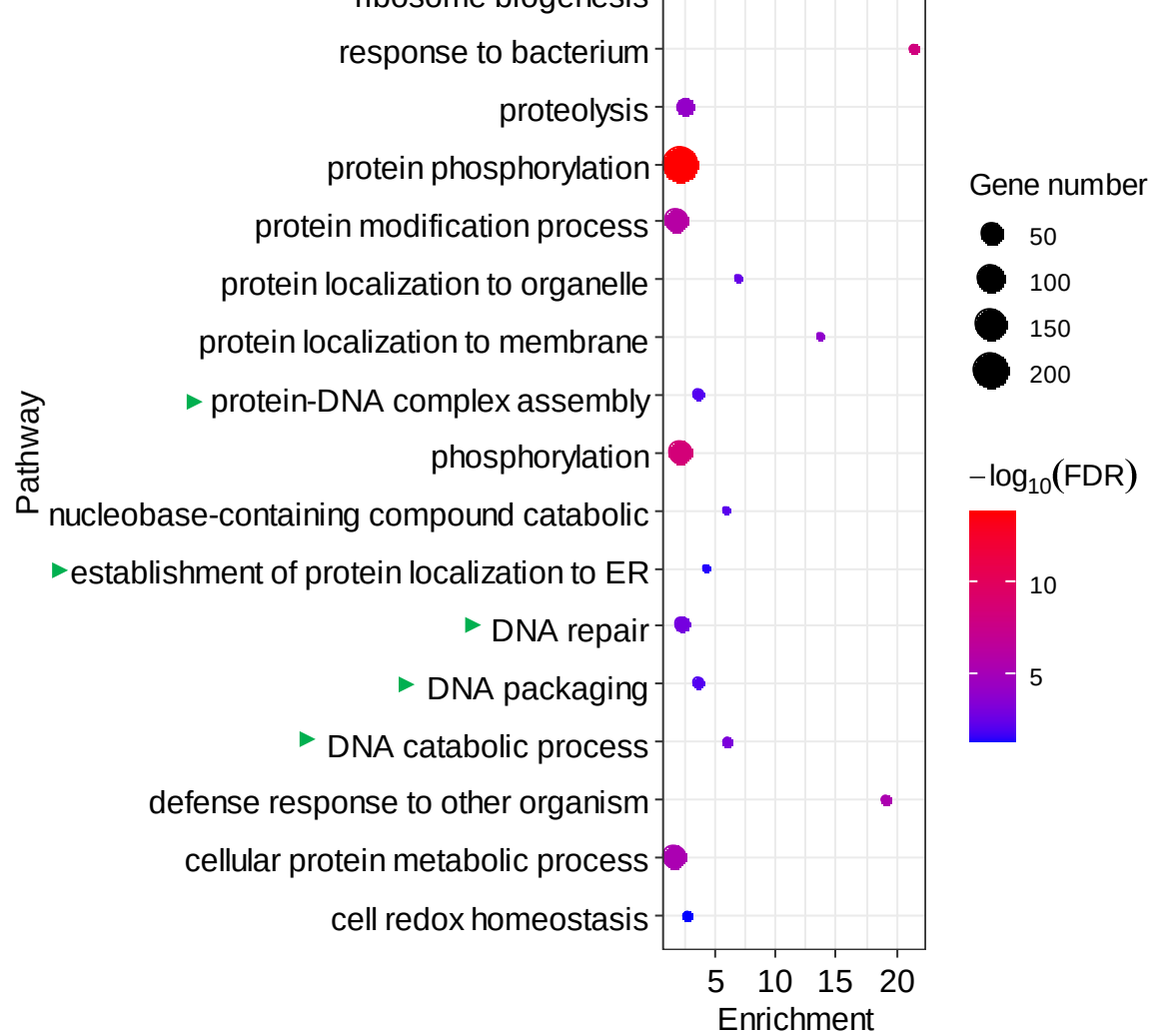
(c)

Functional enrichment of specific genes in flowering cherry

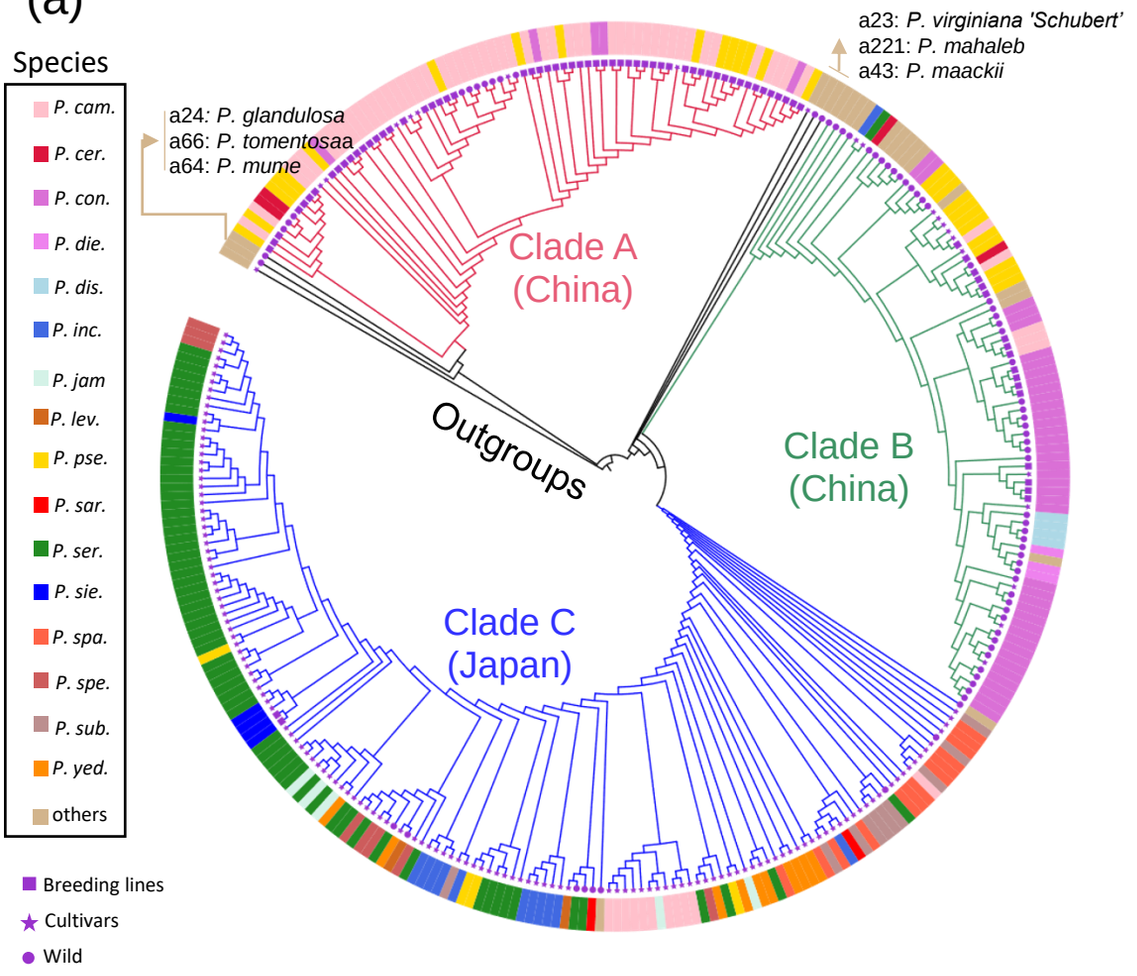


(d)

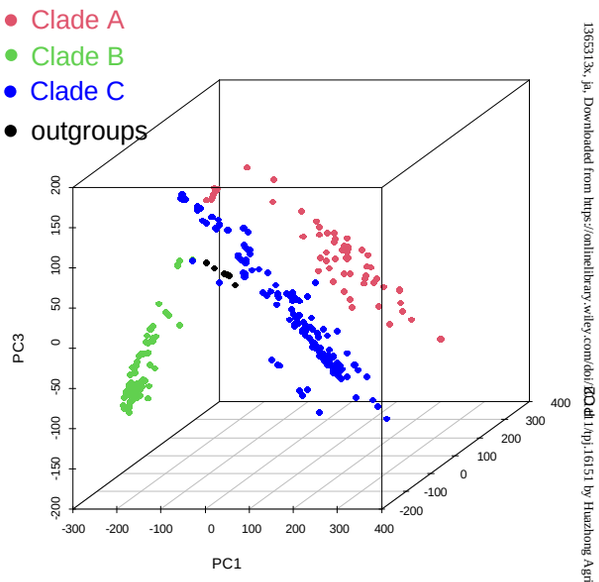
Functional enrichment of expanded genes in flowering cherry



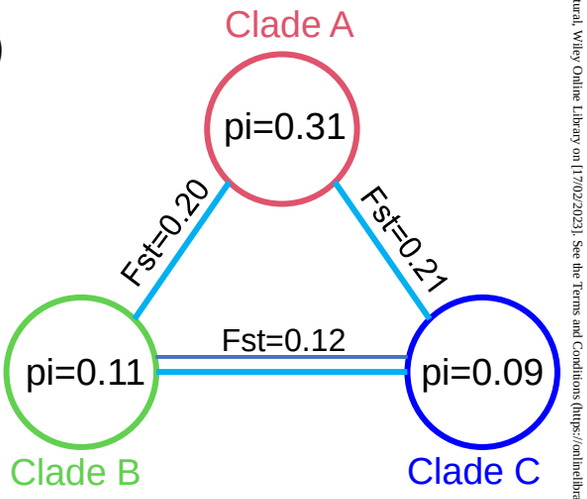
(a)



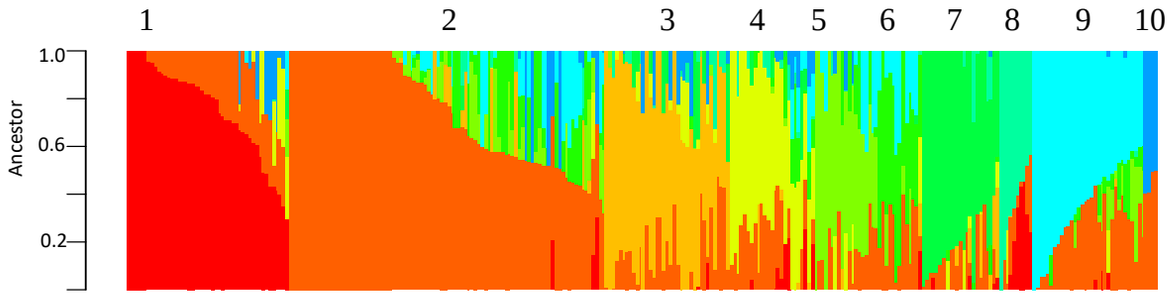
(b)



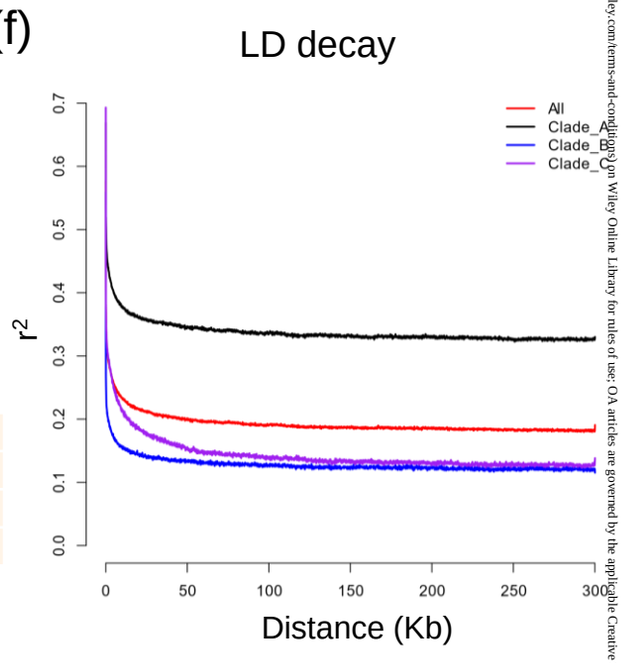
(e)



(c)



(f)

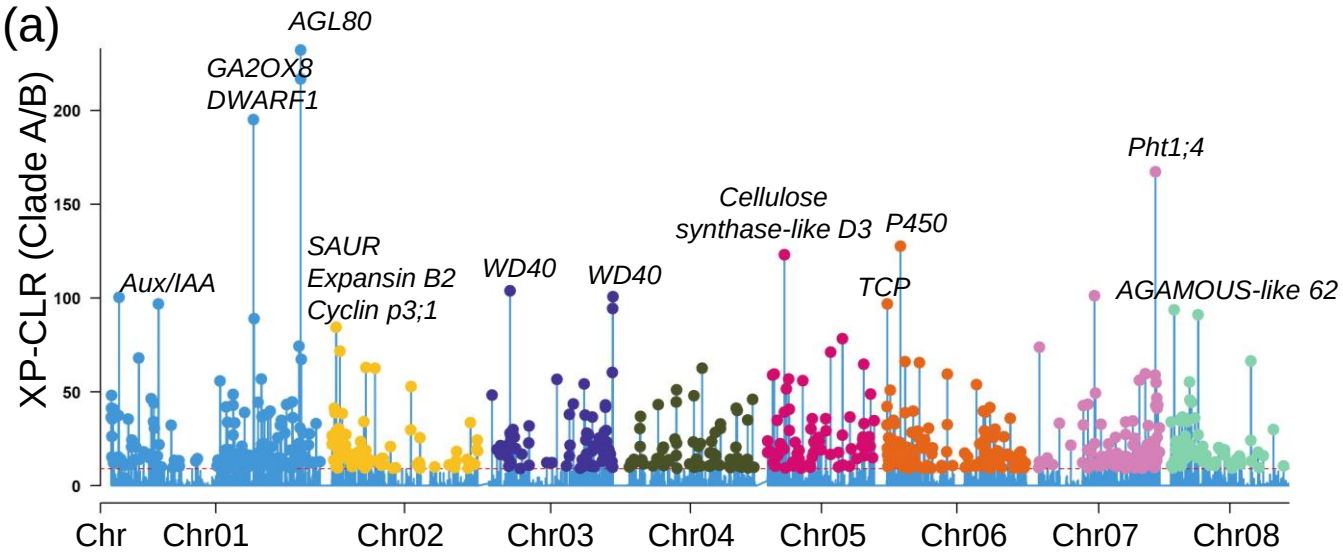


(d)

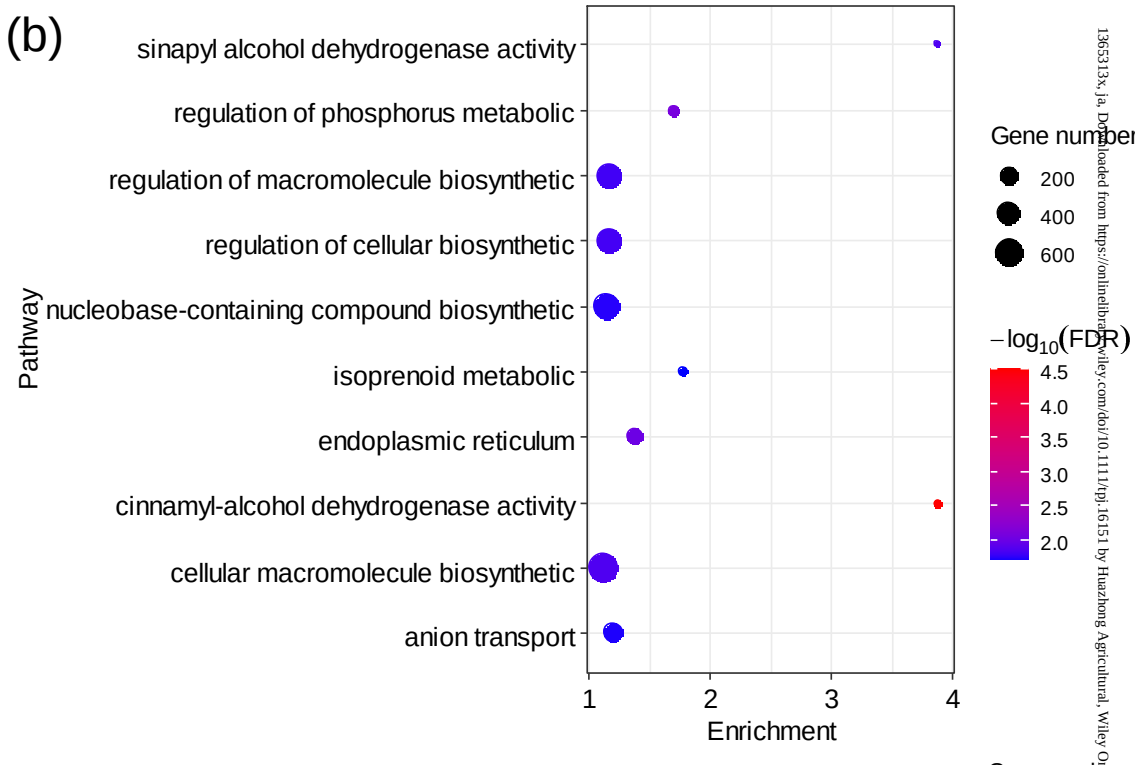
Sub-populations in 3 clades

	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Pop8	Pop9	Pop10
Clade A	47	6	0	8	0	0	0	9	0	4
Clade B	0	12	37	10	0	0	23	0	0	0
Clade C	1	76	0	7	19	13	0	1	33	0

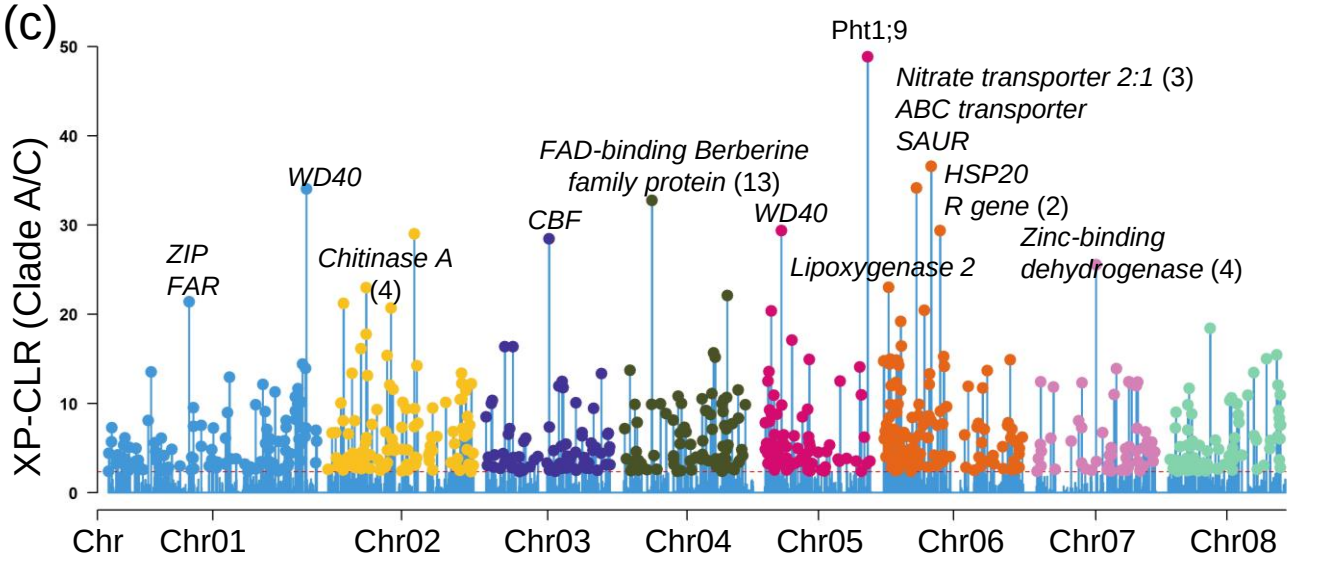
(a)



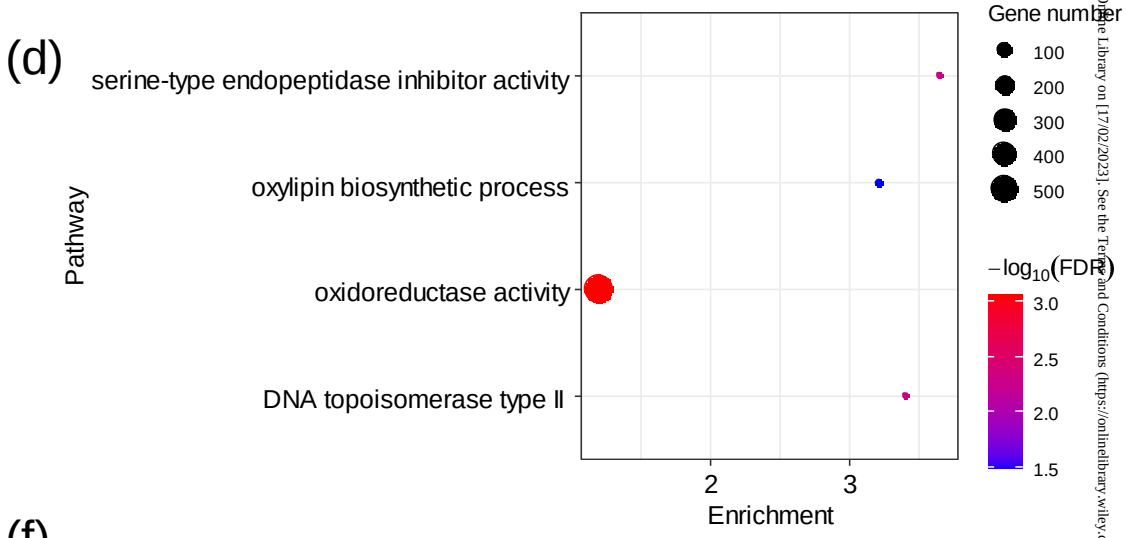
(b)



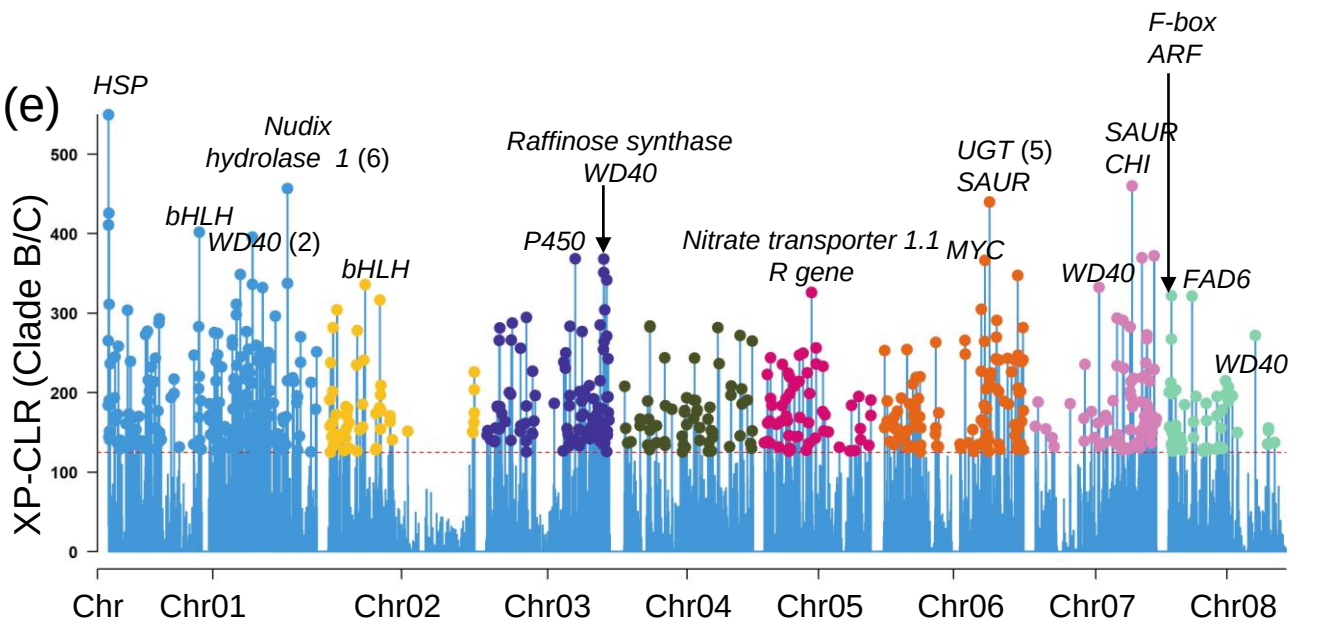
(c)



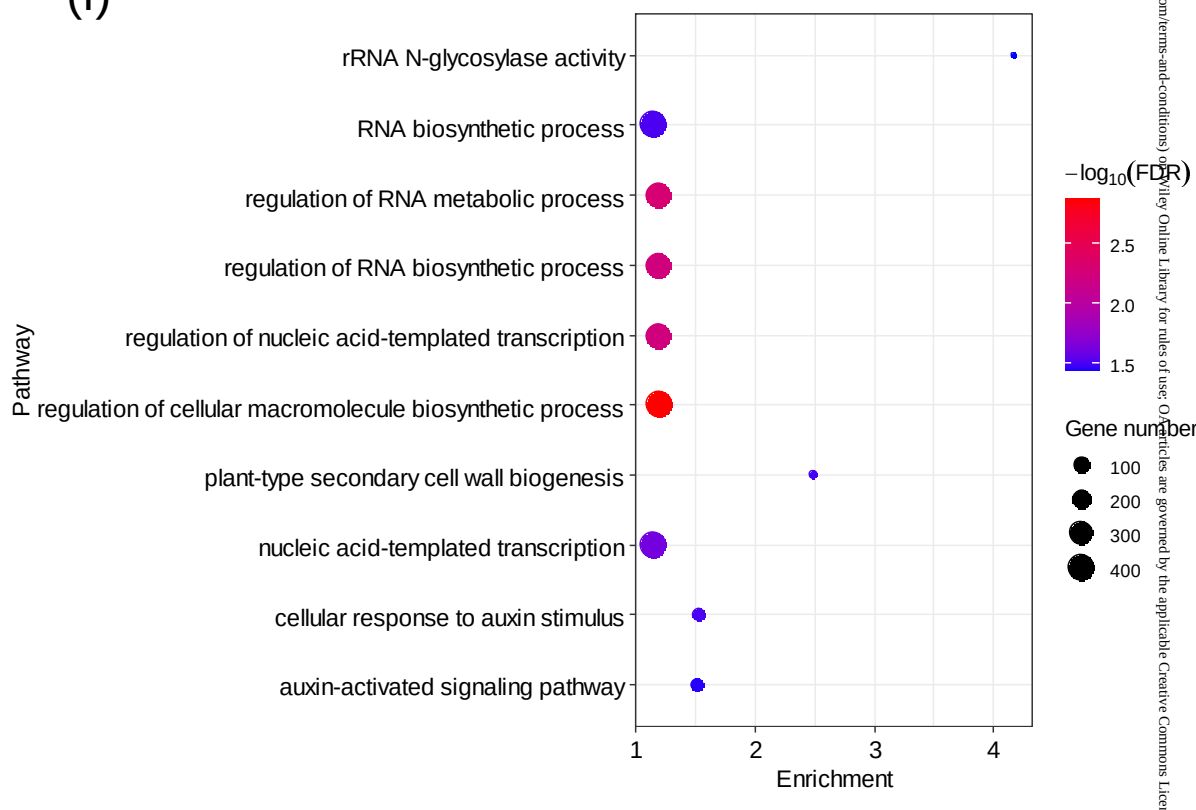
(d)



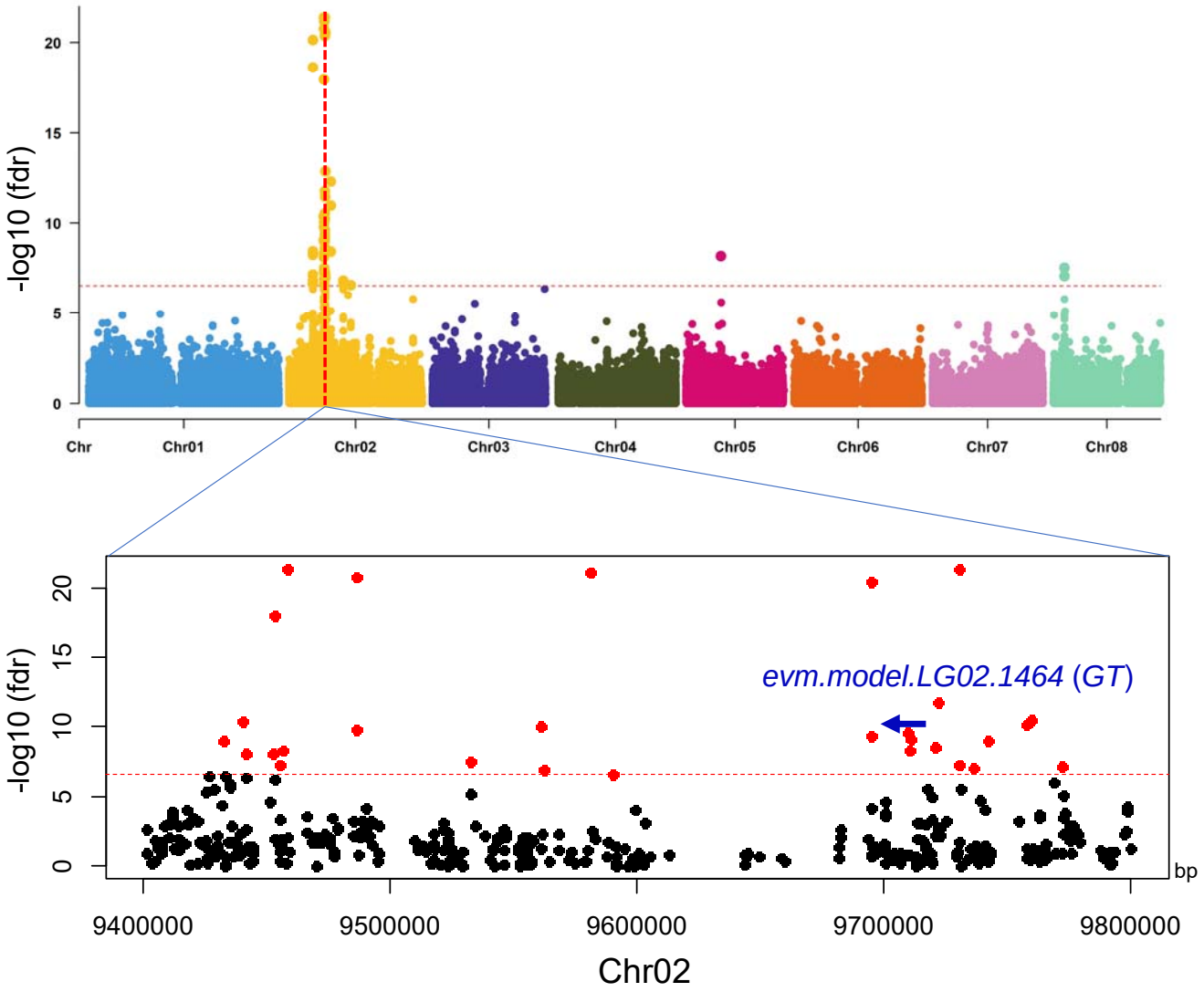
(e)



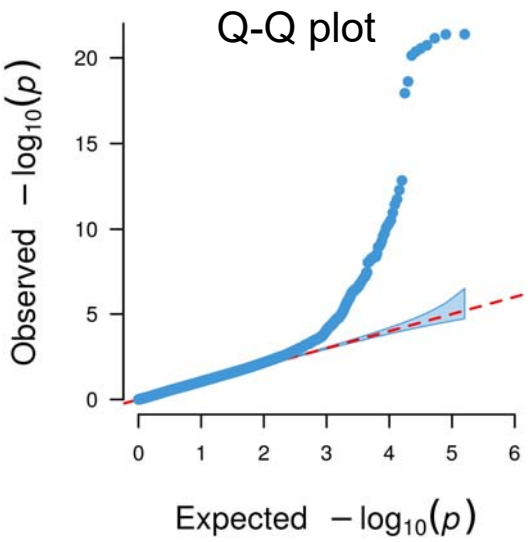
(f)



(a)



(b)



(c)