Comparative population genomics reveals convergent and divergent selection in the Apricot-Peach-Plum-Mei Complex

Xuanwen Yang[1, 2, 3, 5 #], Ying Su[3, 6, #], Siyang Huang[3, #], Qiandong Hou[1], Pengcheng Wei[2], Yani Hao[3, 7], Jiaqi Huang[3, 8], Hua Xiao[3], Zhiyao Ma[3], Xiaodong Xu[3], Xu Wang[3], Shuo Cao[3, 5], Xuejing Cao[3], Mengyan Zhang[3], Xiaopeng Wen[1], Yuhua Ma[4], Yanling Peng[3], Yongfeng Zhou[3, 9, *], Ke Cao[2, *], Guang Qiao[1, *]

[1] Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Institute of Agro-bioengineering/College of Life Sciences, Guizhou University, Guiyang 550025, China

[2] National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Science, Zhengzhou 450009, China

[3] National Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518120, China

[4] Institute of Pomology Science, Guizhou Academy of Agricultural Sciences, Guiyang 550006, China

[5] College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan, 430070, China

[6] Xinjiang Key Laboratory of Biological Resources and Genetic Engineering, College of Life Science and Technology, Xinjiang University, Xinjiang, Urumqi 830046, China

[7] Department of Bioinformatics, School of Biology and Basic Medical Sciences, Suzhou Medical College of Soochow University, Suzhou, 215123, China

[8] College of Life Sciences, Wuhan University, Wuhan, 430072, China

[9] National Key Laboratory of Tropical Crop Breeding, Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, China

[#] These authors contributed equally to the work.

[*] Correspondence authors: Guang Qiao, gqiao@gzu.edu.cn; Ke Cao, wyandck@126.com; Yongfeng Zhou, zhouyongfeng@caas.cn

Abstract

The economically significant genus *Prunus* includes fruit and nut crops that have been domesticated for shared and specific agronomic traits, however, the genomic signals of convergent and divergent selection have not been elucidated. In this study, we aim to detect genomic signatures of convergent and divergent selection by conducting comparative population genomic analyses of the Apricot-Peach-Plum-Mei (APPM) complex, utilizing a haplotype-resolved telomere-to-telomere (T2T) genome assembly and population resequencing data. The haplotype-resolved T2T reference genome for the plum cultivar was assembled through HiFi and Hi-C reads, resulting in two haplotypes with 251.25 Mb and 251.29 Mb in size, respectively. Comparative genomics reveals a chromosomal translocation of approximately 1.17 Mb in the apricot genomes compared to peach, plum, and mei. Notably, the translocation involves the *D* locus, significantly impacting acidity (TA), pH, and sugar content. Population genetic analysis detected substantial gene flow between plum and apricot, with introgression regions enriched in post-embryonic development and pollen germination processes. Comparative population genetic analyses revealed convergent selection for stresses, flower development, and fruit ripening, along with divergent selection shaping crop-specific genes, such as somatic embryogenesis in plum, pollen germination in mei, and hormone regulation in peach. Notably, selective sweeps on chromosome 7 coincide with a chromosomal co-linearity from the comparative genomics, impacting key fruit-softening genes such as *PG*, regulated by *ERF* and *RMA1H1*. Overall, this study provides insights into the genetic diversity, evolutionary history, and domestication of the APPM complex, offering valuable implications for genetic studies and breeding programs of *Prunus* crops.

Keywords: *Prunus* species, introgression, structural variation, convergent evolution, T2T genome

Introduction

Species were predominantly defined by reproductive isolation, as physiological and mating incompatibilities[1]. Nonetheless, many species are known to hybridize occasionally in captivity and the wild, accompanied by introgression (the acquisition of genetic variation from another species[2]), and recombinational speciation (today known as homoploid hybrid speciation). Phenotypic changes are primarily attributed to natural mutations, with occasional interspecific introgression[3,4]. Though identifying genomic introgression is challenging, a wealth of genomic data confirms the occurrence of frequent gene flow between species, such as wheat[5], some of which are adaptive[6–10]. For instance, the domesticated grapevine (*Vitis Vinifera*) has undergone additional genomic introgression from its wild relatives, resulting in enhanced fruit aroma and increased disease resistance[11]. Genomic regions with high recombination rates are more susceptible to introgression than regions with lower recombination rates[12], due to the effects of selection on sites linked to introgressed alleles (i.e., linked selection)[13,14].

Such selection signals were mostly specific to given populations/species (divergent selection), occasionally, it could also be shared among populations/species (convergent selection). Convergent selection takes place when diverse species consistently encounter comparable selective pressures in their environments[15–17], whereas divergent selection results in trait divergence within a single lineage due to diverse selective pressures[18]. If a particular genomic region has positive (i.e., selective sweep[19]) or negative (i.e., purifying selection[20]) effects, both the region itself and neighboring sequences will show reduced genetic diversity. Selective sweeps identified in domesticated populations, including maize[21], cattle[22], grapes[23], and citrus[24] have enhanced our understanding of the genetic architecture and artificial selection during the domestication history of agronomic traits[25,26].

Prunus species are economically important plants in the subfamily Amygdaloideae of the Rosaceae[27] with genome sizes of approximately 230 – 280 Mb[28–30]. The genus includes a range of economically important flower, fruit, and nut crops[31,32], in which apricot, peach, plum, and mei are closely related species with some levels of hybrid and grafting compatibilities among them. Historical gene flow was also reported, forming the Apricot-Peach-Plum-Mei (APPM) complex(2n = 2x = 16). The APPM complex includes four fruit tree species that share a more recent common ancestry within its constituent species compared to cherries[28–30]. Nevertheless, despite its historical significance and traditional classification, our knowledge of the genetic background of this species remains very limited.

P. mume, known as Chinese plum, mei, mume, Japanese plum, and Japanese apricot, is closely related to the apricot[33–37]. *P. brigantina*, resembling apricot in texture but with smooth skin like plums, is a wild species[38,39]. *P. cerasifera* is crucial in polyploid breeding and is one of the parent species in the origin of *P. domestica*[40]. Furthermore, recent breeding programs often utilize interspecific crossing of the subgenus *Prunus*, such as 'Plumcot' (*P. salicina* × *P. armeniaca*)[41,42] or 'Peacotum' (*P. persica*× *P. armeniaca* × *P. salicina*)[43]. During the differentiation of these three species within the genus *Prunus* (*P. mume*, *P. armeniaca*, and *P. salicina*), several hybridization events may have occurred, resulting in significant introgressions associated with unidentified phenotypic changes during the formation of the common ancestor of *P. mume*[35]. The population genetics in the APPM complex can provide a deeper understanding of their convergent and divergent selection and domestication history[32,44], which is vital for the process of *Prunus* breeding. However, the genomic signatures of convergent and divergent selection of favorable traits in these crops remain poorly understood.

To further explain the evolutionary history of *Prunus*, potential introgressions, and encompassing convergent and divergent selection, we explored comparative genomics and population genomics across different *Prunus* species, including apricot, peach, plum, and mei. Furthermore, we explored the domestication genomics of the APPM complex, thereby contributing novel perspectives to the realm of *Prunus* breeding strategies.

Results

The assembly of haplotype-resolved T2T reference genome of plum

We generated a total of 26.84 Gb (~100× coverage) of raw PacBio high-fidelity (HiFi) reads and 30 Gb (~120× coverage) of chromosome conformation capture sequencing (Hi-C) data for assembling the plum (*P. salicina* cv. 'Fengtangli') genome (**Figure 1A-C, Table S1**). Before starting the assembly, we estimated the genome heterozygosity to be 0.97% using a *k*-mer based approach **(Supplementary Data Fig. S1)**. For the initial assembly, the N50 value of contig-level for haplotype 1 and haplotype 2 were 20.11 Mb and 19.66 Mb, respectively, about 14 times that of the *P. salicina* cv. 'Sanyueli' contig-level assembly, with the largest contig reaching a length of 52.18 Mb. After anchoring and ordering, the scaffold N50 sizes reached 28.08 Mb and 28.37 Mb. The Benchmarking universal single-copy orthologs (BUSCOs) assessment revealed that the conservative plant core genes are nearly fully captured at 99.2% for haplotype 1 and 98.9% for haplotype 2, marking an improvement from

the previous rate of 98.64% **(Table 1, Supplementary Data Fig. S1)**. 13 and 15 gaps were identified after initial assembly into scaffold among two haplotypes **(Supplementary Data Table S2)**. By mapping the HiFi data to each of the two haplotypes, we manually filled in all the gaps. Finally, we assembled two gap-free haplotypes of the *P. salicina* cv. 'Fengtangli', named PS_T2T_hap1 (251.25 Mb) and PS_T2T_hap2 (251.29 Mb) **(Figure 1D, Table 1)**. The visualization of Hi-C data using Juicebox[45] demonstrated a high level of consistency across all chromosomes for both haplotypes, proving their accuracy of ordering and orientation (**Figure. 1E).** By remapping the HiFi and Hi-C reads to their respective PS_T2T assemblies, we achieved a mapping ratio of over 98.5%, emphasizing the accuracy and completeness of our genome assembly (**Supplementary Data Table S1**). In comparison, the quality of PS_T2T is similar to those published complete genomes, including maize[46], rice[47], grapes[48,49], pear[50], and kiwifruit[51]. These high-quality genomes provide the opportunity to study the evolutionary genomics among *Prunus* species.

Centromeres are typically composed of repetitive DNA sequences, with variations among different organisms[52]. Tandem Repeats Finder (TRF)[53] found 470 different repeat units in the PS_T2T assemblies. Finally, the 166 bp repeats were the most abundant unit in the whole genome. The transposable elements (TEs) analyses also supported the centromeric feature of 166 bp repetitive unit **(Supplementary Data Fig. S2)**. Therefore, the centromeres are primarily identified based on the 166 bp repeats and are located on all 16 chromosomes of the two haplotypes with lengths ranging from 1.02 to 3.61 Mb in PS_T2T_hap1, 1.88 to 3.36 Mb in PS_T2T_hap2 **(Supplementary Data Table S3)**. Telomere identification was performed by searching the telomere sequences. They consist of repetitive DNA sequences, such as 'CCCATTT' at the 5' end and 'TTTAGGG' at the 3' end in plants. 24 of the expected 32 telomeres (8 chromosomes of two haplotypes) were identified, and 11 and 13 telomeres were found in PS_T2T_hap1 and PS_T2T_hap2 **(Figure 1D, Supplementary Data Fig. S3, Table S4).** Our PS_T2T assemblies filled in the absence of centromere and telomere assembly of the reference genome 'Sanyueli'. Overall, the above results confirm a high-quality genome assembly.

An Extensive de novo TE Annotator (EDTA)[54] was used to generate a high-quality repetitive sequence library, and identified 119.63 and 121.71 Mb of repetitive sequences in the two haplotypes, accounting for 45.41% and 46.19% of PS_T2T_hap1 and PS_T2T_hap2, respectively. For haplotype 1, LTR retrotransposons account for 29.57%, which contained 18.07% Gypsy and 4.50% Copia LTR elements **(Supplementary Data Table S5**). For gene annotation, 28,775 and 28,139 protein-coding genes were predicted for two haplotypes.

BUSCOs assessment using the longest transcribed proteins revealed that two haplotypes captured 96.6% and 94.0% of a BUSCOs reference gene set, respectively. Moreover, 40,270 and 39,205 transcripts were predicted with an average of 1.39 splice variants per gene. Additionally, 48,645 shared genes were obtained, including 24,542 from PS_T2T_hap1 and 24,103 from PS_T2T_hap2, belonging to 20,323 orthologous gene families, representing a core set of gene clusters in the PS_T2T genome **(Supplementary Data Table S6)**. Furthermore, 876 genes were unique to PS_T2T_hap1, and 760 genes were unique to PS_T2T_hap2, likely reflecting genetic distinctions between the two parental strains **(Supplementary Data Table S6**). In conclusion, the depth of our sequencing and retention of more sequences resulted in a more comprehensive and accurate assembly of the diploid genome for the first time, representing a substantial advancement in plum genomics.

Comparative genomics of the Apricot-Peach-Plum-Mei (APPM) species

To further verify the quality of our assemblies, comparative genomics analyses were performed with two haplotypes using *P. salicina* cv. 'Sanyueli' as the reference genome **(Supplementary Data Fig. S4)**. The results showed good collinearity in most regions of the genome. Non-collinearity might be caused by poor sequence quality, low coverage, or misassembly in the previous genome, but it could also be genuinely present, especially across the whole genome scale. With a more complete diploid assembly of *P. salicina* cv. 'Fengtangli', we performed variants analysis between the two haplotypes. The comparison between the two haplotypes of PS_T2T indicates that they share a set of similar genomic features, including closed genome size, parallel repetitive content, and a comparable number of genes. However, a significant amount of variation is still observed between the two haplotypes, including 1,667,700 single nucleotide polymorphism (SNPs), around 389,962 insertions and deletions (InDels, below 50 bp), and 30,698 structural variants (SVs,50 bp or longer[26,55]). This establishes a substantial genetic diversity repository of plum **(Supplementary Data Table S7, Fig. S5)**. Some SVs, such as inversions, deletions, insertions, and translocations were rarely observed in the newly assembled centromeric regions. This could be attributed to the presence of more conserved repeats and greater sequence stability within the highly repetitive mitotic satellites, as well as the lower diversity in mitotic patterns among haplotypes.

To study the evolutionary conservation within the *Prunus* genus genome, we selected high-quality genomes (long-read based) from four species: apricot (*P. armeniaca*)[56], peach (*P.*

persica)[57], plum (*P. salicina,* this study), and mei (*P. mume*)[58] for genome collinear comparison **(Supplementary Data Table S8)**. For the plum, the two haplotypes we assembled exhibit a high degree of collinearity, therefore we chose PS_T2T_hap1 as a representative of the plum species. A strong collinear relationship exists among the four *Prunus* species, indicating that our pseudo-chromosomes derived from anchored and oriented contigs were of high quality **(Figure 2A)**. For instance, chromosome 7 of apricot demonstrated significant collinearity with chromosome 7 in peach, chromosome 8 in plum, and chromosome 8 in mei.

A large structural rearrangement in the form of translocations was characterized among chromosome 7 in apricot compared with the others **(Figure 2A, Supplementary Data Table S9)**. However, an SV was detected in the apricot in the form of a translocation spanning the 5.10 - 6.27 Mb (1.17 Mb) region of the long arm of chromosome 7. This particular segment corresponded with chromosome 5 in peach, chromosome 7 in plum, and chromosome 7 in mei **(Figure 2B, Supplementary Data Table S10)**. Gene Ontology (GO) enrichment analysis indicates that the 1,223 genes located in the translocation exhibit significant enrichment in various biological processes, including regulation of shoot system development, protein glycosylation, and innate immune response **(Figure 2C, Supplementary Data Table S11)**. At the breakpoint of the translocation, both genes on either side may be affected. These genes have been identified as *ABC1K3*, which is essential for the photo-oxidative stress response[59], and *GSTF13*, which plays a key role in the detoxification of certain herbicides[60] **(Figure 2B)**. We also highlighted the gene rearrangement events in the translocation region and identified a very important *D* locus associated with fruit flavor at the translocation of chromosome 5 in peach **(Figure 2B)**.

Additionally, a gene family cluster analysis was performed on the genomes of apricot, peach, plum, and mei. The results revealed that these four *Prunus* species collectively shared 14,651 gene families, and plum shared 2,466 gene families with mei, 2,862 with apricot, and 1,896 with peach, with core genes being excluded from this analysis (**Figure 2D)**. Of these, 130 gene families, comprising 788 genes, were exclusively identified in the plum genome. The count is lower than the private genes noted in apricot (3,450 genes across 619 gene families) and mei (1,471 genes across 350 gene families) but exceeds that of peach, which had 299 genes within 107 distinct gene families (**Figure 2D, Supplementary Data Table S12)**. GO enrichment analysis showed that plum-specific genes were significantly enriched in biological processes such as signal transduction, receptor-mediated endocytosis, and defense response.

Apricot-specific genes are mainly related to the regulatory sources of circadian rhythm and leaf senescence **(Supplementary Data Table S13**).

Genetic diversity, divergence, and population structure in the APPM complex

To further study population genetics of introgression, and divergent and convergent selection in the APPM complex, whole-genome sequencing data were collected from 171 accessions, including 41 peach (4 *P. dulcis* and 37 *P. persica*), 46 apricot (5 *P. mandshurica* and 41 *P. armeniaca*), 36 mei (36 *P. mume*), 48 plum (1 *P. cerasifera* × *P. munsoniana*, 4 *P. cerasifera* and 6 *P. brigantine*, 37 *P. salicina*), and 6 outgroups (3 apples and 3 pears) from worldwide sources **(Supplementary Data Table S14**). After mapping resequencing data to a reference genome, we filtered low-quality variants and retained a total of 8,011,197 high-quality SNPs. Based on the phylogenetic tree, we categorized the 171 accessions into four distinct clades: peach, plum, mei, and apricot accessions (**Figure 3A, Supplementary Data Fig. S6)**. We then investigated the population genetic structure from 2 to 15 clusters (K) based on 3,472,793 SNPs in 171 *Prunus* accessions to estimate the ancestral proportions of each germplasm **(Figure 3A, Supplementary Data Fig. S7**). When K = 4, four clades were separated, but a small mixture of apricots was observed in the plum population. Principal component analysis (PCA) further confirmed these population affiliations, with samples from each clade clustered together (**Figure 3B, C, Supplementary Data Fig. S8)**. To explore the differences between these four clades, we assessed levels of population heterozygosity. Peach exhibited a lower population heterozygosity (1.33%) compared to the other populations: mei (2.95%), apricot (3.21%), and plum (3.31%) (**Figure 3D, Supplementary Data Table S14**).

The fixation index (F_{ST}) is a measure used in population genetics to quantify the level of genetic differentiation among populations. We calculated the values of F_{ST} and found that the highest F_{ST} was 0.419 between mei and peach, 0.362 between apricot and peach, and the lowest F_{ST}, both 0.258 between apricot_mei and apricot_plum. The results indicate a high genetic differentiation in mei and peach. To examine the levels of inbreeding in the four clades, we tested the genetic diversity (π). Among the three clades, peach exhibited the lowest nucleotide diversity ($\pi = 8.82 \times 10^{-4}$), followed by mei ($\pi = 1.55\times 10^{-3}$), apricot ($\pi = 2.04 \times 10^{-3}$) and plum ($\pi = 2.17 \times 10^{-3}$) **(Figure 4A, Supplementary Data Fig. S9)**.

Gene flow in the APPM complex

The F_{ST} and admixture results suggest potential genetic information exchange between plum and apricot, but the specific pattern of gene flow is not clear (**Figure 3A, 4A**). To determine

possible genome-wide introgression in present species clades, we estimated the F-branch (f_b) statistics using Dsuite[61] to quantify potential signals of related gene flow and patterns of allele sharing specific to branches. Peach had a weaker introgression signal with plum (f_b = 0.007, Z= 3.93) and apricot (f_b = 0.004, Z= 3.31), respectively. Strong signals of introgression (f_b = 0.014, Z= 7.42) were identified in the plum and apricot **(Figure 4B)**, suggesting potential introgression events among these species. As with f_b, the triple topology showed that the apricot-plum comparison (average f_d of 7.02×10^{-2}) had higher genome-wide values than the apricot-peach comparison (average f_d of 5.78×10^{-2}, **Supplementary Data Fig. S10, Table S15**). Given the evidence for introgression, we also applied the fdM and f_d statistic based on 20 kb (overlapping 10 kb) windows. The fdM statistic and the histogram confirm a high level of introgression between plum and apricot, as there are 6,463 windows (20 kb) between plum and apricot, and 6,057 windows between mei and plum (**Figure 4C, Supplementary Data Table S15, 16**).

To clarify the direction of gene flow migration between the plum and apricot, we inferred migration events and patterns based on whole genome allele frequency data. Treemix[62] results suggest a migration event between plum and apricot, sharing crucial signals from specific genotypes (m = 1, 2), which is associated with a lower migration weight (**Figure 4D)**. In addition, we analyzed genes within the top 5% of high f_d windows between apricot and plum, identifying a total of 524 genes **(Supplementary Data Table S17)**. GO analysis indicated that these genes were notably enriched in biological processes, including cysteine biosynthetic process, regulation of post-embryonic development, cell wall organization, and pollen germination (**Supplementary Data Fig. S11, Table S18**). Furthermore, we inscribed the genome-wide region of the highest f_d potential introgression between apricot and plum (Chr2: 16.92 -16.94 Mb). As expected, π, population branch statistics (PBS) **(Supplementary Data Fig. S12)**, and F_{ST} in apricot and plum populations were much lower in this region **(Figure 4)**. It is worth noting that the top of the f_d outlier window contained introgressed signals related to the *TOGT1* gene on chromosome 2 (Chr2: 16,922,546-16,923,985 bp), which is associated with phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance[63].

Convergent and divergent selection signals in the APPM complex

To examine the potential mechanism of adaptation to environmental changes during *Prunus*'s evolution, we performed a composite-likelihood based (CLR) test and assessed the evidence for positive selection in different populations[56,64]. We compared the genomes of 41 peach, 46

apricot, 36 mei, and 48 plum accessions to detect selective sweeps[23,64] Our analysis pinpointed genes within the top 1% CLR regions identified by SweeD in apricot, mei, plum, and peach. **(Figure 5A, Supplementary DataTable S19)**. Subsequently, we performed GO enrichment analysis for genes within these top 1% CLR regions and discovered nine functional categories that were shared among all four clades ($P \leq 0.01$), including RNA modification, embryo development ending in seed dormancy, cytokinesis by cell plate formation and cell wall organization **(Figure 5B, Supplementary Data Table S20)**. We identified common genes shared across four clades and genes unique to individual clades. Shared genes located on chromosomes 1, 3, 6, and 7 include those involved in response to biological and abiotic stress (e.g., *EP3*, *BAGP1*, *TIFY9*), flower development and seed-related processes (e.g., *TCP4*, *SPL4*, *TIO*), and fruit ripening (e.g., *PG*) **(Supplementary Data Fig. S13 A-D, Table S21)**. Apricot-specific genes on chromosome 2 are enriched in processes such as callose deposition[65], chloroplast RNA processing, and anther wall tapetum development (e.g., *PME5*, *BHLH10*, *BHLH91*). Mei-specific genes on chromosome 7 are associated with pollen germination, vesicle fusion, growth regulation, and carbohydrate metabolism (e.g., *SWEETIE*, *NDF5*, *BDG4*)[66–68]. Plum-specific genes on chromosome 4 are linked to plastid and chloroplast fission, and somatic embryogenesis. Peaches exhibit specific genes on chromosome 1, including auxin polar transport and hormone signal regulation genes (e.g., *3BETAHSD/D2*, *ACL5*, *APM1*, *AHK5*, *AIP2*, *GCR1*) **(Supplementary Data Fig. S16 A, F, Table S24, S25)**.

Our analyses detect genomic regions associated with convergent selectivity sweeps. Specifically, we identified that the convergent region on chromosome 7, corresponds to a chromosomal co-linearity in peach according to the comparative genomics analysis **(Figure 5F, Supplementary Data Fig. S14, Table S21, 22)**. Evolutionary tree analysis of SNPs within the convergent region on chromosome 7 was consistent with the analysis utilizing all 3,472,793 SNPs **(Supplementary Data Fig. S13E, Figure 3A)**. Further examination focused on individual genes within this region, particularly Polygalacturonase (*PG*), which exhibited a distinct evolutionary tree compared to others. Given *PG*'s known association with fruit ripening and softening, it might be regulated by two convergent genes[69,70]: E3 ubiquitin-protein ligase (*RMA1H1*) and Ethylene-responsive transcription factor (*ERF053*) **(Figure 5F)**. The evolutionary tree results for these genes were largely consistent with the analysis using all 3,472,793 SNPs, thereby indicating the significance of *PG* for further exploration **(Supplementary Data Fig. S13F - H)**. Haplotype analysis of all coding region genes *PG* generated a diverse array of haplotypes **(Supplementary Data Fig. S15)**. A total of 60

haplotypes were identified, which were categorized into five haplotype clades (*PG*-hg1/2/3/4/5) based on a neighbor-joining tree **(Figure 5C, D)**. These clades exhibited non-uniform distribution across the Apricot-Peach-Plum-Mei (APPM) populations **(Figure 5E)**, with distinctive proportions observed in different fruit types.

Additionally, we identified unique genes under divergent selection in the top 1% CLR regions in apricot, mei, plum, and peach populations. Mei possesses two specific *PG* genes, while plum has one unique *PG* gene. In the AP2/ERF transcription factor family, mei harbors two distinct genes, and peach has one exclusive gene. Regarding the response to ethylene, apricot (*ACBP4*) has one unique gene, and peach (*SQE1*) possesses another. In the regulation of the ethylene biosynthetic process, mei (*SWEETIE*) has two exclusive genes, and peach (*ACS8*) has one. For the ethylene-activated signaling pathway, mei (*EEN*, *THO1*, *XCT*) has three unique genes, plum (*SRT1*) has one, and peach (*AHK5*, *EDR1*) has two. Notably, peach has two genes dedicated to the negative regulation of the ethylene-activated signaling pathway (*AHK5*, *RTE1*). *AHK5* plays a dual role in both the ethylene-activated signaling pathway and its negative regulation **(Figure 5F- I, Supplementary Data Fig. S16, Table S26)**.

In conclusion, our comprehensive analysis provides valuable insights into the genetic variations and associations within key genomic regions related to fruit ripening and ethylene response, paving the way for further investigation into the functional roles of these genes in fruit development and quality regulation.

Discussion

The APPM complex shares a recent evolutionarily history and convergent and divergent signals of selection. The candidate genes could be applied in genetic studies and breeding programs for the APPM crops. In this research, by integrating HiFi and Hi-C reads, the haplotype-resolved T2T reference genome of plum was successfully assembled for the first time. The average N50 value of this genome is 19.85 Mb, which is 14 times higher than previously published genomes[30,71]. Comparative genomics enables the identification of shared and unique genomic regions across *Prunus* species, revealing patterns of gene acquisition, loss, and duplication that underlie their adaptive strategies[72,73]. Through genome collinearity analysis, there is a strong shared relationship was evident among the genomes of *Prunus* species[28–30]. Significantly, this study reports chromosomal translocation of approximately 1.17 Mb (Chr7: 5.10 - 6.27 Mb) for the first time, revealing structural changes in the genomes of apricot compared to peach, plum, and mei[56,58]. The acidity and sugar of peach fruit are

controlled by *D* locus, which is short for 'Doux', meaning 'sweet' in French[74–77], and is mapped to a 509-kb interval on Chr5: 703 - 1,212 kb[78]. Over the past several decades, it has remained a captivating subject of genetic investigation. Additionally, as we explore the analysis of the *D* locus in *Prunus* species, the translocation (Chr5: 611,362 - 1,950,661 bp) in peach merits consideration **(Figure 2B, Supplementary Data Table S9 - 11)**.

Phylogenetic and population structure analyses suggest a close shared ancestry among plum, apricot, and mei, with the possibility of independent evolution driven by specific environmental or anthropogenic factors during the evolutionary process. Our findings indicate that *P. brigantina* is aligned with the plum category, as supported by chloroplast DNA sequences[39], but diverges from this alignment in nuclear DNA sequences, being closer to apricot[38], suggesting it might be a hybrid descendant of both plum and apricot species, with the admixture results reflecting the influence of different parental species (**Figure 3A-C, Supplementary Data Fig. S6, 7, Table S14**). Peaches exhibit notably lower heterozygosity values (**Figure 3D**). Historically, peaches have been associated with reduced genetic variability due to their self-compatible (SC) mating system[79,80]. However, in the *Prunus* genus, self-incompatibility (SI) is generally the rule, with most species being partially or fully self-incompatible, thereby facilitating gene flow[66,81]. For introgression to take place, two species must have overlapping geographic ranges and must be sufficiently closely related so that at least some hybrid offspring are fertile. Despite most hybrids being unfit, some recombinant hybrids must have high fitness in the adaptive landscape. The observed high diversity in the *Prunus* genus could also be attributed to interspecific hybridization and additional introgression through backcrossing among closely related *Prunus* species[40,82].

Our findings present compelling evidence of genetic exchange between plum and apricot, which contradicts previous studies[38]. According to speciation with gene flow models, a key prediction is that loci involved in speciation should exhibit high relative divergence (F_{ST})[83–85]. Notably, the lowest F_{ST} value (0.258) was observed between apricot_mei and apricot_plum, indicating a high level of gene flow between apricot and either mei or plums. As seen in many other species, such as three-spined sticklebacks, regions of the genome with high rates of recombination facilitate the decoupling of deleterious introgressed alleles from neighboring loci, promoting gene flow and resulting in lower levels of divergence (F_{ST}) in these regions[1,86]. Mei displays high genetic diversity, originating from independent domestication events, and a relatively small proportion of the genome is affected by selection[56]. The f_b statistic results reflected no evidence for introgression between the apricot and mei clades (f_b = 0), but the f_b value (0.014) between the apricot and plum clades was consistent with the

possibility of introgression (low F_{ST}). We have utilized ABBA-BABA statistics (e.g., f_b, f_d, and fdM) and TreeMix to estimate potential migration events, revealing ample evidence for introgression from apricot to plum (**Figure 4C, D, Supplementary Data Table S15, 16**). Given the introgression between apricot and plum, an important question is whether it has had an adaptive basis. Adaptive introgression has played a significant role in crop domestication, as observed in crops like maize, barley, potato, and rice[87]. Our analyses offer compelling evidence that introgression has contributed to essential agronomic traits in plums, with the introgression enriched with 524 pivotal genes related to important biological processes, including the regulation of postembryonic development and pollen germination (**Supplementary Data Table S17, 18**). Studies conducted in other taxa have also demonstrated that introgression can promote species adaptation through various mechanisms[1,88].

The APPM complex shares a recent evolutionary history, and it is essential to uncover the convergent and divergent signals of genomic selection in their genomes during domestication and adaptation for interspecific hybrid breeding. In our study, we identified a series of shared and specific genes among these four clades, revealing the complex genetic landscape during their evolution and adaptation. Each gene plays a distinct functional role in response to various environmental cues and developmental processes **(Figure 5, Supplementary Data Fig. S13, S16)**. Notably, convergent selectivity sweeps on chromosome 7 coincide with a chromosomal co-linearity from the comparative genomics **(Supplementary Data Table S21, 22)**, impacting crucial fruit softening genes like *PG*. The regulatory network involves specific genes like *MdPG1* and *FcPG12*, regulated by ERFs such as *MdCBF2* and *FcERF5*, contributing to fruit softening[69,70]. CRISPR/Cas9 applications in tomatoes demonstrate the potential for delaying fruit softening[89]. The ubiquitin-mediated proteolysis pathway, represented by E3 ligases like *PRT6*, influences gene expression, impacting ERF regulation[90]. The transcriptional repressor *SlERF.F12*, known for recruiting co-repressors and histone deacetylases, influences fruit ripening in tomatoes. These findings emphasize the complex molecular mechanisms governing fruit development and postharvest traits[69,70,89–91]. Additionally, the identification of a peach-specific divergent selection region on chromosome 1[92], influenced by domestication and associated with increased fruit size, provides valuable insights into the transition from wild to cultivated varieties in *Prunus* crops. These candidate genes hold potential for future genetic studies and breeding programs, contributing to ongoing improvements in *Prunus* characteristics.

The potential for hybridization and introgression between two populations represents a

multifaceted process shaped by genetic divergence, selection, recombination, and demographic factors. Adaptive introgression of a specific allele depends on its advantages, genomic location, impact on recipient population fitness, and recent demographic history[1].

The outcomes of hybridization are predominantly determined by long-term influences, including the fitness of early-generation hybrids and the emergence of novel allele combinations, which serve as catalysts for evolutionary change by introducing fitness variability. Under variable environmental conditions, this process enables populations to overcome challenges, explore new evolutionary paths, and persist.

Materials and methods

For full materials and methods, see **Supplementary Data**.

Plant materials, and sequencing

P. salicina cv. 'Fengtangli' is a high-quality plum variety. Plant materials were collected from Guizhou Province. The Hi-C library was prepared with NEBNext® Ultra™ II kit and sequenced on Illumina HiSeq X Ten (150PE). Two replicates per group ensured robust Hi-C results. For HiFi sequencing, PacBio Sequel II with CCS was used, and SMRTlink software processed raw data for high-quality sequences, adhering to specific quality parameters.

Genome De Novo Assembly and Quality Assessment

The HiFi and Hi-C data were assembled using HiFiasm[93,94], producing two contig-level haplotype genomes. Genome heterozygosity was estimated using a k-mer-based approach by GenomeScope[95]. RagTag[96] anchored and removed short contigs, aligning with *P. salicina* cv. 'Sanyueli' serves as the reference genome. Juicer[97] and 3D-DNA[98] were employed for scaffold-level assembly by HiC data. For better genome quality, the obtained results were manually adjusted using Juicebox[45] before running 3D-DNA again to obtain the genome at the scaffold level. To verify the correctness of gap filling, Minimap2[99] was used to compare the original HiFi data for sequence comparison. The resulting sequence positions of gaps were located using IGV[99], and the gaps were subsequently filled in the genome using Minimap2. Genome completeness relied on BUSCOs with the embryophyta_odb10 database[100], and continuity was assessed via contig N50 length.

Identification of Telomeres and Centromeres

To identify telomeres, we used TIDK to recognize telomeric sequences with 'CCCATTT' at the 5' end and 'TTTAGGG' at the 3' end. To detect centromeric regions, we scanned candidate repeats from 30 to 500 bp along the genome using TRF v4.09[53]. Then, the centromeric repeat units are identified by comparing the results of the TE distribution. Finally, the 166 bp repeats were the most abundant unit in the whole genome.

Genome annotation and identified transposable elements

The raw RNA sequence data were used Trimmomatic[101] to eliminate adapter sequences. HISAT2[102] and StringTie[103] were employed to remove bases with more than 10% ambiguous nucleotides (N) and quality values ≤5. Gene annotation followed the Genome-Wide-Annotation-Pipeline (https://github.com/unavailable2374/Genome-Wide-Annotation-Pipeline). Functional gene annotation was conducted using Interproscan[104].

Comparative genomics and gene families

For whole-genome alignment, Minimap2 was used to align the genomes, and the BAM file was indexed using SAMtools[105]. We used SyRI[106] for genome collinearity alignment and Plotsr for visualization. Additionally, we aligned the PS_T2T genome to *P. salicina* cv. 'Sanyueli' reference using MUMmer, extracted detailed comparisons with the 'show coords -T -q -H' parameter and created dot plots for visualization with Gnuplot.

Mcscan (Python version) was used for collinearity analysis of apricot[56], peach[57], plum, and mei[58] genomes and detection of structural changes. Use Orthofinder to analyze the gene families of apricots, peaches, plums, and plums[107]. David was used for GO enrichment analysis on shared and species-specific gene sets.

Phylogenetic and population structure analyses

We analyzed illumina raw reads from 177 samples (SRA at NCBI) **(Supplementary Data Table S14)**. Illumina short reads were aligned to the TJSM genome using BWA-MEME, sorted with SAMtools[108], and PCR duplicates removed with GATK[109]. SNPs were called and filtered using GTX[110,111] and VCFtools[112], leaving 3,749,618 SNPs for heterozygosity analysis, calculated as (N(NM) - O(HOM))/ N(NM), where N(NM) represents observed variants, and O(HOM) represents observed homozygous variants[113]. PLINK v1.9[114] was used for a phylogenetic tree, population structure, and PCA on samples with MAF ≥ 0.05 and missing rate ≤ 20%. FastTree2[115] built the phylogenetic tree, and ADMIXTURE[116] explored population structure (K from 2 to 15). Heterozygous sites and π for SNPs in each group were

computed using VCFtools v0.1.15. Fixation indices (F_{ST}) were calculated with 50-kb nonoverlapping windows to identify domestication and differentiation regions.

Introgression analysis with ABBA-BABA tests and TreeMix

We analyzed introgression events using D-statistic, f_d, fdM, and fb statistics with an 8,011,997 SNPs dataset. For a triplet of taxa P1, P2, and P3, and an outgroup, that follows the phylogeny of (((P1, P2), P3), Outgroup), a D-statistic significantly different from zero indicates P3 exchanged gene with P1 (D value 0) or P2 (D value >0). To assess introgression divergence, we filtered windows with genotyped variants < 100, computing f_d and fdM using the ABBABABAwindows.py. D-statistics based on SNP frequency differences were also computed to identify gene exchange. TreeMix[62] inferred population relatedness and migration events, highlighting results using the plotting_func.R script. PBS was calculated using PBScan, each with 20 kb sliding windows, and windows with the top 5% of values were selected as highly divergent regions.

Detection genome scanning for selective sweep signals

We used SweeD[64] with default parameters to calculate CLR scores, identifying potential regions under positive selection in peach, apricot, plum, and mei. The top 1% regions of CLR scores were considered as potential positive selection, and gene functions within these regions were annotated. Heterozygous sites, π, and F_{ST} were calculated with VCFtools v0.1.15 in 50-kb nonoverlapping windows. PBS was calculated using PBScan, each with 20 kb sliding windows, and windows with the top 5% of values were selected as highly divergent regions.

Data availability

The 'PS_T2T' raw genome sequencing reads were available from the NCBI under project ID PRJNA1000098. All raw sequence data and assembly have been deposited in the National Genomics Data Center (NGDC) Genome Sequence Archive (GSA) (https://ngdc.cncb.ac.cn/gsa/), with BioProject number PRJCA018700. The assembly and annotation have been deposited in zenodo: https://zenodo.org/records/10570999.

Acknowledgements

We are grateful to the Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, for providing the high-performance computing platforms during this study. This work was financially supported by the Guizhou Provincial Science and

Technology Projects of China ([2022]Zhongdian018, YQK[2023]008), the National Guidance Foundation for Local Science and Technology Development of China (Grant No. 2023-009), and the National Natural Science Fund for Excellent Young Scientists Fund Program (Overseas) to Yongfeng Zhou.

Author contributions

In this project, G.Q., K.C., and Y.Z. were responsible for conceiving and designing the research. X.Y., Y.S., and S.H. contributed to writing the manuscript. X.Y. and S.H. were responsible for collecting and updating the genome assemblies, while Y.S. performed the gene annotation. All authors contributed to the data collection and/or bioinformatics analyses. Y.M. and Q.H. provided the experimental materials. All authors have reviewed and approved the final manuscript.

Conflicts of interest

The authors declare that they have no competing interests.

Supplementary Data

Supplementary information accompanies the manuscript on the Horticulture Research website http://www.nature.com/hortres.

References

1 Edelman NB, Mallet J. Prevalence and adaptive impact of introgression. *Annu Rev Genet* 2021; **55**: 265–283.

2 Anderson E, Hubricht L. The evidence for introgressive hybridization. *Am J Bot* 1938; **25**: 396–402.

3 Suarez-Gonzalez A, Lexer C, Cronk QCB. Adaptive introgression: a plant perspective. *Biology Letters* 2018; **14**: 20170688.

4 Suarez-Gonzalez A, Hefer CA, Lexer C, Douglas CJ, Cronk QCB. Introgression from *Populus balsamifera* underlies adaptively significant variation and range boundaries in *P. trichocarpa*. *New Phytol* 2018; **217**: 416–427.

5 He F, Pasam R, Shi F *et al.* Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. *Nat Genet* 2019; **51**: 896–904.

6 Cheng J, Kao H, Dong S. Population genetic structure and gene flow of rare and endangered *Tetraena mongolica* Maxim. revealed by reduced representation sequencing. *BMC Plant Biol* 2020; **20**: 391.

7 Xiao H, Liu Z, Wang N *et al.* Adaptive and maladaptive introgression in grapevine domestication. *PNAS* 2023; **120**: e2222041120.

8 Hämälä T, Savolainen O. Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*. *Mol Biol Evol* 2019; **36**: 2557–2571.

9 Vonlanthen P, Bittner D, Hudson AG *et al.* Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* 2012; **482**: 357–362.

10 Todesco M, Pascual MA, Owens GL *et al.* Hybridization and extinction. *Evol Appl* 2016; **9**: 892–908.

11 Morales-Cruz A, Aguirre-Liguori JA, Zhou Y *et al.* Introgression among North American wild grapes (*Vitis*) fuels biotic and abiotic adaptation. *Genome Biology* 2021; **22**: 254.

12 Aeschbacher S, Selby JP, Willis JH, Coop G. Population-genomic inference of the strength and timing of selection against gene flow. *PNAS* 2017; **114**: 7061–7066.

13 Burri R. Interpreting differentiation landscapes in the light of long-term linked selection. *Evol Lett* 2017; **1**: 118–131.

14 Cutter AD, Payseur BA. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* 2013; **14**: 262–274.

15 Yeaman S, Hodgins KA, Lotterhos KE *et al.* Convergent local adaptation to climate in distantly related conifers. *Science* 2016; **353**: 1431–1433.

16 Birkeland S, Gustafsson ALS, Brysting AK, Brochmann C, Nowak MD. Multiple genetic trajectories to extreme abiotic stress adaptation in arctic Brassicaceae. *Mol Biol Evol* 2020; **37**: 2052–2068.

17 Xu S, Wang J, Guo Z, He Z, Shi S. Genomic convergence in the adaptation to extreme environments. *Plant Commun* 2020; **1**: 100117.

18 Lin Z, Qin P, Zhang X *et al.* Divergent selection and genetic introgression shape the genome landscape of heterosis in hybrid rice. *PNAS* 2020; **117**: 4623–4631.

19 Stephan W. Selective Sweeps. *Genetics* 2019; **211**: 5–13.

20 Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *eLife* 2018; **7**: e36317.

21 Clark RM, Linton E, Messing J, Doebley JF. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *PNAS* 2004; **101**: 700–707.

22 Qanbari S, Pausch H, Jansen S *et al.* Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet* 2014; **10**: e1004148.

23 Zhou Y, Massonnet M, Sanjak JS, Cantu D, Gaut BS. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *PNAS* 2017; **114**: 11715–11720.

24 Wang N, Cao S, Liu Z *et al.* Genomic conservation of crop wild relatives: A case study of citrus. *PLoS Genet* 2023; **19**: e1010811.

25 Huang X, Huang S, Han B, Li J. The integrated genomics of crop domestication and breeding. *Cell* 2022; **185**: 2828–2839.

26 Gaut BS, Seymour DK, Liu Q, Zhou Y. Demography and its effects on genomic variation in crop domestication. *Nature Plants* 2018; **4**: 512–520.

27 Abdulrahman SS, Daştan SD, Shahbaz SE, Selamoglu Z. Phylogenetic analysis of *Prunus* genus using nuclear and chloroplast gene markers as a bioorganic structure profiling. *J Mol Struct* 2023; **1284**: 135300.

28 Tan Q, Li S, Zhang Y *et al.* Chromosome-level genome assemblies of five *Prunus* species and genome-wide association studies for key agronomic traits in peach. *Hortic Res* 2021; **8**: 213.

29 Liu C, Feng C, Peng W *et al.* Chromosome-level draft genome of a diploid plum (*Prunus salicina*). *Gigascience* 2020; **9**: giaa130.

30 Fang Z, Lin-Wang K, Dai H *et al.* The genome of low-chill Chinese plum 'Sanyueli' (*Prunus salicina* Lindl.) provides insights into the regulation of the chilling requirement of flower buds. *Mol Ecol Resour* 2022; **22**: 1919–1938.

31 Xiang Y, Huang C-H, Hu Y *et al.* Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Mol Biol Evol* 2017; **34**: 262–281.

32 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* 2014; **76**: 34–48.

33 Zhuo X, Zheng T, Li S *et al.* Identification of the *PmWEEP* locus controlling weeping traits in *Prunus mume* through an integrated genome-wide association study and quantitative trait locus mapping. *Hortic Res* 2021; **8**: 1–14.

34 Zhang Q, Zhang H, Sun L *et al.* The genetic architecture of floral traits in the woody plant *Prunus mume*. *Nat Commun* 2018; **9**: 1702.

35 Numaguchi K, Akagi T, Kitamura Y, Ishikawa R, Ishii T. Interspecific introgression and natural selection in the evolution of Japanese apricot (*Prunus mume*). *TPJ* 2020; **104**: 1551–1567.

36 Yingsakmongkon S, Miyamoto D, Sriwilaijaroen N *et al. In vitro inhibition* of human influenza a virus infection by fruit-juice concentrate of Japanese plum (*Prunus mume* SIEB. *et* ZUCC). *Biol Pharm Bull* 2008; **31**: 511–515.

37 Ni X, Ni Z, Ouma KO, Gao Z. Mutations in *PmUFGT3* contribute to color variation of fruit skin in

Japanese apricot (*Prunus mume* Sieb. et Zucc.). *BMC Plant Bio* 2022; **22**: 304.

38 Liu S, Decroocq S, Harte E *et al.* Genetic diversity and population structure analyses in the Alpine plum (*Prunus brigantina* Vill.) confirm its affiliation to the Armeniaca section. *Tree Genet Genomes* 2021; **17**: 2.

39 Reales A, Sargent DJ, Tobutt KR, Rivera D. The genome of low-chill Chinese plums, *Prunus* L. section *Prunus* (Rosaceae), according to coding and non-coding chloroplast DNA sequences. *Tree Genet Genomes* 2010; **6**: 37–45.

40 Zhebentyayeva T, Shankar V, Scorza R *et al.* Genetic characterization of worldwide *Prunus domestica* (plum) germplasm using sequence-based genotyping. *Hortic Res* 2019; **6**: 12.

41 Guerrero BI, Guerra ME, Rodrigo J. Simple sequence repeat (SSR)-based genetic diversity in interspecific Plumcot-type (*Prunus salicina* × *Prunus armeniaca*) hybrids. *Plants* 2022; **11**: 1241.

42 Szymajda M, Studnicki M, Kuras A, Żurawicz E. Cross-compatibility in interspecific hybridization between three *Prunus* species. *S Afr J Bot* 2022; **146**: 624–633.

43 Rimpika, Sharma D. Advances in Breeding of Peach, Plum and Apricot. In: B. Küden A, Küden A (eds). *Prunus - Recent Advances*. IntechOpen, 2022 doi:10.5772/intechopen.100284.

44 Hodel RGJ, Zimmer E, Wen J. A phylogenomic approach resolves the backbone of *Prunus* (Rosaceae) and identifies signals of hybridization and allopolyploidy. *Mol Phylogenet Evol* 2021; **160**: 107118.

45 Durand NC, Robinson JT, Shamim MS *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* 2016; **3**: 99–101.

46 Chen J, Wang Z, Tan K *et al.* A complete telomere-to-telomere assembly of the maize genome. *Nat Genet* 2023; **55**: 1221–1231.

47 Shang L, He W, Wang T *et al.* A complete assembly of the rice Nipponbare reference genome. *Molecular Plant* 2023; **16**: 1232–1236.

48 Shi X, Cao S, Wang X *et al.* The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. *Hortic Res* 2023; **10**: uhad061.

49 Wang X, Tu M, Wang Y *et al.* Telomere-to-telomere and gap-free genome assembly of a susceptible grapevine species (Thompson Seedless) to facilitate grape functional genomics. *Hortic Res* 2024; **11**: uhad260.

50 Sun M, Yao C, Shu Q *et al.* Telomere-to-telomere pear (*Pyrus pyrifolia*) reference genome reveals segmental and whole genome duplication driving genome evolution. *Hortic Res* 2023; **10**: uhad201.

51 Yue J, Chen Q, Wang Y *et al.* Telomere-to-telomere and gap-free reference genome assembly of the kiwifruit Actinidia chinensis. *Hortic Res* 2022; **10**: uhac264.

52 Perea-Resa C, Blower MD. Centromere biology: transcription goes on stage. *Mol Cell Biol* 2018; **38**: e00263-18.

53 Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999; **27**: 573–580.

54 Ou S, Su W, Liao Y *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* 2019; **20**: 275.

55 Zhou Y, Gaut BS. Large chromosomal variants drive adaptation in sunflowers. *Nat Plants* 2020; **6**: 734–735.

56 Groppi A, Liu S, Cornille A *et al.* Population genomics of apricots unravels domestication history and adaptive events. *Nat Commun* 2021; **12**: 3956.

57 Wang J, Cao K, Li Y *et al.* Genome variation and LTR-RT analyses of an ancient peach landrace reveal mechanism of blood-flesh fruit color formation and fruit maturity date advancement. *Hortic Res* 2024; **11**: uhad265.

58 Zheng T, Li P, Zhuo X *et al.* The chromosome-level genome provides insight into the molecular mechanism underlying the tortuous-branch phenotype of *Prunus mume*. *New Phytol* 2022; **235**: 141–156.

59 Qin X, Duan Z, Zheng Y *et al.* ABC1K10a, an atypical kinase, functions in plant salt stress tolerance. *BMC Plant Biol* 2020; **20**: 270.

60 Sylvestre-Gonon E, Law SR, Schwartz M *et al.* Functional, structural and biochemical features of plant serinyl-glutathione transferases. *Front Plant Sci* 2019; **10**: 608.

61 Malinsky M, Matschiner M, Svardal H. Dsuite - Fast *D* -statistics and related admixture evidence from VCF files. *Mol Ecol Resour* 2021; **21**: 584–595.

62 Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 2012; **8**: e1002967.

63 Chong J, Baltz R, Schmitt C, Beffa R, Fritig B, Saindrenan P. Downregulation of a pathogen-responsive tobacco UDP-Glc:phenylpropanoid glucosyltransferase reduces scopoletin glucoside accumulation, enhances oxidative stress, and weakens virus resistance. *Plant Cell* 2002; **14**: 1093–1107.

64 Pavlidis P, Živković D, Stamatakis A, Alachiotis N. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. *Mol Biol Evol* 2013; **30**: 2224–2234.

65 German L, Yeshvekar R, Benitez-Alfonso Y. Callose metabolism and the regulation of cell walls and plasmodesmata during plant mutualistic and pathogenic interactions. *Plant Cell Environ* 2023; **46**: 391–404.

66 Abdallah D, Baraket G, Perez V, Salhi Hannachi A, Hormaza JI. Self-compatibility in peach [*Prunus persica* (L.) Batsch]: patterns of diversity surrounding the S-locus and analysis of SFB alleles. *Hortic Res* 2020; **7**: 1–15.

67 Muñoz-Sanz JV, Zuriaga E, Badenes ML, Romero C. A disulfide bond A-like oxidoreductase is a strong candidate gene for self-incompatibility in apricot (*Prunus armeniaca*) pollen. *J Exp Bot* 2017; **68**: 5069–5078.

68 Ahmad MH, Rao MJ, Hu J *et al.* Systems and breakdown of self-incompatibility. *Crit Rev Plant Sci* 2022; **41**: 209–239.

69 Wu M, Luo Z, Cao S. Promoter variation of the key apple fruit texture related gene *MdPG1* and the upstream regulation analysis. *Plants* 2023; **12**: 1452.

70 Wang Y, Fan Z, Zhai Y, Huang H, Vainstein A, Ma H. Polygalacturonase gene family analysis identifies *FcPG12* as a key player in fig (*Ficus carica* L.) fruit softening. *BMC Plant Biology* 2023; **23**: 320.

71 Huang Z, Shen F, Chen Y, Cao K, Wang L. Chromosome-scale genome assembly and population

genomics provide insights into the adaptation, domestication, and flavonoid metabolism of Chinese plum. *TPJ* 2021; **108**: 1174–1192.

72 Feng Y, Neme R, Beh LY *et al.* Comparative genomics reveals insight into the evolutionary origin of massively scrambled genomes. *eLife* 2022; **11**: e82979.

73 Lin H, Yu M, Wang X, Zhang X-H. Comparative genomic analysis reveals the evolution and environmental adaptation strategies of vibrios. *BMC Genomics* 2018; **19**: 135.

74 Dirlewanger E, Moing A, Rothan C *et al.* Mapping QTLs controlling fruit quality in peach (*Prunus persica* (L.) Batsch): *Theor Appl Genet* 1999; **98**: 18–31.

75 Etienne C, Rothan C, Moing A *et al.* Candidate genes and QTLs for sugar and organic acid content in peach [*Prunus persica* (L.) Batsch]. *Theor Appl Genet* 2002; **105**: 145–159.

76 Peng Q, Wang L, Ogutu C *et al.* Functional analysis reveals the regulatory role of *PpTST1* encoding tonoplast sugar transporter in sugar accumulation of peach fruit. *Int J Mol Sci* 2020; **21**: 1112.

77 Wang Q, Cao K, Cheng L *et al.* Multi-omics approaches identify a key gene, *PpTST1*, for organic acid accumulation in peach. *Hortic Res* 2022; **9**: uhac026.

78 Wang L, Jiang X, Zhao L *et al.* A candidate *PpRPH* gene of the *D* locus controlling fruit acidity in peach. *Plant Mol Biol* 2021; **105**: 321–332.

79 Miller PJ, Parfitt DE, Weinbaum SA. Outcrossing in peach. *HortScience* 1989; **24**: 359–360.

80 Cao K, Zheng Z, Wang L *et al.* Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. *Genome Biol* 2014; **15**: 415.

81 Gradziel TM. Transfer of self-fruitfulness to cultivated almond from peach and wild almond. *Horticulturae* 2022; **8**: 965.

82 Rimpika, Sharma DP. Advances in breeding of Peach, Plum and Apricot. In: *Prunus - Recent Advances*. IntechOpen, 2021 doi:10.5772/intechopen.100284.

83 Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 2014; **23**: 3133–3157.

84 Zhou Y, Zhang L, Liu J, Wu G, Savolainen O. Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. *Mol Ecol* 2014; **23**: 3504–3522.

85 Malinsky M, Challis RJ, Tyers AM *et al.* Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 2015; **350**: 1493–1498.

86 Roesti M, Moser D, Berner D. Recombination in the threespine stickleback genome—patterns and consequences. *Mol Ecol* 2013; **22**: 3014–3027.

87 Janzen GM, Wang L, Hufford MB. The extent of adaptive wild introgression in crops. *New Phytol* 2019; **221**: 1279–1288.

88 Taylor SA, Larson EL. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat Ecol Evol* 2019; **3**: 170–177.

89 Nie H, Shi Y, Geng X, Xing G. CRISRP/Cas9-mediated targeted mutagenesis of tomato polygalacturonase gene (*SlPG*) delays fruit softening. *Front Plant Sci* 2022; **13**: 729128.

90 Suka IE, Roslan NF, Chew BL *et al.* The effects of PROTEOLYSIS6 (*PRT6*) gene suppression on the expression patterns of tomato ethylene response factors. *J Plant Biochem Biotechnol* 2022; **31**: 115–123.

91 Deng H, Chen Y, Liu Z *et al.* SlERF.F12 modulates the transition to ripening in tomato fruit by recruiting the co-repressor TOPLESS and histone deacetylases to repress key ripening genes. *The Plant Cell* 2022; **34**: 1250–1272.

92 Li Y, Cao K, Zhu G *et al.* Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biol* 2019; **20**: 36.

93 Cheng H, Jarvis ED, Fedrigo O *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol* 2022; **40**: 1332–1335.

94 Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 2021; **18**.

95 Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* 2020; **11**: 1432.

96 Alonge M, Lebeigle L, Kirsche M *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol* 2022; **23**: 258.

97 Durand NC, Shamim MS, Machol I *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* 2016; **3**: 95–98.

98 Dudchenko O, Batra SS, Omer AD *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 2017; **356**: 92–95.

99 Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief BIoinform* 2013; **14**: 178–192.

100 Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M (ed). *Gene Prediction*. Springer New York: New York, NY, 2019, pp 227–245.

101 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–2120.

102 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019; **37**: 907–915.

103 Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 2015; **33**: 290–295.

104 Jones P, Binns D, Chang H-Y *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014; **30**: 1236–1240.

105 Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.

106 Goel M, Sun H, Jiao W-B, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* 2019; **20**: 277.

107 Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019; **20**: 238.

787 108 Li H, Handsaker B, Wysoker A *et al.* The Sequence Alignment/Map format and SAMtools.
788 *Bioinformatics* 2009; **25**. doi:10.1093/bioinformatics/btp352.

789 109 McKenna A, Hanna M, Banks E *et al.* The genome analysis toolkit: a MapReduce framework for
790 analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**: 1297–1303.

791 110 Yang R, Guo X, Zhu D *et al.* Accelerated deciphering of the genetic architecture of agricultural
792 economic traits in pigs using a low-coverage whole-genome sequencing strategy. *GigaScience* 2021;
793 **10**: giab048.

794 111 Bu L, Wang Q, Gu W *et al.* Improving read alignment through the generation of alternative reference
795 via iterative strategy. *Sci Rep* 2020; **10**: 18712.

796 112 Danecek P, Auton A, Abecasis G *et al.* The variant call format and VCFtools. *Bioinformatics* 2011;
797 **27**: 2156.

798 113 Zhou Y, Zhang Z, Bao Z *et al.* Graph pangenome captures missing heritability and empowers tomato
799 breeding. *Nature* 2022; **606**: 527–534.

800 114 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising
801 to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.

802 115 Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large
803 alignments. *PLoS ONE* 2010; **5**: e9490.

804 116 Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry
805 estimation. *BMC Bioinformatics* 2011; **12**: 246.

806

Figure Legends

Figure 1. Overview of the PS_T2T reference genome. (A-C) The pictures used in this study show the tree, fruits, and flowers of the 'Fengtangli' plum. (D) Overview of genome assembly. Collinearity between 'Sanyueli' and two haplotypes of PS_T2T. Gray lines represent the collinearity blocks with a length of 15,000 bp, while orange lines represent the potential inversions. Centromeres and telomeres are indicated by black boxes and black dots, respectively. (E) Hi-C interaction matrix based on the PS_T2T diploid assembly.

Figure 2. Comparative genomic analysis of the APPM genomes. (A) Synteny analysis at the gene level among four *Prunus* species. Light gray represents syntenic blocks. Dark gray and red lines denote co-linearity, with red arrows indicating a translocation. (B) Co-linearity comparison of four genomes demonstrates apricot translocation spanning 1.17 Mb (Chr7: 5.10 - 6.27 Mb) shared with three other species. Green syntenic blocks indicate *D* loci associated with agronomic traits. (C) Top enriched biological processes for genes in the 1.17 Mb translocation region. (D) Venn diagram comparing shared and private gene families of four *Prunus* species.

Figure 3 Population structure and heterozygosity of the APPM complex. (A) Phylogenetic tree of all accessions based on whole-genome SNPs. Different clades are represented by colored branches: 41 peach, 46 apricot, 36 mei, and 48 plum accessions. The estimated admixture proportions ranged from K= 3 to K= 5. (B-C) Principal component analysis (PCA) of 171 samples in *Prunus*. (D) Heterozygosity in each of the Apricot-Peach-Plum-Mei (APPM) complex.

Figure 4 Genetic differentiation and extensive introgression among four species in the APPM complex. (A) The heatmap indicates the pairwise differentiation (F_{ST}) statistics between the populations. The values within the blue circles denote the nucleotide diversity (π) for each respective population. (B) Heatmap representation of the f_b statistics across different populations. The grey boxes correspond to pairs that could not be tested due to topological constraints occurring in the system. (C) Distribution of window counts based on the fdM values for a 20 kb window. The green and pink colours represent the number of windows greater than and less than 0, respectively. (D) An inferred graphical representation highlighting migration events, based on genome-wide allele frequencies among the four clades. The yellow arrow indicates the direction of introgression. (E) An example of regions that are inferred to have introgressed from the apricot into plum. F_{ST}, π, and f_d values were evaluated in a 20 kb window. The population branch statistic (PBS) was calculated by comparing plum, apricot clade, and outgroup. The prominently displayed red vertical line

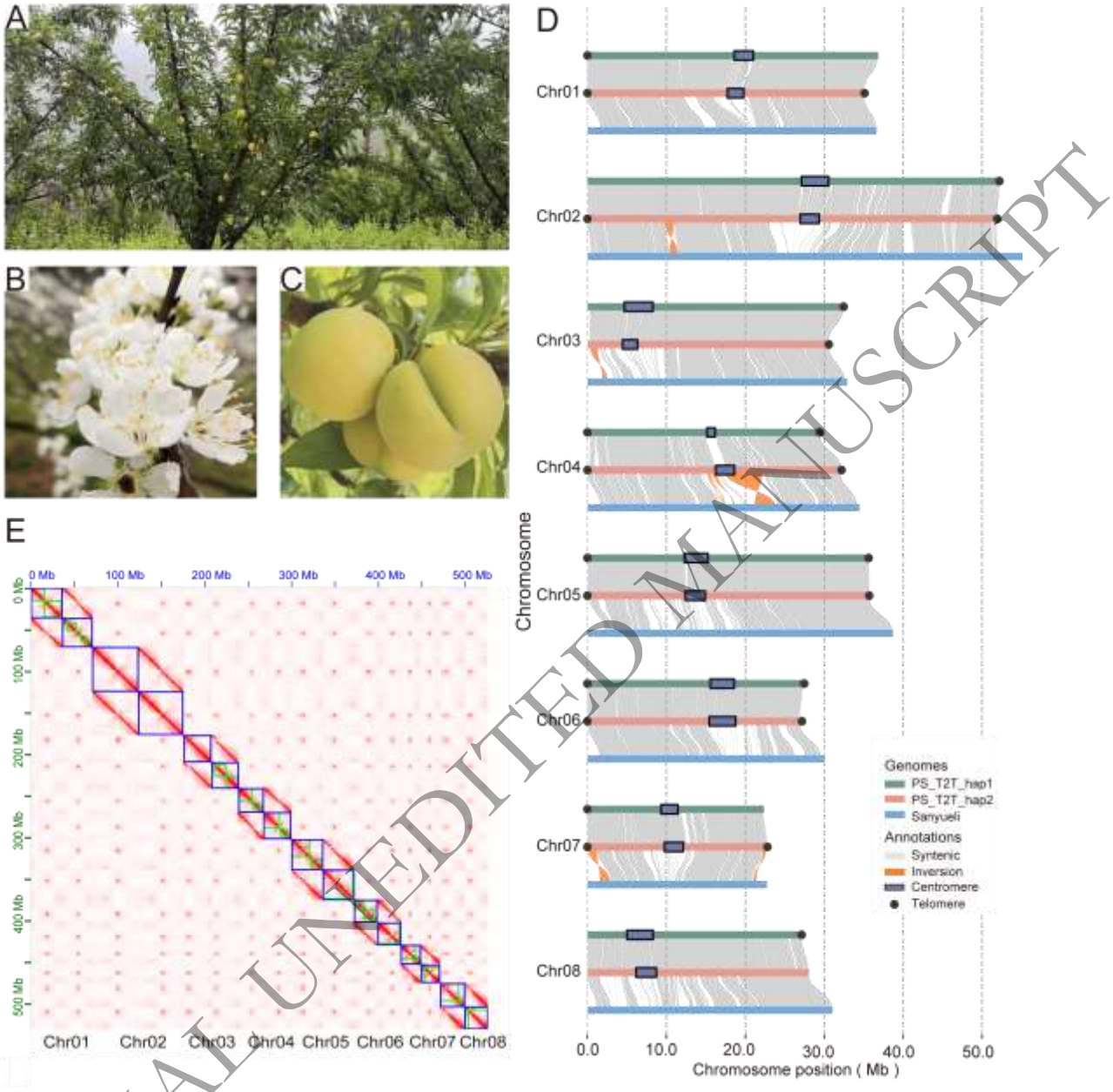
demarcates the region (Chr2: 16.92-16.94 Mb) believed to be under selection.

Figure 5. Convergent and divergent selection signals in the APPM complex. (A) Venn diagram of genes which were calculated in the top 1% CLR regions detected by SweeD in apricot, mei, plum, and peach. (B) Shared GO terms enriched in the top 1% CLR regions detected by SweeD. GO terms with a P-value < 0.01 (Fisher's exact test) are indicated with two stars. (C) Phylogenetic tree of functional haplotype sequences in highlighted genes shared across four clades: chromosome 7 (*PG*). (D) Visualization of variant positions above the *PG* gene model: the black line represents the genome, and rectangles represent exons. (E) Population frequencies of *PG* haplotype clades in the Apricot-Peach-Plum-Mei (APPM) Complex. (F-I) The dashed lines mark the regions at the top 1%. The red italicized text indicates functional genes common to all four clades, while purple, blue, orange, and green represent functional genes unique to apricot, mei, plum, and peach, respectively.

854 **Table 1 Statistics for the *Prunus salicina* cv. 'Fengtangli' genome assembly and**
855 **annotation.**

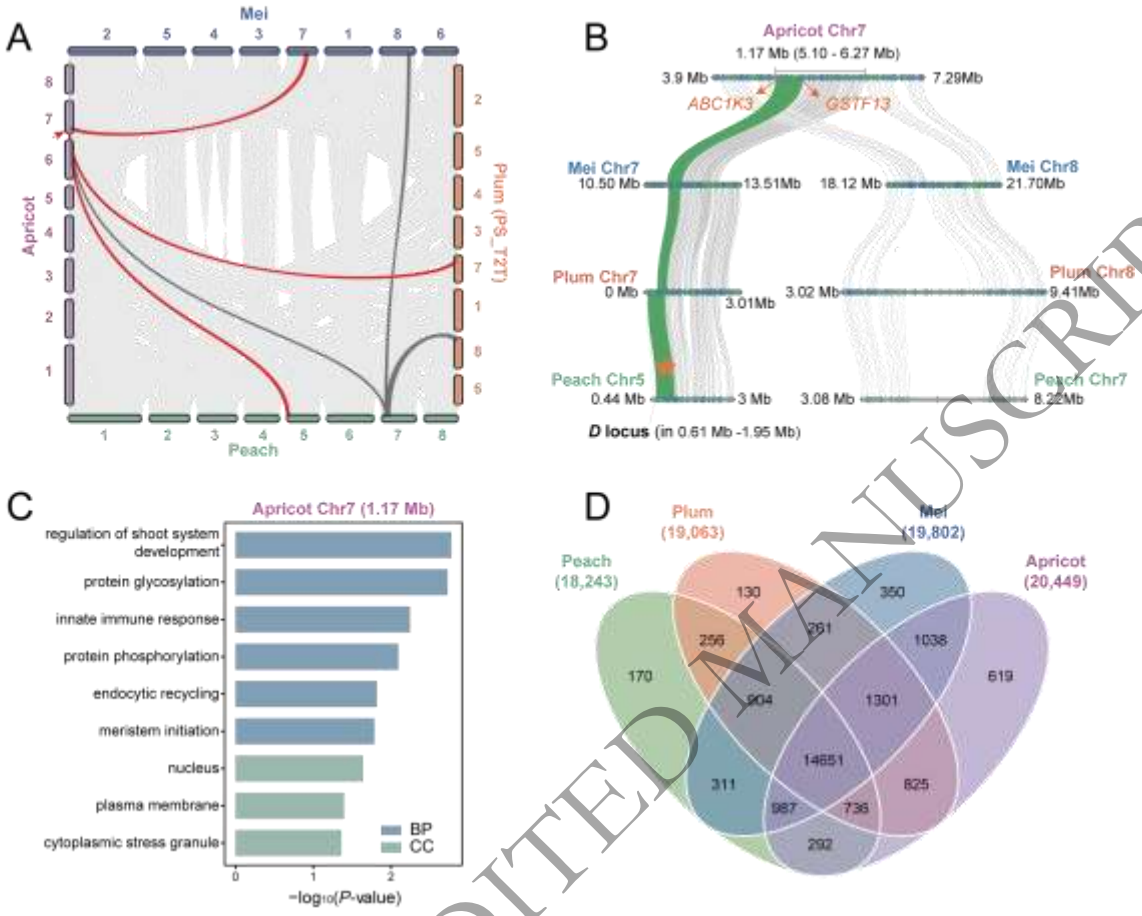
Genomic features	PS_T2T_hap1	PS_T2T_hap2	Sanyueli
Estimated genome size (Mb)	217.09	217.09	309.35
Assembled genome size (Mb)	251.25	251.29	282.38
Number of contigs (≥1 kb)	269	165	502
N50 contig length (bp)	20,111,805	19,658,610	1,370,198
BUSCOs of contig	99.1	98.7	-
Largest contig (bp)	52,178,471	51,917,112	8,384,049
Complete BUSCOs (%)	99.2	98.9	98.64
Complete and single-copy BUSCOs (%)	97.5	97	93.68
Complete and duplicated BUSCOs (%)	1.7	1.9	4.96
Fragmented BUSCOs (%)	0.6	0.7	0.37
Missing BUSCOs (%)	0.2	0.4	0.99
Number of gaps	0	0	536
Number of scaffolds	8	8	18
N50 scaffold length (Mb)	32.46	32.18	34.46
Mapping rate by reads from short-insert libraries (%)	99.8	99.8	96.6
Telomeres	11/16	13/16	0/16
Centromere	8/8	8/8	0/8
Number of genes	28,775	28,139	29,712
TE (%)	45.41%	46.19%	-

Figure 1



856

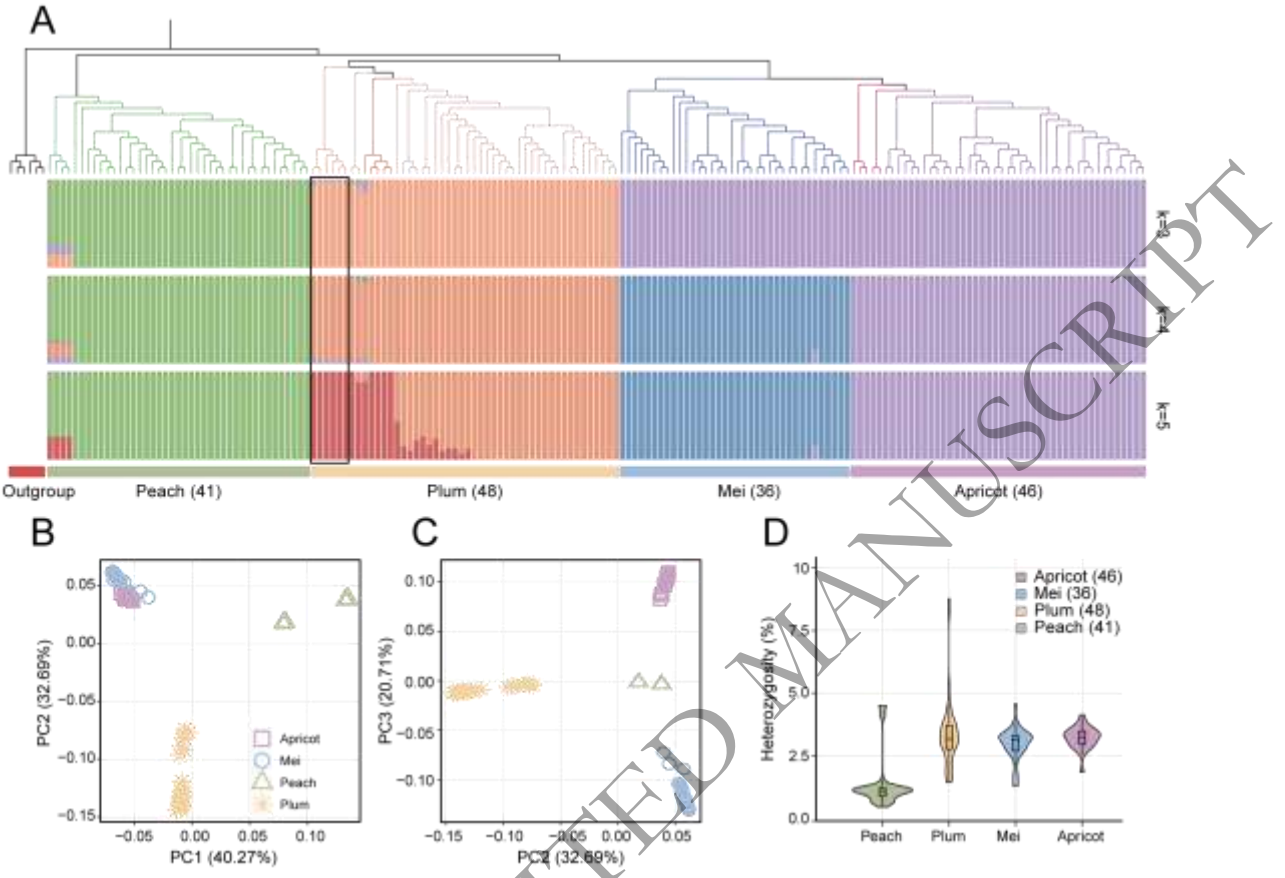
Figure 2



857

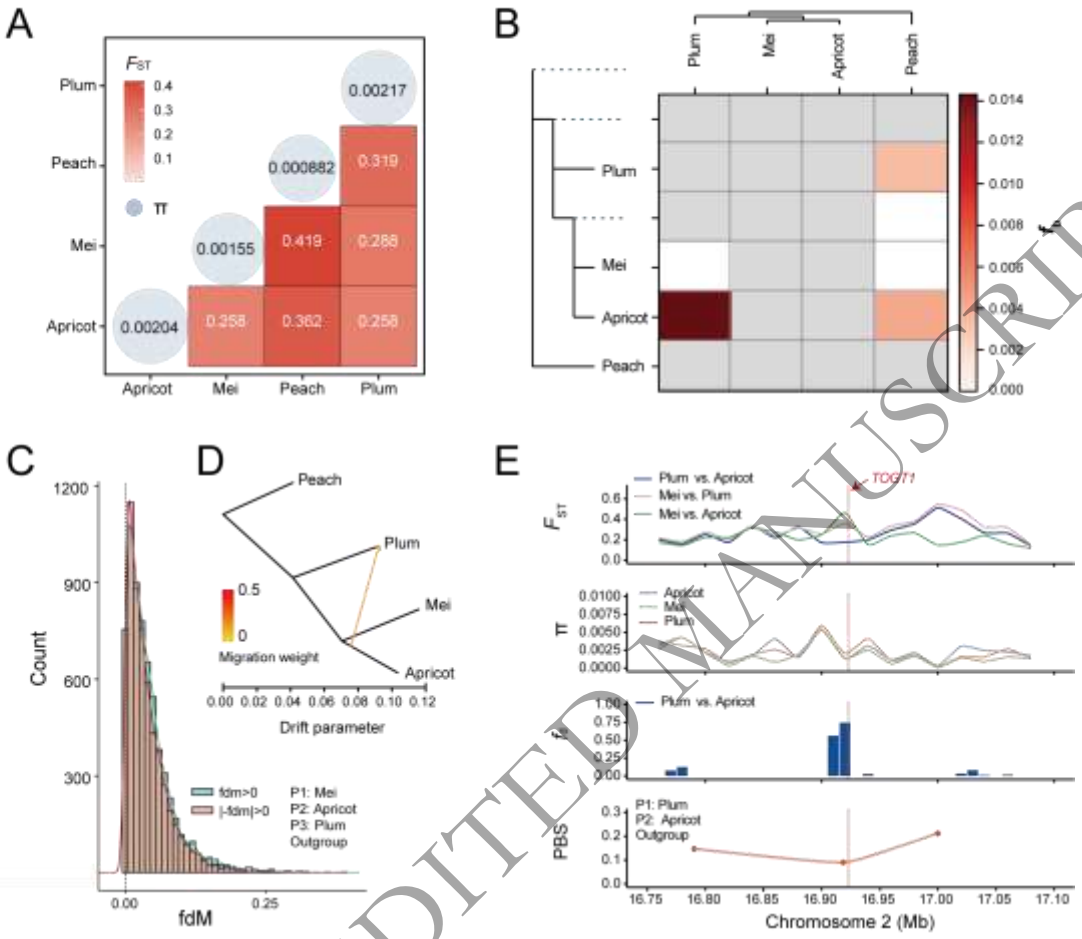
858

Figure 3



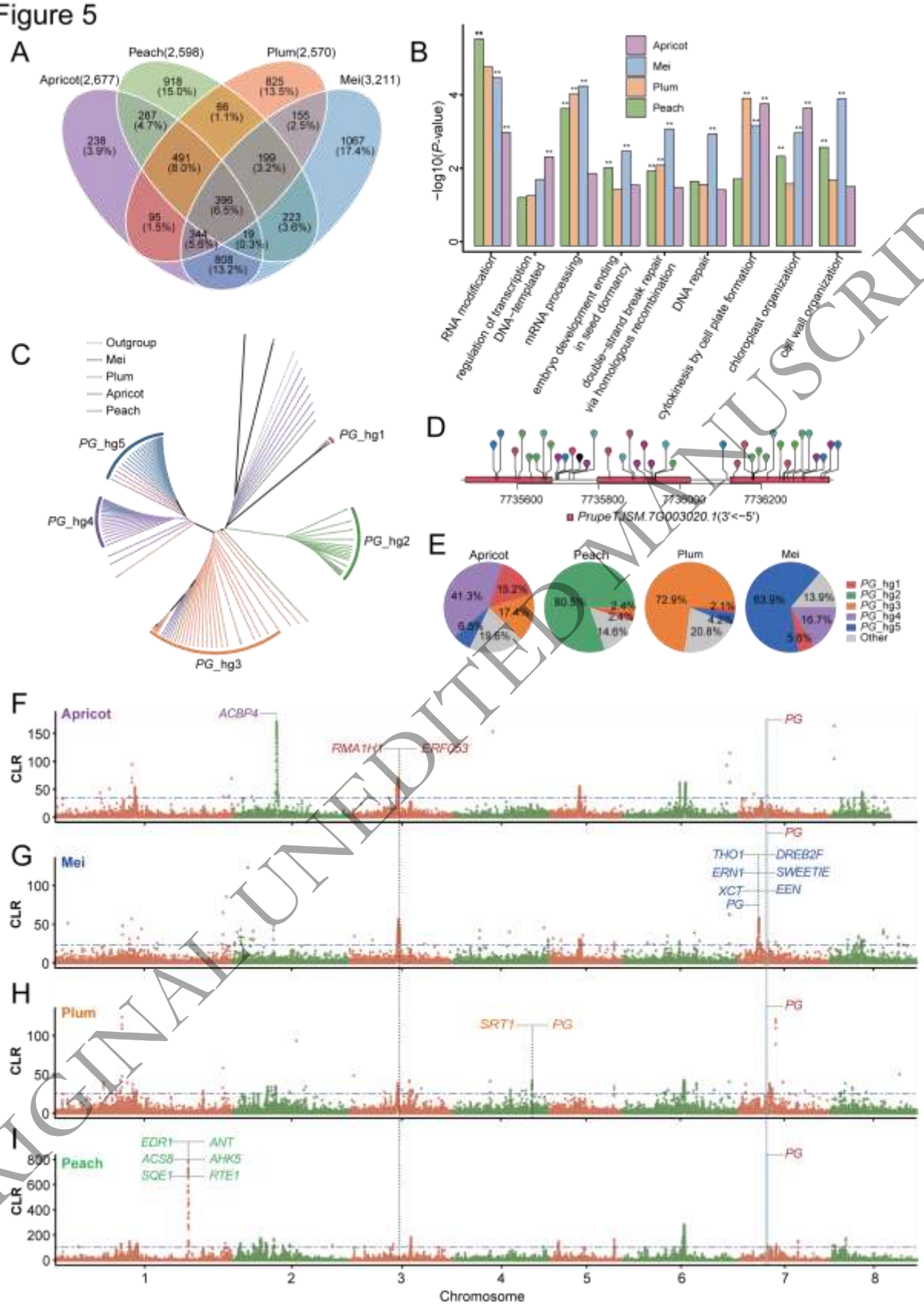
859

Figure 4



860

Figure 5



861

862

863