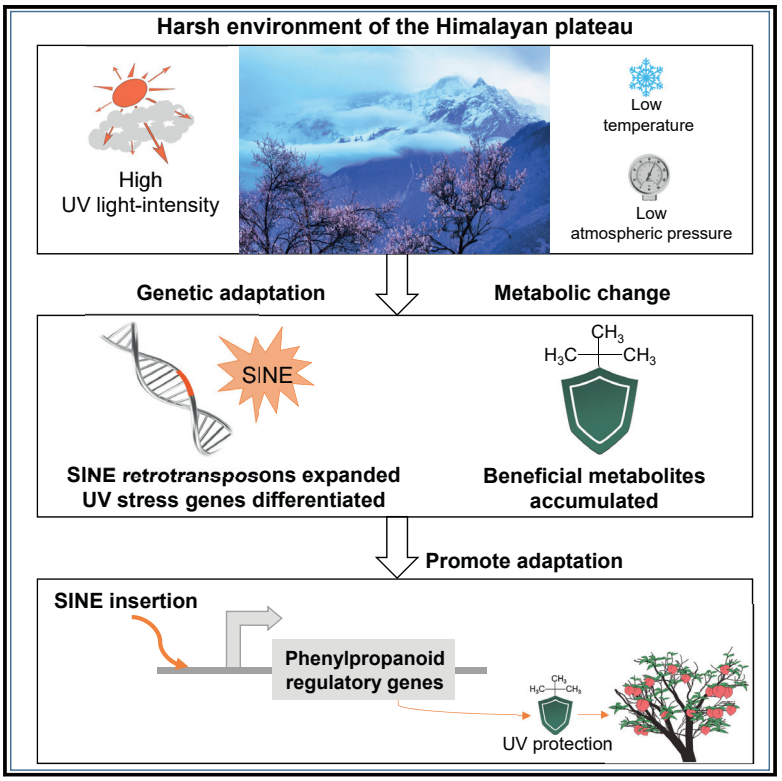
Article

Current Biology

Genomic basis of high-altitude adaptation in Tibetan *Prunus* fruit trees

Graphical abstract



Highlights

- Wild *Prunus* germplasm is collected from the high altitudes of the Himalayas
- SINE retrotransposons expand in the genomes of three Tibetan *Prunus* species
- UV response and phenylpropanoid metabolism associate with high-altitude adaptation
- Specific SINE insertions change the expression of altitude-related genes

Authors

Xia Wang, Shengjun Liu, Hao Zuo, ..., Xiuxin Deng, Xiuli Zeng, Qiang Xu

Correspondence

zengxiuli@taaas.org (X.Z.), xuqiang@mail.hzau.edu.cn (Q.X.)

In brief

The origin and development of plants in Himalayas is a fascinating topic. Wang et al. sequence genomes and determine metabolites of more than 300 *Prunus* accessions collected in this area. The results indicate that SINE transposons promote the adaptation of plants to high altitudes by affecting the nearby genes to enhance beneficial metabolites.

Wang et al., 2021, Current Biology *31*, 3848–3860
September 13, 2021
https://doi.org/10.1016/j.cub.2021.06.062

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

Genomic basis of high-altitude adaptation in Tibetan *Prunus* fruit trees

Xia Wang,[1,5,8] Shengjun Liu,[1,5,8] Hao Zuo,[1,5,8] Weikang Zheng,[1,5,8] Shanshan Zhang,[2,4,8] Yue Huang,[1,5] Gesang Pingcuo,[2,4] Hong Ying,[2,4] Fan Zhao,[2,4] Yuanrong Li,[2,4] Junwei Liu,[1,5] Ting-Shuang Yi,[6] Yanjun Zan,[7] Robert M. Larkin,[1] Xiuxin Deng,[1,3,5] Xiuli Zeng,[2,4,*] and Qiang Xu[1,3,5,9,*]

[1]Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan 430070, China
[2]Qinghai-Tibet Plateau Fruit Trees Scientific Observation Test Station (Ministry of Agriculture and Rural Affairs), Lhasa, Tibet 850032, China
[3]Hubei Hongshan Laboratory, Wuhan 430070, China
[4]Institute of Vegetables, Tibet Academy of Agricultural and Animal Husbandry Sciences, Lhasa, Tibet 850002, China
[5]Key Laboratory of Horticultural Crop (Fruit trees) Biology and Genetic Improvement (Ministry of Agriculture and Rural Affairs), Huazhong Agricultural University, Wuhan 430070, China
[6]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
[7]Department of Forestry Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå 90736, Sweden
[8]These authors contributed equally
[9]Lead contact
*Correspondence: zengxiuli@taaas.org (X.Z.), xuqiang@mail.hzau.edu.cn (Q.X.)
https://doi.org/10.1016/j.cub.2021.06.062

SUMMARY

The Great Himalayan Mountains and their foothills are believed to be the place of origin and development of many plant species. The genetic basis of adaptation to high plateaus is a fascinating topic that is poorly understood at the population level. We comprehensively collected and sequenced 377 accessions of *Prunus* germplasm along altitude gradients ranging from 2,067 to 4,492 m in the Himalayas. We *de novo* assembled three high-quality genomes of Tibetan *Prunus* species. A comparative analysis of *Prunus* genomes indicated a remarkable expansion of the SINE retrotransposons occurred in the genomes of Tibetan species. We observed genetic differentiation between Tibetan peaches from high and low altitudes and that genes associated with light stress signaling, especially UV stress signaling, were enriched in the differentiated regions. By profiling the metabolomes of Tibetan peach fruit, we determined 379 metabolites had significant genetic correlations with altitudes and that in particular phenylpropanoids were positively correlated with altitudes. We identified 62 Tibetan peach-specific SINEs that colocalized with metabolites differentially accumualted in Tibetan relative to cultivated peach. We demonstrated that two SINEs were inserted in a locus controlling the accumulation of 3-*O*-feruloyl quinic acid. SINE1 was specific to Tibetan peach. SINE2 was predominant in high altitudes and associated with the accumulation of 3-*O*-feruloyl quinic acid. These genomic and metabolic data for *Prunus* populations native to the Himalayan region indicate that the expansion of SINE retrotransposons helped Tibetan *Prunus* species adapt to the harsh environment of the Himalayan plateau by promoting the accumulation of beneficial metabolites.

INTRODUCTION

The Great Himalayan Mountains and their foothills are believed to be the centers of origin and genetic diversity for many cultivated plant species.[1,2] The genetics of adaptation to the extreme Himalayan climate have been reported for several animal species.[3–11] However, for plants, especially perennials, the genetic basis of adaptation to such conditions is largely unknown. The low-altitude Himalayas are an optimal agro-climatic zone for the production of tree species.[12,13] Particular primitive *Prunus* species were uniquely distributed in this region.[14,15]

Prunus L. (Rosaceae) consists of over 200 species that include economically important fruit crops,[16] such as peach (*Prunus persica*), almond (*P. dulcis*), plum (*P. salicina*), apricot (*P. armeniaca*), Mei (*P. mume*), and sweet cherry (*P. avium*). *Prunus* species are widely distributed in the temperate zone of the Northern Hemisphere and in the subtropical and tropical forests of Asia, Africa, South America, and Australia.[16] Wild peach, plum, apricot, and Mei (*P. mira*, *P. salicina* "Tibet," *P. armeniaca* "Tibet," and *P. mume* "Tibet") were found on the Tibetan Plateau during long-term field investigations. All cultivated peaches (*P. persica* and *P. ferganensis*) and wild peach species (*P. mira*, *P. davidiana*, *P. tangutica*, *P. mongolica*, and *P. kansuensis*) belong to the *Persica* section of the subgenus *Amydalus*.[17] *P. dulcis* belongs to the *Amygdalus* section of the subgenus *Amygdalus*.[17] Tibetan peach (*P. mira*) is probably an ancient progenitor of cultivated peach[18,19] and is endemic to the middle- to high-altitude regions of the Himalayas, at approximately 2,000

to 4,500 m above sea level (a.s.l.).[20] The population history and genetic diversity of Tibetan peach was shaped by the harsh environments of the Tibetan Plateau as Tibetan peach colonized the region and expanded its range.[18,21]

Natural selection acting on phenotypes and their plasticity results in evolution of populations and local adaptation.[22,23] Increasing evidence has demonstrated that altitudinal gradients are an informative environmental factor that can be used to investigate responses of plants to high-altitude climates.[24] For instance, populations of *Arabidopsis thaliana* native to different altitudes in the western Himalayas respond differently to common garden environments.[25] A comparison of the genomes from two *Eutrema* species, one high-altitude species from the eastern Qinghai-Tibet Plateau and the other from the lowlands, indicated that genes related to reproduction, DNA damage repair, and cold tolerance were specifically duplicated in the high-altitude species.[26] Recent work on Tibetan semi-wild wheat indicated that high-altitude environments can trigger the extensive reshaping of its genome.[27]

Plants have evolved multiple metabolic pathways that play vital roles in their responses to abiotic stress, such as light[28] and low temperature stress.[29] A metabolite-based genome-wide association study of qingke, a domesticated Tibetan highland barley, demonstrated that various phenylpropanoids were co-selected with particular varieties that are more tolerant to UV-B stress.[30] These findings provide evidence for the metabolic basis of the response to intense light in this species. Physiological and metabolomic studies of alpine plants have also demonstrated that adaptation strategies for survival at high altitudes involve changes in hormone synthesis and signal transduction in *Pedicularis punctata* and *Plantago major* in the Himalayan region[31] and in *Herpetospermum pedunculosum*[32] and *Potentilla saundersiana*[33] in the northwestern Tibetan Plateau.

The expansion of transposable elements (TEs) was reported to drive ecological adaptation.[34,35] For example, the proliferation of *BARE*-1 retrotransposons in wild barley grown in dry environments was correlated with an increase in genome size.[36] In apple, a major burst of retrotransposon activity that occurred approximately 21.0 million years ago (Mya) coincided with the uplift of the Tianshan Mountains, which is the postulated center of origin of apples.[37] In *Crucihimalaya himalaica*, a close relative of *Arabidopsis* that is ecologically adapted to high altitudes, LTR retrotransposons proliferated shortly after the dramatic uplift of the Himalayas that led to climatic change from the Late Pliocene to the Pleistocene.[38]

The genetic basis of the adaptation of the Tibetan peach and *Prunus* species to the harsh environment of the Himalayas has remained unexplored. In this study, we generated chromosome-scale genomes of the Tibetan *Prunus* species. Additionally, we performed a comparative genomics and population analysis of Tibetan peach accessions from a continuous series of altitudes to investigate the genetic and metabolic basis underpinning the adaptation to high altitudes for Tibetan peach.

RESULTS

Collection of *Prunus* germplasm and survey of their habitual environments on the Tibetan Plateau

From 2017 to 2019, we conducted a survey of Tibetan *Prunus* germplasm around the Himalayan region, including Lhasa (3,567–4,492 m a.s.l.), Nyingchi (2,118–3,494 m a.s.l), Xigaz (3,006–3,806 m a.s.l.), and Shannan (2,744–4,033 m a.s.l.) (Figure S1; Data S1A). A total of 377 *Prunus* accessions that included 346 peach accessions (299 *P. mira*, 44 *P. persica*, 2 *P. ferganensis*, and 1 *P. davidiana*), 22 *P. avium* accessions, 7 *P. armeniaca* "Tibet" accessions, 1 *P. mume* "Tibet" accession, and 1 *P. salicina* "Tibet" accession were collected in this study (Figure S2; Data S1A, S2, and S3). The locations of the sampling sites in Tibet are shown in Figure 1A (Data S1A).

To analyze environmental characteristics at different altitudes in the Himalayan region, data on 76 meteorological variables, including sunshine, temperature, humidity, and atmospheric pressure-related variables, were collected from different altitudes. The intensity of sunlight was positively correlated with altitude. Atmospheric pressure, temperature, and humidity were negatively correlated with altitude. In regions located between 4,000 m and 4,800 m a.s.l., the annual duration of sunshine was as high as 2,900 h (Data S1B). In contrast, in regions located between 2,300 m and 3,000 m a.s.l., the annual duration of sunshine was 1,727 h (Data S1B), which is significantly less.

Pan-genome of *Prunus*

Three high-quality genomes of Tibetan *Prunus* species were *de novo* assembled. Tibetan peach (*P. mira*) is the highest quality genome among the currently available *Prunus* genomes, with seven gaps per chromosome on average and thus serves as a reference genome for *Prunus*. The assembled genome of Tibetan peach is 242.67 Mb with a contig N50 of 12.14 Mb, accounting for 97% of the estimated genome size (Data S1C and S1D). We also *de novo* assembled the genomes of *P. mume* "Tibet" and *P. armeniaca* "Tibet," with assembly sizes of 241.72 Mb and 266.25 Mb, respectively, and contig N50 of 3.35 Mb and 1.75 Mb, respectively (Data S1C). We annotated 27,270, 31,116, and 28,973 gene models for *P. mira*, *P. mume* "Tibet," and *P. armeniaca* "Tibet," respectively (Data S1C).

Seven published high-quality genomes of *Prunus* species, including two cultivated peaches (*P. persica*[39] and *P. ferganensis*[21]), a species closely related to peach (*P. dulcis* var. Texas[40]), plum (*P. salicina* var. Zhongli No. 6; https://www.rosaceae.org/Analysis/9019655), Mei (*P. mume*[41]), apricot (*P. armeniaca*[42]), and cherry (*P. avium*[43]) were selected as representatives of non-Tibetan *Prunus* species. Three genome assemblies from this study (*P. mira*, *P. mume* "Tibet" and *P. armeniaca* "Tibet") and genotype information from *P. salicina* "Tibet" based on DNA sequencing were used as representatives of Tibetan *Prunus* species. In addition, two representative Rosaceae species (*Fragaria vesca*[44] and *Rubus occidentalis*[45]) and *Vitis vinifera*[46] were also used. A phylogenetic tree based on the single-copy genes shared by these species showed that the wild Tibetan accessions and cultivars of four *Prunus* species formed four separate sister pairs, namely *P. mira* versus *P. persica* and *P. ferganensis*, *P. salicina* "Tibet" versus *P. salicina*, *P. armeniaca* "Tibet" versus *P. armeniaca*, and *P. mume* "Tibet" versus *P. mume* (Figure 1B; Data S4). Molecular dating based on fossils[47–49] showed that the sampled *Prunus* species diverged approximately 27.6 Mya (Figure 1B) and thus the divergence of *Prunus* species coincided with the rapid elevation of the Himalayas from 1,000 m to 2,300 m.[50] The peach lineage diverged from almond (*P. dulcis*) during the

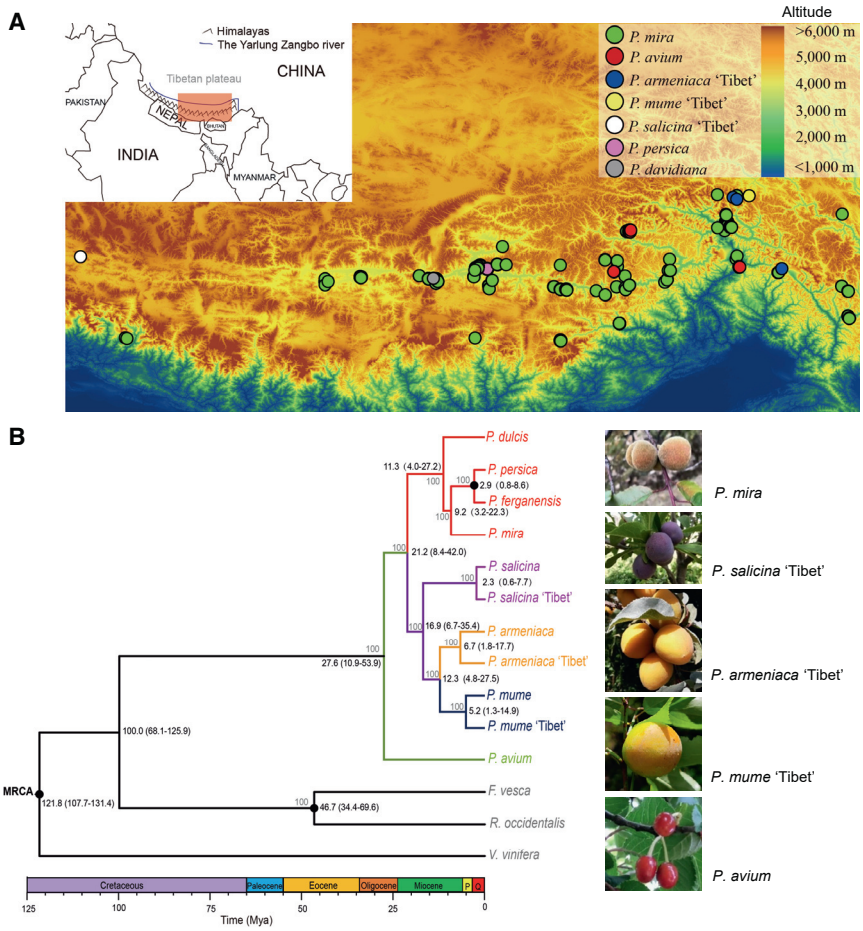


Figure 1. Locations and phylogeny of *Prunus* germplasm collected on the Tibetan Plateau

(A) Geographic distribution of the Tibetan *Prunus* accessions collected in this study. Accessions are represented with different colored dots that indicate different species categories. The altitudes of different sites in the region are indicated with different colors.

(B) Maximum likelihood (ML) tree and estimated divergence times of the Tibetan and cultivated *Prunus* taxa. The ML tree was inferred from a matrix comprising 2,589 shared single-copy genes. Bootstrap values are indicated along the branches. The divergence times at the nodes were estimated using three fossil calibrations indicated with solid black dots. Median age estimates and 95% highest posterior densities (Ma) are shown for each node. Q and P represent the Quaternary Period and the Pliocene Epoch, respectively. Pictures of the fruits (from top to bottom) from Tibetan peach, Tibetan plum, Tibetan apricot, Tibetan Mei, and Tibetan cherry are shown at the right. See also Figures S1 and S2, Data S1A, S1B, S2, S3, and S4.

Miocene (approximately 11.3 Mya; Figure 1B). Cultivated peach (*P. persica*) diverged from Tibetan peach (*P. mira*) approximately 9.2 Mya (Figure 1B).

We performed pan-genome analyses using the above-mentioned 10 high-quality *Prunus* genomes. The *Prunus* pan-genome that we constructed was 350.95 Mb and contained 29,017 protein-coding genes. A total of 12,239 core gene families shared by ten *Prunus* genomes were identified (Figure 2A; Data S1E). The gene families that were absent from at least one species of peach, plum, Mei, apricot, and cherry were defined as dispensable gene families (Figure 2A; Data S1E). The genomes of the Tibetan and cultivated *Prunus* species were comparatively analyzed (Figure 2B and S3; Data S1F). Regarding the presence/absence variation (PAV), we identified 6,434 insertions longer than 50 bp in Tibetan peach relative to cultivated peach that were 9.33 Mb in length and accounted for 6.5% of the Tibetan peach genome. We found a total of 199 genes in genomic regions that are specific to the Tibetan peach (Data S1G and S1H). These genes mainly contribute to DNA repair, response to DNA damage stimulus, response to UV-C, and response to fungus. A comparison of DNA methylation levels in Tibetan peach and cultivated peach indicated that these genomes contained 5,728 differentially methylated regions (DMRs) and that these DMRs tended to co-exist with PAVs. The Spearman's rank correlation coefficient between the number of DMRs and the length of the PAVs within each one Mb window was 0.56 (p value = 1.63×10^{-20}; Figure 2B; Data S1F). Remarkably, 67.65% of the sequences that were present in Tibetan peach and absent in cultivated peach were TEs, mainly DNA-EnSpm (22.85%) and LTR-Gypsy TEs (6.91%; Data S1I). We found that the proportion of short interspersed nuclear elements (SINEs) in the *P. mira*-specific regions was 0.30% (Data S1I). Conversely, we found a much lower proportion of SINEs in the *P. persica*-specific regions (0.01%; Data S1I).

Expansion of SINE retrotransposons in Tibetan *Prunus* genomes

We compared the types and contents of TEs found in Tibetan species to cultivated species because TEs were enriched in regions containing PAVs in these species. We compared three pairs of species that included peach, Mei, and apricot species. The SINE-type TE content was remarkably higher in the Tibetan genomes (0.55%, 0.69%, and 0.42% of the genomes for Tibetan peach, Tibetan Mei, and Tibetan apricot, respectively) relative to the corresponding cultivated genomes (0.20%, 0.42%, and 0.06% of the genomes for cultivated peach, Mei, and apricot, respectively) (Data S1J). In contrast, for other types of TEs, we did not observe significant differences in the overall content in Tibetan species relative to the repsective cultivated species (Data S1J).

Members of the SINE family were classified into canonical SINEs and noncanonical SINEs based on their conserved domains. Tibetan peach maintained a high content of noncanonical

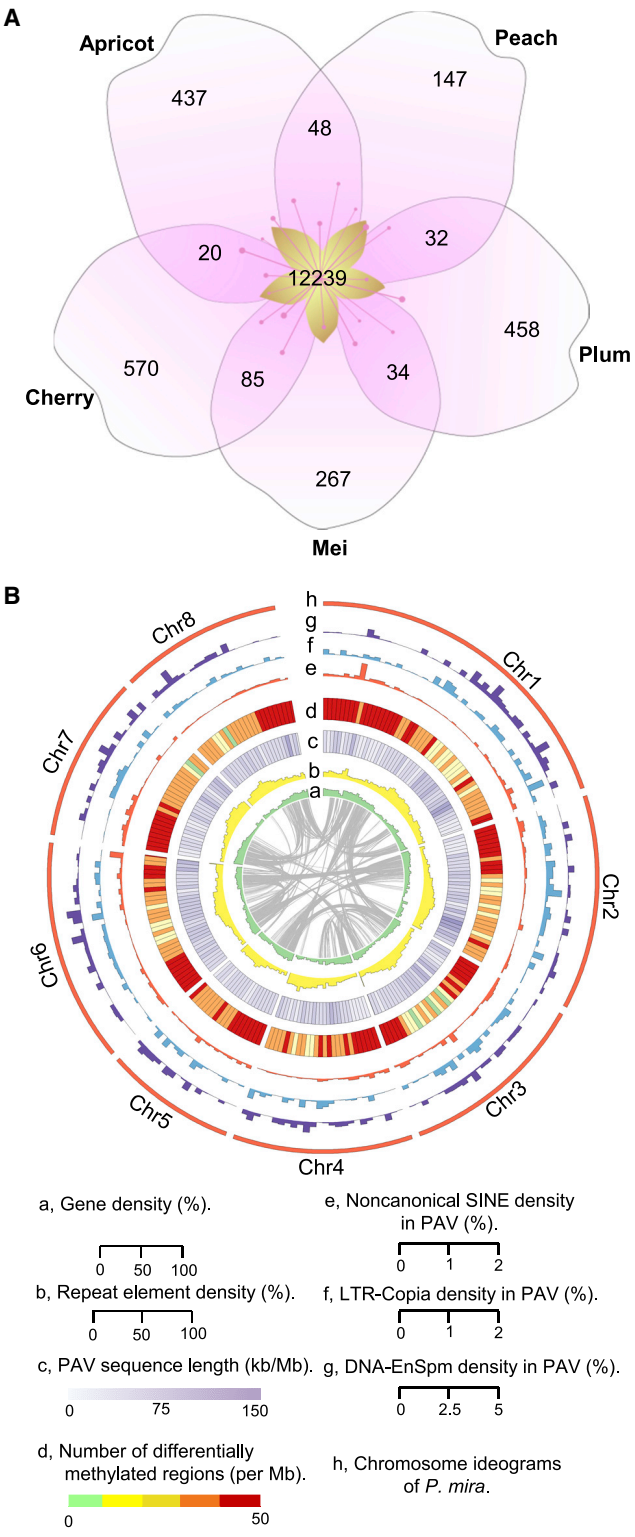


Figure 2. Comparative genomic analysis of *Prunus* species

(A) Numbers of core and dispensable gene families in representative *Prunus* species. Peach (*P. mira*, *P. persica*, and *P. ferganensis*), plum (*P. salicina*), Mei (*P. mume* "Tibet" and *P. mume*), cherry (*P. avium*), and apricot (*P. armeniaca* "Tibet" and *P. armeniaca*) were used as representative *Prunus* catagories.

SINEs (Data S1K). Indeed, the noncanonical SINEs in Tibetan peach expanded 11.64-fold relative to almond, a species closely related to peach, and 2.76-fold relative to cultivated peach (Figure 3A; Data S1K–S1M).

The 2,218 noncanonical SINE insertions were unevenly distributed throughout the Tibetan peach genome (Figure 2B; Data S1K). There were three hotspots on chromosome 1, chromosome 3, and chromosome 6 (Figure 2B; Data S1K). The CG methylation levels in 2-kb regions that flanked noncanonical SINEs were substantially decreased in Tibetan peach relative to cultivated peach (Figure 3B; Data S1N). We estimated the effects of noncanonical SINE insertions on gene expression within 10-kb regions and found that there was a dramatic increase in the expression of genes around 2-kb distance from the insertion sites of the noncanonical SINEs relative to the other TEs (Figure 3C; Data S1O).

Genetic differentiation of Tibetan peach populations at high and low altitudes

A total of 388 accessions, including 304 Tibetan peaches, 56 cultivated peaches, 8 wild peach accessions (2 *P. kansuensis*, 4 *P. davidiana*, 1 *P. tangutica*, and 1 *P. mongolica*), 15 *P. dulcis* and 5 *P. ledebouriana* accessions, were used in a population analyses (Data S1A). A total of 21,350,006 SNPs were identified in this population (Data S1P). A SNP-based phylogenetic tree indicated an ancient phylogenetic status for Tibetan peaches that is consistent with their wild habitats (Figure 4; Data S5). The cultivated peach population had 1.55-fold more putative deleterious mutations relative to Tibetan peach (Figures S4A and S4B; Data S1Q).

Based on the principal component analyses (PCAs) of genome-wide SNPs, the Tibetan peach accessions were divided into two groups separated by 3,500 m (Figure 5A; Data S1R). A pool of 66 Tibetan peach accessions collected from relatively high altitudes (3,800–4,492 m a.s.l.) and a pool of 67 Tibetan peach accessions collected from relatively low altitudes (2,067–3,200 m a.s.l.) were used to detect signals associated with the adaptation to high altitudes (Data S1A).

The levels of genetic divergence between the high- and low-altitude populations across the genome were uneven (Figure 5B; Data S1S). We identified genomic regions that were highly divergent at both the single-base level (F_{ST} > 0.18) and the haplotype level (hapFLK > 1.39; Figure S4C and Data S1S). On the basis of this pairwise comparison, we concluded that a total of 9.02 Mb of genomic regions containing 1,368 genes probably contributed to the adaptation to the high-altitude environment (Data S1T). The significantly divergent genes were enriched in functions associated with the regulation of response to stimuli, especially the responses to light and radiation, signal transduction, pollen-pistil recognition, stomatal opening, and metabolic processes of aminoglycans (Figure 5C; Data S1U). Furthermore, 2.77 Mb of genomic regions that harbored 394 genes with both significantly high genetic divergence (F_{ST} > 0.18) and extended haplotype

(B) Landscape of presence/absence variation (PAV) and differential methylation between Tibetan peach and cultivated peach. The PAV indicates specific regions that are present in Tibetan peach and absent from cultivated peach. The lines in the center of the circle indicate pairs of homologous genes on different chromosomes of Tibetan peach. See also Figure S3 and Data S1E–S1I.

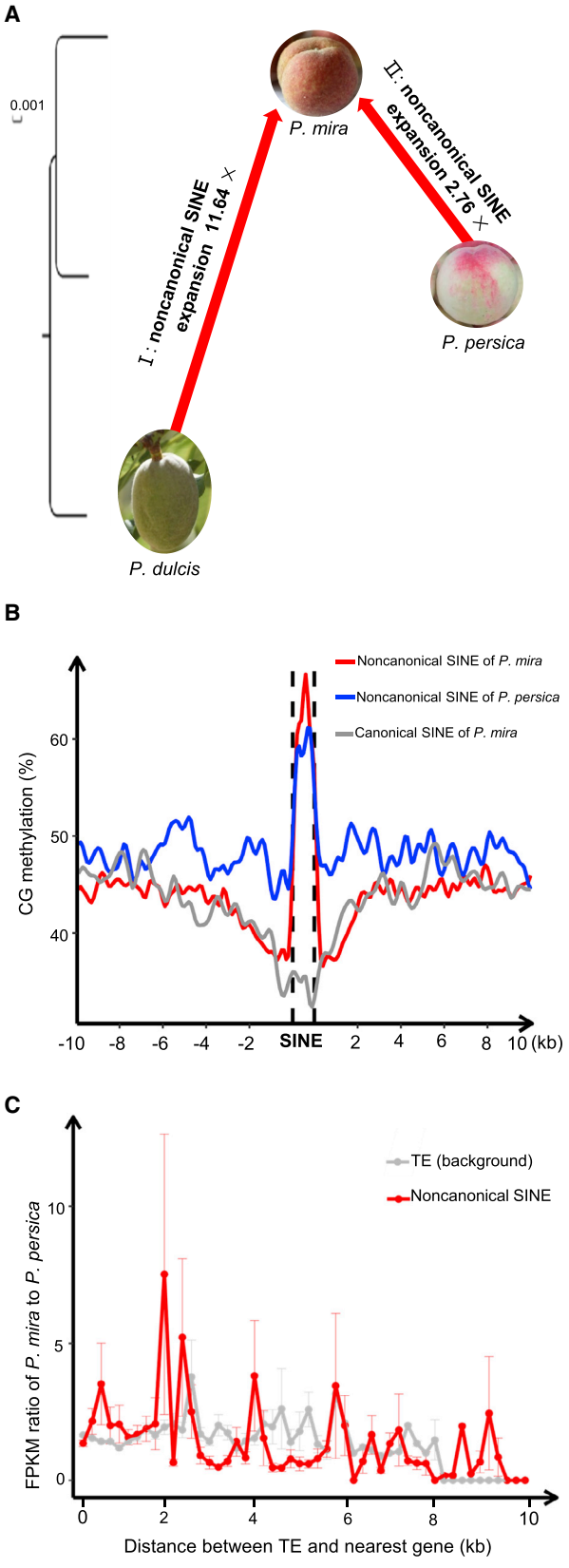


Figure 3. SINE expansion in Tibetan peach

(A) Expansion of noncanonical SINEs in Tibetan peach (*P. mira*) realtive to almond (*P. dulcis*) and cultivated peach (*P. persica*). A ML phylogenetic tree containing Tibetan peach, cultivated peach, and almond is shown (left).

(B) Distribution of DNA methylation levels in SINE elements in Tibetan and cultivated peach. Average CG methylation levels were calculated in 10 intervals in the SINE region and in 100 intervals in the regions that were located 10 kb upstream and downstream of the SINE region.

(C) Influence of noncanonical SINEs and other TE insertions on the expression of nearby genes in Tibetan peach (*P. mira*) relative to cultivated peach (*P. persica*). The 10-kb regions flanking noncanonical SINEs and background TEs were divided into 50 equally long bins. The data are represented as mean ± SD. See also Data S1K–S1O.

homozygosity (XP-EHH > 1.39) were expected to be under positive selection in the high-altitude population relative to the low-altitude population (Figure 5B; Data S1V).

Remarkable genotypic divergence was observed on genes encoding the light signaling regulator FAR1/FHY3 between Tibetan peaches and cultivated peaches (Figure S4D; Data S1W). When we compared the DNA methylation levels of these genes, we found significant CHG-type hypomethylation in the gene bodies of members of the *FAR1/FHY3* gene family in Tibetan peach relative to cultivated peach (Figure S4E; Data S1X). We also found selection signals in the gene encoding theUV light receptor UVR8 in the high-altitude population (Data S1V). Moreover, the population-based environment-genotype association analysis revealed associations between annual sunshine duration and several UV-response related genes including genes that encode a UV-B-induced protein (*Pmira5g018050*), a RING-type E3 ubiquitin transferase (*Pmira5g005820*), and the DNA repair endonuclease UVH1 (*Pmira5g016550*) (Figures S4F–S4J; Data S1Y). Presumably, these genes help the high-altitude Tibetan peach to achieve full UV-B tolerance.[51,52] Meanwhile, comparative genomic analysis revealed that positively selected genes in Tibetan peach were enriched in DNA repair, response to DNA damage stimulus, and negative regulation of programmed cell death (Data S1Z). These data provide genetic evidence that the high light intensity and, in particular, the high fluence rate of UV light on the Tibetan Plateau helped to fix genetic changes in the high-altitude population.

Genetic basis of changes in metabolite levels in high altitudes

Metabolites were investigated because genes related to the response to light stress and UV-B radiation were differentiated between high- and low-altitude populations. Tibetan peaches produce fruit with wide variations in color, flavor, and other traits influenced by metabolism (Figure S1B; Data S1A). We quantified the levels of 1,768 metabolites in fruit from 319 accessions, including 275 Tibetan peach accessions and 44 cultivated peach accessions (Data S1AA). The PCA result based on the levels of 1,768 metabolites in the Tibetan peach population indicated that the metabolite constitutions of the fruit were largely congruent with the classification of ecotypes based on altitude (Figure S5A; Data S1AB).

A particular metabolite was proposed to facilitate adaptation to high altitude if variation in the levels of the metabolite was mostly determined by genetics and if the metabolite was significantly correlated with altitude. We partitioned the variation of

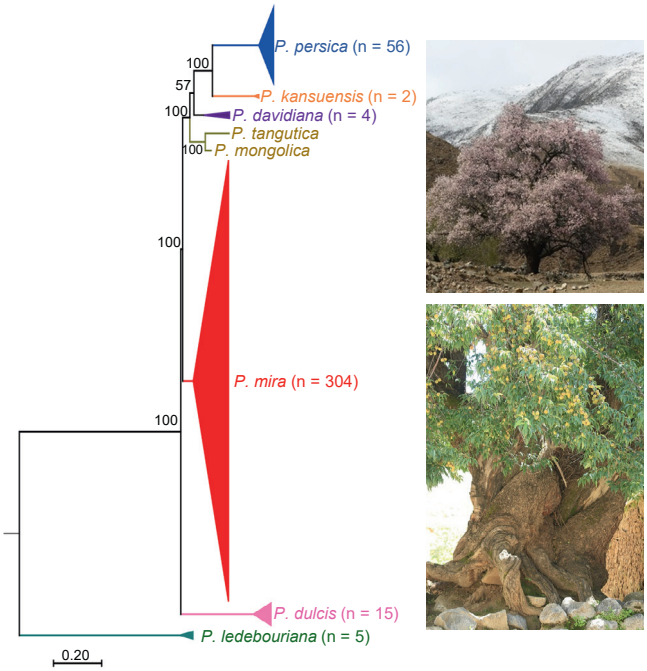


Figure 4. Phylogeny of Tibetan peach, cultivated peach, and closely related species

The maximum likelihood phylogenetic tree was constructed using 388 accessions of peach and closely related species. Bootstrap values are shown along the nodes. The number of accessions in each clade is indicated in parentheses. See also Figures S1 and S4 and Data S1P, S1Q, and S5.

each metabolite into contributions from genetic and environmental components by estimating heritability and polygenic score for each accession (Figure 6A and S5B). Calculating a polygenic score (i.e., a breeding value) involves aggregating the estimated effects of genome-wide variants to predict the contribution of an individual's genome to a phenotypic trait.[53] First, a set of 1,697 metabolites with variations that were dominated by genetic components ($h^2 > 0.1$) were selected (Data S1AA). Then, the correlation between the polygenic score of each metabolite and the altitude was evaluated by fitting a linear model. A total of 379 metabolites that yielded r^2 values > 0.25 and p values < 0.05 from the linear regression analysis were defined as metabolites that might contribute to the long-term response to high altitudes (Figure 6A; Data S1AA). The levels of 361 of these metabolites—including 67 annotated metabolites—were significantly correlated with altitude (the absolute value of Spearman's rank correlation coefficient, $|r| \geq 0.3$; Figure 6A; Data S1AA). Among the remaining 1,389 metabolites, 313 showed a significant correlation with altitude (absolute value of Spearman's rank correlation coefficient, $|r| \geq 0.3$), but their heritabilities were less (Data S1AA). Thus, we defined fluctuations in the levels of these metabolites as short-term responses to high altitude that promoted metabolic acclimation and proper development (Figure 6A; Data S1AA).

Phenylpropanoids and particular organic acids were overrepresented among these 379 high-altitude adaptation-related metabolites that accumulated to significantly higher levels at high altitudes relative to low altitudes. These metabolites included neohesperidin, hesperidin, 3-*O*-feruloyl quinic acid, chlorogenic acid, and shikimic acid (Data S1AA). In contrast, particular organic acids, lipids, and terpenes accumulated to significantly higher levels at low altitudes relative to high altitudes. These metabolites included rosmarinic acid, punicic acid, and roseoside (Data S1AA).

A set of 510,989 high-quality SNPs from 275 diverse Tibetan peach accessions was used in a metabolic GWAS (mGWAS) to determine the genetic basis underlying the metabolic variation. mGWAS association signals for the 379 high-altitude related metabolites were identified (Data S1AC). A total of 337 loci were detected for 150 metabolites based on the genome-wide significance threshold (Data S1AD). Among the hotspots for metabolites associated with high altitude, a region around 19 Mb on chromosome 3 was significantly divergent in the high-altitude and low-altitude Tibetan peach populations (Figures S6A and S6B; Data S1AE). This region is responsible for the variability in the levels of six quinic acid-related metabolites, including chlorogenic acid, neochlorogenic acid, coumarylglucaric acid, 3-*O*-feruloyl quinic acid, 6,7-dihydroxycoumarin 7-*O*-quinic acid, and 1-*O*-caffeoyl quinic acid (Figures 6B and S6C–S6G; Data S1AF–AG).

SINE insertions related to the accumulation of phenylpropanoids in high-altitude Tibetan peach

Regarding the increased proportion of SINE elements in the Tibetan peach genome relative to the cultivated peach genome, we identified 2,218 Tibetan peach-specific SINEs (Data S1AH). A total of 1,461 metabolites accumulated to different levels in Tibetan peach accessions relative to cultivated peach accessions (Data S1AA). Furthermore, we found that 62 Tibetan peach-specific SINEs co-localized with the mGWAS hits underlying the variation in the levels of 53 differentially accumulated metabolites (Figures S5C and S7A; Data S1AI). There was a 16 to 19 Mb region on chromosome 3 that was a remarkable hotspot for both SINE insertions and mGWAS loci associated with the accumulation of phenylpropanoids and flavonoids (Figure 7A).

Based on the genome sequencing data, we detected a Tibetan peach-specific SINE1 insertion that was 2.3 kb upstream of *Pmira3g06670*, which encodes a NAC transcription factor and contributed to the variation in the accumulation of 3-*O*-feruloyl quinic acid (Figures 7B and S7B; Data S1AJ and S6). We detected another SINE2 insertion that predominanted in high-altitude Tibetan peaches and was 1.5 kb upstream of *Pmira3g006670* (Figure 7B). Independent PCR-based experiments were performed to validate the SINE insertions. The SINE1 insertion was absent from all 16 of the cultivated peach accessions and was present in 15 of the 16 Tibetan peach accessions that were tested (Figures 7C, 7D, and S7C). Next, we used a PCR-based assay to survey the frequency of SINE2 insertions in 51 high-altitude Tibetan peaches and 29 low-altitude Tibetan peaches. Consistent with the genomic sequencing data analysis, we found that genotypes with SINE2 insertions were predominantly from the high-altitude Tibetan peach population (76.47%; Figures 7E, 7F, and S7D). In contrast, we found that the predominant genotype in low-altitude Tibetan peach population lacked this particular SINE insertion (68.97%; Figures 7E, 7F, and S7D).

The expression level of this candidate gene was significantly higher in the Tibetan peaches that harbor this SINE2 insertion

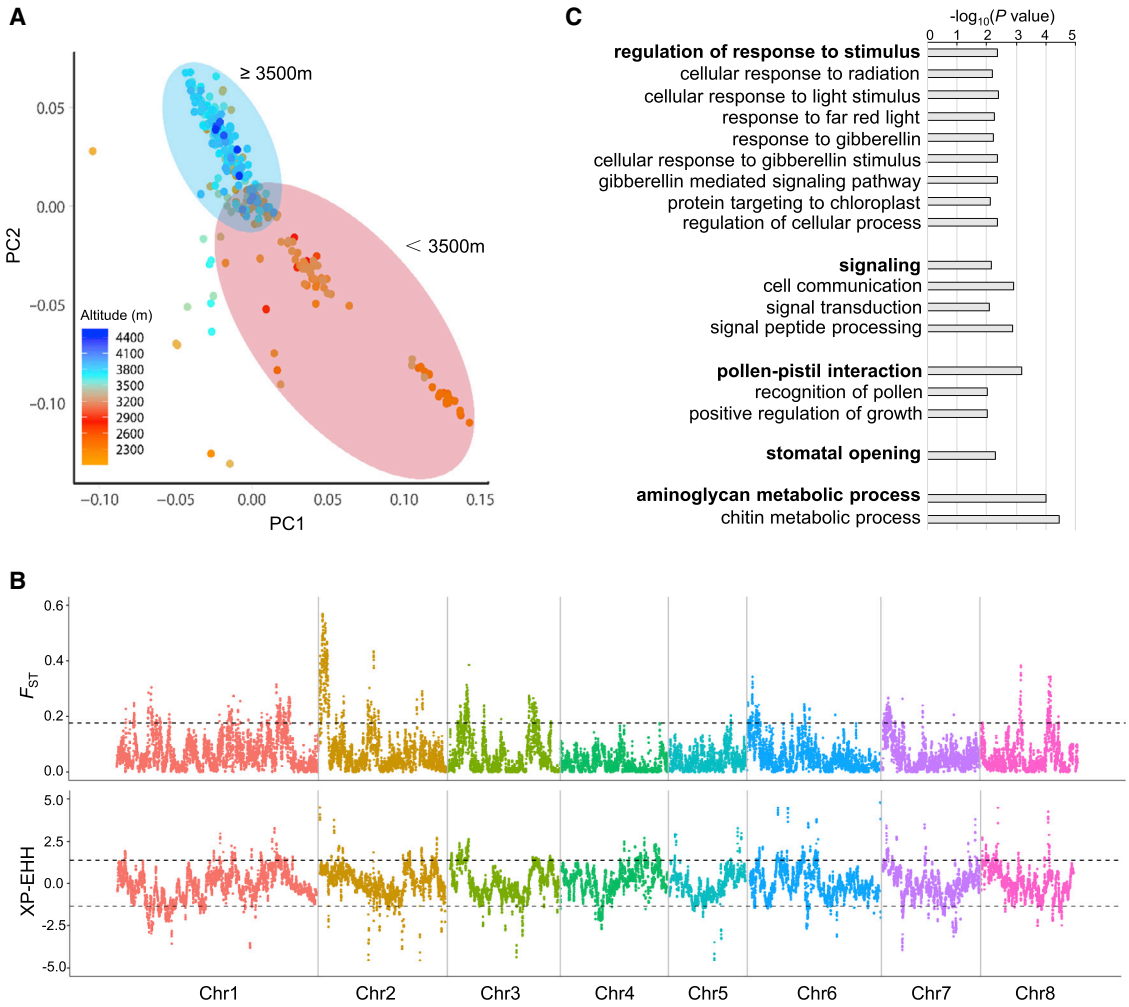


Figure 5. Comparison of high-altitude and the low-altitude Tibetan peach populations

(A) Principal component analysis (PCA) of SNPs from 304 Tibetan peach accessions. The altitudes of the different accessions are indicated with different colors. Blue circles include Tibetan peach accessions mostly from regions ≥3,500 m. Red circles include Tibetan peach accessions mostly from regions <3,500 m.

(B) Distribution of genetic differentiation (F_{ST}) and cross-population extended heterozygosity haplotype (XP-EHH) across the genomes of the high-altitude and low-altitude populations. A total of 66 Tibetan peach accessions collected from ≥ 3,800 m were treated as the high-altitude population. A total of 67 Tibetan peach accessions collected from ≤ 3,200 m were treated as the low-altitude population. F_{ST} and XP-EHH were calculated in 50-kb sliding windows with 10-kb steps. The regions above the dashed line in the F_{ST} value distribution are in the 5% right tail of the empirical distribution (F_{ST} is 0.18). The region above the dashed line in the distribution of XP-EHH corresponds to the 5% right tail of the empirical distribution (XP-EHH is 1.39).

(C) Gene Ontology (GO) enrichment analysis of genes in regions that were significantly differentiated in high-altitude relative to low-altitude Tibetan peach populations. Statistically significant enrichment was determined using the Fisher's exact test. p values are indicated. See also Figure S4 and Data S1R–S1Y.

relative to the Tibetan peaches that lack this SINE2 insertion (Figure 7G; Data S1AK). Moreover, the Tibetan peach accessions that harbor this SINE2 insertion accumulated significantly higher levels of 3-*O*-feruloyl quinic acid than Tibetan peach accessions that lacked this insertion (Figure 7H; Data S1AL). These results provide evidence that SINE retrotransposon polymorphism existed between high- and low-altitude peaches and probably affected the expression of nearby genes which can regulate the accumulation of phenylpropanoids.

DISCUSSION

Many plants are native to the region of Tibet and its southern and southeastern mountains. To the best of our knowledge, this is the first report on the collection and genetic analysis of a large natural population of plants that is continuously distributed across a broad range of altitudes on the Himalayan plateau. We *de novo* assembled high-quality genomes of Tibetan *Prunus* species and sequenced 377 accessions of *Prunus* germplasm. This dataset provides a rare gene bank for adaptation genomics and will contribute to the identification of adaptive loci that affect the levels of fruit metabolites. We found that SINE retrotransposons expanded in Tibetan *Prunus* species, and Tibetan peach-specific SINEs co-localized with altitude-associated metabolites, in the particular case of phenylpropanoids.

A survey of the meteorological variables in the Himalayan region showed significant correlations between light, temperature,

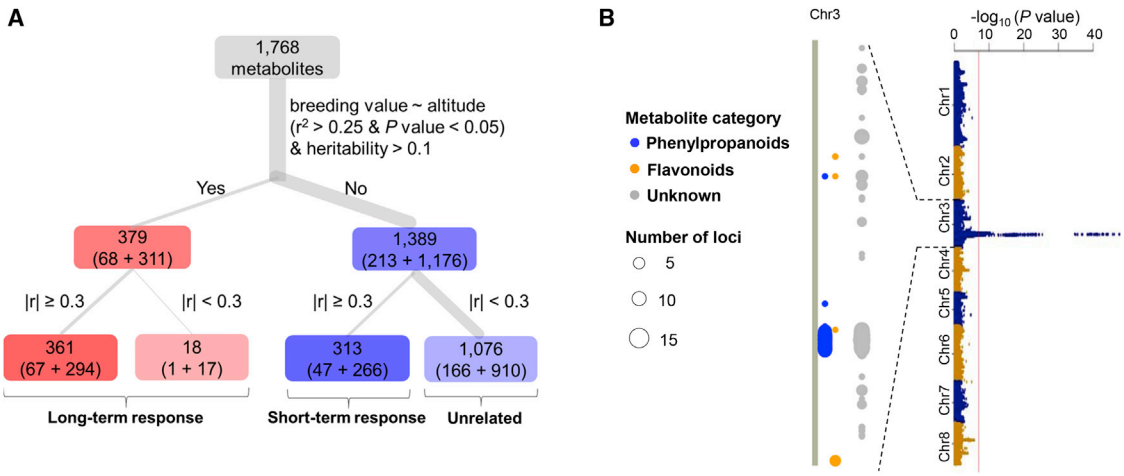


Figure 6. Genetic profiling of metabolites associated with high altitude

(A) Classification of 1,768 metabolites according to their responses to high altitudes. The width of the gray line represents the number of metabolites in each category. The two numbers in parentheses in each category are the numbers of annotated metabolites and unknown metabolites, respectively. The absolute value of the Spearman's rank correlation coefficient (|r|) was calculated for level of metabolite and altitude.

(B) Distribution of GWAS mapping loci for the 379 altitude-related metabolites on chromosome 3 (left) and a Manhattan plot of the mGWAS for chlorogenic acid (right). See also Figures S5 and S6 and Data S1AA and S1AC–S1AG.

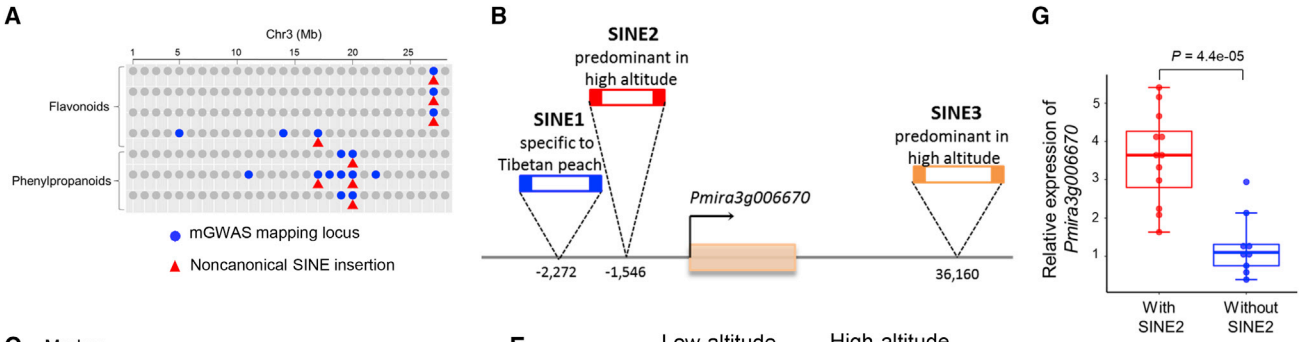
humidity, atmospheric pressure, and altitude. UV stress-related genes and metabolites were significantly differentiated in high-altitude relative to low-altitude populations, consistent with the dramatic increases in light intensity that occur as altitude increases. Genes associated with cold tolerance were not as differentiated as genes associated with responses to light. These data are probably explained by the fact that we studied plants growing from 2,067 to 4,492 m in the Himalayas. The average annual temperature at these altitudes ranges from −2°C to 8°C. Cultivated peach is a temperate perennial species that can tolerate these temperatures.[12,54] Our metabolite analysis provides compelling evidence that the high levels of phenylpropanoids and flavonoids that are associated with high altitudes may contribute to UV stress tolerance in Tibetan peach. Similar findings were reported previously for *Arabidopsis* and rice.[55,56] The loci controlling levels of quinic acid harbored natural selection signatures. Therefore, quinic acids are probably crucial metabolites for the adaptation of Tibetan peach to high-altitude conditions.

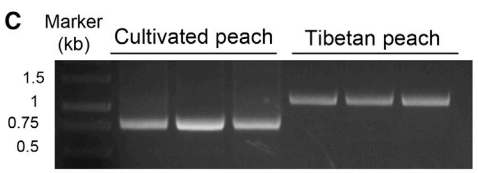
We found that SINE-type TEs expanded in Tibetan peach and that their expansion probably plays an important role in the adaptation to high-plateau environments from genomic and population perspectives. Our data indicate that SINE insertions into genes that promote the accumulation of phenylpropanoids may be one of the adaptive mechanisms used to cope with UV light stress. Compared with cultivated peach, several integrations of noncanonical SINEs into loci associated with species-specific metabolites were detected in Tibetan peach. Further divergence of a noncanonical SINE insertion in high-altitude relative to low-altitude Tibetan peach populations was also found. Thus, extensive insertion of SINEs in the genome of Tibetan peach may have driven its adaptation to stressful environments on the high plateau. Consistent with this interpretation, stress-induced activation of SINEs may play a prominent role in the genomic evolution of wheat.[57] We speculate that SINE retrotransposons were activated during the diversification of *Prunus* species and that SINE retrotransposons contributed to the adaptation of *Prunus* species to changing environments throughout history. We estimate that the native distribution of Tibetan peaches and other *Prunus* species at altitudes ranging from 2,000 to 4,500 m in the Himalayas occurred during the Himalayan uplift approximately 15–23 Mya.[50] This estimate is also consistent with the time frame of *Prunus* diversification based on molecular dating (Figure 1B). Taken together, the data provide evidence that the activation of retrotransposons contributes to the adaptation of plants to high plateau environments. Future characterization of genes and metabolites affected by these retrotransposons offers a promising approach both for increasing our understanding of the mechanisms that contribute to adaptation to high altitudes and for the genetic improvement of crops by breeding.

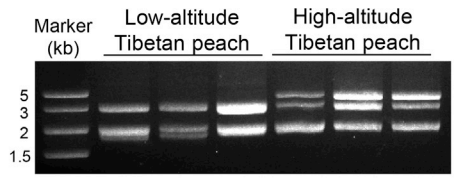
STAR★METHODS

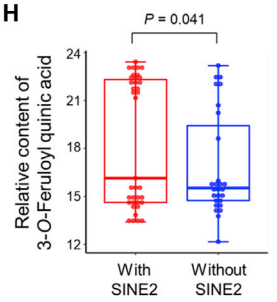
Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Materials collection and sequencing
  - *De novo* assembly of three *Prunus* genomes
  - Repeat element and protein-coding gene annotation for three *Prunus* genomes
  - Phylogenetic tree construction and estimation of the divergence times of *Prunus* species









D

	With SINE1	Without SINE1
Cultivated peach	0 (0%)	16 (100%)
Tibetan peach	15 (93.75%)	1 (6.25%)

F

	With SINE2	Without SINE2
Low-altitude Tibetan peach	9 (31.03%)	20 (68.97%)
High-altitude Tibetan peach	39 (76.47%)	12 (23.53%)

Figure 7. Example of SINE insertions associated with phenylpropanoids accumulation in the high-altitude Tibetan peach

(A) Co-localization of Tibetan peach-specific SINE insertions and mGWAS hits associated with differentially accumulated flavonoids and phenylpropanoids in Tibetan peach relative to cultivated peach populations on chromosome 3.

(B) Diagram of three SINE insertions around the candidate gene for controlling variation in 3-*O*-feruloyl quinic acid levels.

(C) Experimental validation of SINE1 insertion polymorphisms in cultivated and Tibetan peach.

(D) Genotype frequency of the SINE1 insertions in cultivated and Tibetan peach.

(E) Experimental validation of SINE2 insertion polymorphisms in high-altitude and low-altitude Tibetan peach.

(F) Genotype frequency of the SINE2 insertion in high-altitude and low-altitude Tibetan peach.

(G) Boxplot of relative expression of *Pmira3g006670* in Tibetan peaches with and without the SINE2 insertion. Relative expression was quantified using qRT-PCR. n = 3 replicates were analyzed from 12 Tibetan peaches with the SINE2 insertion and 9 Tibetan peaches without the SINE2 insertion.

(H) Box and beeswarm plots of 3-*O*-feruloyl quinic acid content in Tibetan peaches with and without the SINE2 insertion.

In (G) and (H), boxes indicate the median and interquartile range, and whiskers indicate maximum and minimum values. Statistically significant differences in (G) and (H) were determined using the Student's t test. p values are indicated. See also Figure S7 and Data S1AH–AL and S6.

- Comparative genomic analyses of *Prunus*
- Methylome analyses
- Phylogenetic analyses of *Prunus* populations
- Identification of deleterious mutations in Tibetan and cultivated peach accessions
- Detection of selection signatures for high-altitude adaptation
- Metabolomics profiling and analyses
- mGWAS analysis
- High-altitude adaptation-related metabolites in the Tibetan peach population
- Identification of Tibetan peach-specific SINEs
- Experimental validation for SINE insertions

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.06.062.

ACKNOWLEDGMENTS

We are grateful to W. Zhang, Z. Rao, B. Hu, and J. Fu for their help in performing experiments and to L. Wang and G. Hu for their suggestions during the analysis of our data. We thank H. Kuang, W. Xie, L. Guo and D. Duanmu for helpful discussions. We thank Wuhan MetWare Biotechnology Co., Ltd. (www.metware.cn) for the metabolic platform. The background geographic dataset in this study is provided by the National Tibetan Plateau Data Center (http://data.tpdc.ac.cn). This project was supported by the National Key Research and Development Program of China (2018YFD1000101), the National Natural Science Foundation of China (31925034, 31860536, and 31872052), the Tibet Finance Department of Agricultural Guidance (XZNKY-2019-C-055), the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502), China Postdoctoral Science Foundation (2019M660183), and China National Postdoctoral Program for Innovative Talents (BX201900134).

AUTHOR CONTRIBUTIONS

Q.X. and X.Z. conceived and designed the project and the strategy. X.Z. and S.Z. collected and evaluated the samples with contributions from G.P., H.Y., F.Z., Y.L., and J.L. X.W. performed the comparative population analysis and

metabolic GWAS. S.L. performed comparative genomics and transposon analyses. H.Z. performed phylogenetic analysis and population structure analysis. W.Z. analyzed metabolomics data and performed the PCR-based experimental validation. Y.H. assembled and annotated three genomes. Y.Z. was involved in the genetic analysis of metabolites. Q.X. coordinated the project with help from X.D., X.Z., Y.Z., and T.S.Y. X.W. and Q.X. wrote the manuscript with contributions from R.M.L., Y.Z., and T.S.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 2, 2020
Revised: February 25, 2021
Accepted: June 22, 2021
Published: July 26, 2021

REFERENCES

1. Vavilov, N.I. (1994). Origin and geography of cultivated plants (Cambridge University Press).
2. Luo, D., Yue, J.P., Sun, W.G., Xu, B., Li, Z.M., Comes, H.P., and Sun, H. (2016). Evolutionary history of the subnival flora of the Himalaya-Hengduan Mountains: first insights from comparative phylogeography of four perennial herbs. J. Biogeogr. *43*, 31–43.
3. Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature *512*, 194–197.
4. Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., Cao, C., Hu, Q., Kim, J., Larkin, D.M., et al. (2012). The yak genome and adaptation to life at high altitude. Nat. Genet. *44*, 946–949.
5. Qu, Y., Zhao, H., Han, N., Zhou, G., Song, G., Gao, B., Tian, S., Zhang, J., Zhang, R., Meng, X., et al. (2013). Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. Nat. Commun. *4*, 2071–2079.
6. Ge, R.L., Cai, Q., Shen, Y.Y., San, A., Ma, L., Zhang, Y., Yi, X., Chen, Y., Yang, L., Huang, Y., et al. (2013). Draft genome sequence of the Tibetan antelope. Nat. Commun. *4*, 1858–1864.
7. Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., Wang, T., Yeung, C.K.L., Chen, L., Ma, J., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat. Genet. *45*, 1431–1438.
8. Wang, M.S., Li, Y., Peng, M.S., Zhong, L., Wang, Z.J., Li, Q.Y., Tu, X.L., Dong, Y., Zhu, C.L., Wang, L., et al. (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol. Biol. Evol. *32*, 1880–1889.
9. Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., Yang, S., Jia, J., Kong, X., Wei, Z., et al. (2014). Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Res. *24*, 1308–1315.
10. Yang, X., Liu, H., Ma, Z., Zou, Y., Zou, M., Mao, Y., Li, X., Wang, H., Chen, T., Wang, W., and Yang, R. (2019). Chromosome-level genome assembly of *Triplophysa tibetana*, a fish adapted to the harsh high-altitude environment of the Tibetan Plateau. Mol. Ecol. Resour. *19*, 1027–1036.
11. Li, J.T., Gao, Y.D., Xie, L., Deng, C., Shi, P., Guan, M.L., Huang, S., Ren, J.L., Wu, D.D., Ding, L., et al. (2018). Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. Proc. Natl. Acad. Sci. USA *115*, 8406–8411.
12. Rana, S.K., Gross, K., and Price, T.D. (2019). Drivers of elevational richness peaks, evaluated for trees in the east Himalaya. Ecology *100*, e02548.
13. Manish, K., Pandit, M.K., Telwala, Y., Nautiyal, D.C., Koh, L.P., and Tiwari, S. (2017). Elevational plant species richness patterns and their drivers across non-endemics, endemics and growth forms in the Eastern Himalaya. J. Plant Res. *130*, 829–844.
14. Wu, Z.Y. (1983). Flora of Tibet (Science Press).
15. Yü, T.T. (1984). The origin and evolution of Rosaceae. J. Syst. Evol. *22*, 431–444.
16. Potter, D. (2011). Prunus. In Wild crop relatives: Genomic and breeding resources: Temperate fruits, C. Kole, ed. (Springer Berlin Heidelberg), pp. 129–145.
17. Faust, M., and Timon, B. (1995). Origin and dissemination of peach. Hortic. Rev. *17*, 331–379.
18. Yu, Y., Fu, J., Xu, Y., Zhang, J., Ren, F., Zhao, H., Tian, S., Guo, W., Tu, X., Zhao, J., et al. (2018). Genome re-sequencing reveals the evolutionary history of peach fruit edibility. Nat. Commun. *9*, 5404–5416.
19. Cao, K., Zheng, Z., Wang, L., Liu, X., Zhu, G., Fang, W., Cheng, S., Zeng, P., Chen, C., Wang, X., et al. (2014). Comparative population genomics reveals the domestication history of the peach, *Prunus persica*, and human influences on perennial fruit crops. Genome Biol. *15*, 415–429.
20. Bao, W., Wuyun, T., Li, T., Liu, H., Jiang, Z., Zhu, X., Du, H., and Bai, Y.E. (2017). Genetic diversity and population structure of *Prunus mira* (Koehne) from the Tibet plateau in China and recommended conservation strategies. PLoS ONE *12*, e0188685.
21. Cao, K., Peng, Z., Zhao, X., Li, Y., Liu, K., Arus, P., Zhu, G., Deng, S., Fang, W., Chen, C., et al. (2020). Pan-genome analyses of peach and its wild relatives provide insights into the genetics of disease resistance and species adaptation. bioRxiv. https://doi.org/10.1101/2020.07.13.200204.
22. Laitinen, R.A.E., and Nikoloski, Z. (2019). Genetic basis of plasticity in plants. J. Exp. Bot. *70*, 739–745.
23. Chapman, M.A., Hiscock, S.J., and Filatov, D.A. (2013). Genomic divergence during speciation driven by adaptation to altitude. Mol. Biol. Evol. *30*, 2553–2567.
24. Halbritter, A.H., Fior, S., Keller, I., Billeter, R., Edwards, P.J., Holderegger, R., Karrenberg, S., Pluess, A.R., Widmer, A., and Alexander, J.M. (2018). Trait differentiation and adaptation of plants along elevation gradients. J. Evol. Biol. *31*, 784–800.
25. Singh, A., and Roy, S. (2017). High altitude population of *Arabidopsis thaliana* is more plastic and adaptive under common garden than controlled condition. BMC Ecol. *17*, 39–54.
26. Guo, X., Hu, Q., Hao, G., Wang, X., Zhang, D., Ma, T., and Liu, J. (2018). The genomes of two *Eutrema* species provide insight into plant adaptation to high altitudes. DNA Res. *25*, 307–315.
27. Guo, W., Xin, M., Wang, Z., Yao, Y., Hu, Z., Song, W., Yu, K., Chen, Y., Wang, X., Guan, P., et al. (2020). Origin and adaptation to high altitude of Tibetan semi-wild wheat. Nat. Commun. *11*, 5085–5096.
28. Dong, T., Han, R., Yu, J., Zhu, M., Zhang, Y., Gong, Y., and Li, Z. (2019). Anthocyanins accumulation and molecular analysis of correlated genes by metabolome and transcriptome in green and purple asparaguses (*Asparagus officinalis,* L.). Food Chem. *271*, 18–28.
29. Zhang, J., Luo, W., Zhao, Y., Xu, Y., Song, S., and Chong, K. (2016). Comparative metabolomic analysis reveals a reactive oxygen species-dominated dynamic model underlying chilling environment adaptation and tolerance in rice. New Phytol. *211*, 1295–1310.
30. Zeng, X., Yuan, H., Dong, X., Peng, M., Jing, X., Xu, Q., Tang, T., Wang, Y., Zha, S., Gao, M., et al. (2020). Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of qingke. Mol. Plant *13*, 112–127.
31. Rahman, I.U., Afzal, A., Iqbal, Z., Hart, R., Abd Allah, E.F., Alqarawi, A.A., Alsubeie, M.S., Calixto, E.S., Ijaz, F., Ali, N., et al. (2020). Response of plant physiological attributes to altitudinal gradient: Plant adaptation to temperature variation in the Himalayan region. Sci. Total Environ. *706*, 135714.
32. Zhao, Y., Xu, F., Liu, J., Guan, F., Quan, H., and Meng, F. (2019). The adaptation strategies of *Herpetospermum pedunculosum* (Ser.) Baill at

altitude gradient of the Tibetan plateau by physiological and metabolomic methods. BMC Genomics *20*, 451–465.

33. Ma, L., Sun, X., Kong, X., Galvan, J.V., Li, X., Yang, S., Yang, Y., Yang, Y., and Hu, X. (2015). Physiological, biochemical and proteomics analysis reveals the adaptation strategies of the alpine plant *Potentilla saundersiana* at altitude gradient of the Northwestern Tibetan Plateau. J. Proteomics *112*, 63–82.
34. Li, Z.W., Hou, X.H., Chen, J.F., Xu, Y.C., Wu, Q., González, J., and Guo, Y.L. (2018). Transposable elements contribute to the adaptation of *Arabidopsis thaliana*. Genome Biol. Evol. *10*, 2140–2150.
35. Rey, O., Danchin, E., Mirouze, M., Loot, C., and Blanchet, S. (2016). Adaptation to global change: a transposable element-epigenetics perspective. Trends Ecol. Evol. *31*, 514–526.
36. Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., and Schulman, A.H. (2000). Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA *97*, 6603–6607.
37. Daccord, N., Celton, J.M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., et al. (2017). High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. Nat. Genet. *49*, 1099–1106.
38. Zhang, T., Qiao, Q., Novikova, P.Y., Wang, Q., Yue, J., Guan, Y., Ming, S., Liu, T., De, J., Liu, Y., et al. (2019). Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. Proc. Natl. Acad. Sci. USA *116*, 7137–7146.
39. Verde, I., Jenkins, J., Dondini, L., Micali, S., Pagliarani, G., Vendramin, E., Paris, R., Aramini, V., Gazza, L., Rossini, L., et al. (2017). The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. BMC Genomics *18*, 225–242.
40. Alioto, T., Alexiou, K.G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., Dhingra, A., Duval, H., Fernández I Martí, Á., Frias, L., et al. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. Plant J. *101*, 455–472.
41. Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., Tao, Y., Wang, J., Yuan, Z., Fan, G., et al. (2012). The genome of *Prunus mume*. Nat. Commun. *3*, 1318–1325.
42. 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*, 128–139.
43. Wang, J., Liu, W., Zhu, D., Hong, P., Zhang, S., Xiao, S., Tan, Y., Chen, X., Xu, L., Zong, X., et al. (2020). Chromosome-scale genome assembly of sweet cherry (*Prunus avium* L.) cv. Tieton obtained using long-read and Hi-C sequencing. Hortic. Res. *7*, 122–132.
44. Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., et al. (2011). The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. *43*, 109–116.
45. VanBuren, R., Wai, C.M., Colle, M., Wang, J., Sullivan, S., Bushakra, J.M., Liachko, I., Vining, K.J., Dossett, M., Finn, C.E., et al. (2018). A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. Gigascience *7*, 1–9.
46. Jaillon, O., Aury, J.M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., et al.; French-Italian Public Consortium for Grapevine Genome Characterization (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature *449*, 463–467.
47. Su, T., Wilf, P., Huang, Y., Zhang, S., and Zhou, Z. (2015). Peaches preceded humans: fossil evidence from SW China. Sci. Rep. *5*, 16794.
48. Chandler, M. (1963). The lower tertiary floras of southern England III (London: British Museum).
49. Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. *34*, 1812–1819.
50. Spicer, R.A., Su, T., Valdes, P.J., Farnsworth, A., Wu, F.X., Shi, G., Spicer, T.E.V., and Zhou, Z. (2021). Why 'the uplift of the Tibetan Plateau' is a myth. Natl. Sci. Rev. *8*, https://doi.org/10.1093/nsr/nwaa091.
51. Li, J., Yang, L., Jin, D., Nezames, C.D., Terzaghi, W., and Deng, X.W. (2013). UV-B-induced photomorphogenesis in *Arabidopsis*. Protein Cell *4*, 485–492.
52. Müller-Xing, R., Xing, Q., and Goodrich, J. (2014). Footprints of the sun: memory of UV and light stress in plants. Front. Plant Sci. *5*, 474–485.
53. Blanc, J., and Berg, J.J. (2020). How well can we separate genetics from the environment? eLife *9*, e64948.
54. Li, Y., Cao, K., Zhu, G., Fang, W., Chen, C., Wang, X., Zhao, P., Guo, J., Ding, T., Guan, L., et al. (2019). Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. Genome Biol. *20*, 36–53.
55. Kusano, M., Tohge, T., Fukushima, A., Kobayashi, M., Hayashi, N., Otsuki, H., Kondou, Y., Goto, H., Kawashima, M., Matsuda, F., et al. (2011). Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of *Arabidopsis* to UV-B light. Plant J. *67*, 354–369.
56. Park, H.L., Lee, S.W., Jung, K.H., Hahn, T.R., and Cho, M.H. (2013). Transcriptomic analysis of UV-treated rice leaves reveals UV-induced phytoalexin biosynthetic pathways and their regulatory networks in rice. Phytochemistry *96*, 57–71.
57. Ben-David, S., Yaakov, B., and Kashkush, K. (2013). Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. Plant J. *76*, 201–210.
58. Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods *13*, 1050–1054.
59. Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569.
60. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.
61. Adey, A., Kitzman, J.O., Burton, J.N., Daza, R., Kumar, A., Christiansen, L., Ronaghi, M., Amini, S., Gunderson, K.L., Steemers, F.J., and Shendure, J. (2014). In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. Genome Res. *24*, 2041–2049.
62. Kaplan, N., and Dekker, J. (2013). High-throughput genome scaffolding from *in vivo* DNA interaction frequency. Nat. Biotechnol. *31*, 1143–1147.
63. Chen, Y., Nie, F., Xie, S.Q., Zheng, Y.F., Bray, T., Dai, Q., Wang, Y.X., Xing, J.f., Huang, Z.J., Wang, D.P., et al. (2020). Fast and accurate assembly of Nanopore reads via progressive error correction and adaptive read selection. bioRxiv. https://doi.org/10.1101/2020.02.01.930107.
64. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.
65. Liu, H., Wu, S., Li, A., and Ruan, J. (2020). SMARTdenovo: a de novo assembler using long noisy reads. Preprints, 10.20944/preprints202009.0207.v1.
66. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746.

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

68. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95.

69. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98.

70. Simão, 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.

71. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv.

72. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

73. Han, Y., and Wessler, S.R. (2010). MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. *38*, e199.

74. Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics *9*, 18–31.

75. Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. *35*, W265-8.

76. Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. *176*, 1410–1422.

77. Wenke, T., Döbel, T., Sörensen, T.R., Junghans, H., Weisshaar, B., and Schmidt, T. (2011). Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. Plant Cell *23*, 3117–3128.

78. Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir, A., and Mesirov, J.P. (2017). Variant review with the integrative genomics viewer. Cancer Res. *77*, e31–e34.

79. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457.

80. Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics *Chapter 4*.

81. Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. *32*, W309-12.

82. Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source *ab initio* eukaryotic gene-finders. Bioinformatics *20*, 2878–2879.

83. Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics *6*, 31–41.

84. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

85. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. *31*, 5654–5666.

86. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.

87. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157–170.

88. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

89. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

90. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

91. Harris, R.S. (2007). Improved pairwise alignment of genomic DNA. PhD thesis (The Pennsylvania State University).

92. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12.

93. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

94. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.

95. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

96. Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. *45* (W1), W122–W129.

97. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res. *46* (W1), W200–W204.

98. Marchler-Bauer, A., and Bryant, S.H. (2004). CD-Search: protein domain annotations on the fly. Nucleic Acids Res. *32*, W327-31.

99. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

100. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. *13*, R87.

101. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

102. 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.

103. Felsenstein, J. (1989). PHYLIP-phylogeny inference package (version 3.2). Cladistics *5*, 164–166.

104. Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. *14*, 708–715.

105. Hubisz, M.J., Pollard, K.S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. Brief. Bioinform. *12*, 41–51.

106. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila* melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.

107. Choi, Y., and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics *31*, 2745–2747.

108. 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., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

109. Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. *78*, 629–644.

110. Gautier, M., Klassmann, A., and Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol. Ecol. Resour. *17*, 78–90.

111. Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M., and Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics *193*, 929–941.

112. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. J. Stat. Softw. *25*, 1–18.

113. Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome *4*, https://doi.org/10.3835/plantgenome2011.08.0024.

114. Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. Nat. Methods *8*, 833–835.

115. Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. *8*, 4321–4325.

116. Cao, X., Duan, W., Wei, C., Chen, K., Grierson, D., and Zhang, B. (2019). Genome-wide identification and functional analysis of carboxylesterase and methylesterase gene families in peach (*Prunus persica* L. Batsch). Front. Plant Sci. *10*, 1511–1523.

117. Schmidt, M.H., Vogel, A., Denton, A.K., Istace, B., Wormit, A., van de Geest, H., Bolger, M.E., Alseekh, S., Maß, J., Pfaff, C., et al. (2017). De novo assembly of a new *Solanum pennellii* accession using nanopore sequencing. Plant Cell *29*, 2336–2348.

118. Yoshioka, Y., Matsumoto, S., Kojima, S., Ohshima, K., Okada, N., and Machida, Y. (1993). Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc. Natl. Acad. Sci. USA *90*, 6562–6566.

119. Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C.M., Zhang, C., Tian, Y., Liu, G., Gul, H., et al. (2019). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat. Commun. *10*, 1494–1506.

120. Liu, M.J., Zhao, J., Cai, Q.L., Liu, G.C., Wang, J.R., Zhao, Z.H., Liu, P., Dai, L., Yan, G., Wang, W.J., et al. (2014). The complex jujube genome provides insights into fruit tree biology. Nat. Commun. *5*, 5315–5326.

121. van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., and Page, J.E. (2011). The draft genome and transcriptome of *Cannabis sativa*. Genome Biol. *12*, R102.

122. Osipowski, P., Pawełkowicz, M., Wojcieszek, M., Skarzyńska, A., Przybecki, Z., and Pląder, W. (2020). A high-quality cucumber genome assembly enhances computational comparative genomics. Mol. Genet. Genomics *295*, 177–193.

123. Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., Xing, S., Du, J., Ma, S., and Tian, Z. (2018). *De novo* assembly of a Chinese soybean genome. Sci. China Life Sci. *61*, 871–884.

124. Vlasova, A., Capella-Gutiérrez, S., Rendón-Anaya, M., Hernández-Oñate, M., Minoche, A.E., Erb, I., Câmara, F., Prieto-Barja, P., Corvelo, A., Sanseverino, W., et al. (2016). Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. Genome Biol. *17*, 32–49.

125. Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K.L., Yandell, M., Gundlach, H., et al. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. BMC Genomics *15*, 312–325.

126. De Vega, J.J., Ayling, S., Hegarty, M., Kudrna, D., Goicoechea, J.L., Ergon, Å., Rognli, O.A., Jones, C., Swain, M., Geurts, R., et al. (2015). Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. Sci. Rep. *5*, 17394–17403.

127. R Core Team (2020). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Illumina reads of 346 peach accessions, 22 cherry accessions, seven apricot accessions and one plum accession	This paper	NCBI SRA:PRJNA655666; National Genomics Data Center: PRJCA004083
Illumina reads of transcriptomes from 11 *P. mira* accession	This paper	NCBI SRA: PRJNA662872; National Genomics Data Center: PRJCA004161
Illumina reads of DNA methylation sequencing data of two Tibetan peach accession and one cultivated peach accession	This paper	NCBI SRA: PRJNA662872; National Genomics Data Center: PRJCA004161
Illumina data, 10 × data, HiC data, PacBio data, Nanopore data for genome assembly of *P. mira*, *P. mume* 'Tibet', and *P. armeniaca* 'Tibet'	This paper	NCBI SRA: PRJNA668277 and PRJNA673568; National Genomics Data Center: PRJCA004519
Illumina reads of transcriptomes for genome annotation of *P. mira*, *P. mume* 'Tibet', and *P. armeniaca* 'Tibet'	This paper	NCBI SRA: PRJNA668277; National Genomics Data Center: PRJCA004519
Genome assembly data for *P. mira* (v.2), *P. mira* (v.1), *P. mume* 'Tibet', and *P. armeniaca* 'Tibet'	This paper	NCBI GenBank: JADEVJ000000000, JADDOM000000000, JADEVK000000000, and JADEVL000000000; National Genomics Data Center: PRJCA004519
SNP dataset of peach accessions	This paper	Figshare: https://doi.org/10.6084/m9.figshare.14447502.v1
Statistics of the SNP dataset of peach accessions	This paper	Figshare: https://doi.org/10.6084/m9.figshare.14479164.v1
Original metabolomics data	This paper	Figshare: https://doi.org/10.6084/m9.figshare.14707605.v2
Oligonucleotides		
Forward primer for PCR amplification of SINE1 insertion: AGTTCTTCCATCTTGCCTCATT	This paper	N/A
Reverse primer for PCR amplification of SINE1 insertion: GATCCACCAGTTCCAACACAA	This paper	N/A
Forward primer for PCR amplification of SINE2 insertion: CATTCTTCATCACTCAACTACTGAC	This paper	N/A
Reverse primer for PCR amplification of SINE2 insertion: CTTCTCCGTTACAACTCCGATT	This paper	N/A
Forward primer for PCR amplification of SINE3 insertion: GGCTATGGCAAGGCAAGATAA	This paper	N/A
Reverse primer for PCR amplification of SINE3 insertion: AACCGTTGATTCGCACAATTAG	This paper	N/A
Software and Algorithms		
Code used in the analysis of methylome	This paper	Figshare: https://doi.org/10.6084/m9.figshare.14706162.v1
FALCON pipeline v1.0	[58]	https://github.com/PacificBiosciences/pb-assembly
Quiver v5.1.0	[59]	https://github.com/PacificBiosciences/GenomicConsensus
Pilon v1.22	[60]	https://github.com/broadinstitute/pilon
fragScaff v180112	[61]	https://github.com/adeylab/fragScaff
Lachesis v201701	[62]	https://github.com/shendurelab/LACHESIS

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Nextdenovo v2.0-beta.1	NextOmics; Jiang Hu	https://github.com/Nextomics/NextDenovo
Necat (default)	63	https://github.com/xiaochuanle/NECAT
Canu v1.9	64	https://github.com/marbl/canu/
SMARTdenovo (default)	65	https://github.com/ruanjue/smartdenovo
Racon v1.3.1	66	https://github.com/isovic/racon
NextPolish v1.0.4	67	https://github.com/Nextomics/NextPolish
3D-DNA pipeline v180922	68	https://github.com/aidenlab/3d-dna
juicer v1.11.08	69	https://github.com/aidenlab/Juicebox
BUSCO v.3.0.2	70	https://busco.ezlab.org/
BWA v0.7.17	71	http://bio-bwa.sourceforge.net/bwa.shtml
SAMtools v1.9	72	http://www.htslib.org/
MITE-Hunter (default)	73	https://mite-hunter.com/
LTRharvest v1.6.1	74	https://www.zbh.uni-hamburg.de/en/forschung/gi/software/ltrharvest.html
LTR_FINDER v1.07	75	http://tlife.fudan.edu.cn/tlife/ltr_finder/
LTR_retriever (default)	76	https://github.com/oushujun/LTR_retriever
SINE-Finder v1.0.1	77	https://figshare.com/articles/dataset/SINE-Finder_py/15023385
IGV v2.6.2	78	http://software.broadinstitute.org/software/igv/
RepeatModeler2 v1.0.11	79	http://www.repeatmasker.org/RepeatModeler/
RepeatMasker v4.0.9	80	http://www.repeatmasker.org/
AUGUSTUS v3.3	81	http://bioinf.uni-greifswald.de/augustus/
GlimmerHMM v3.0.4	82	http://ccb.jhu.edu/software/glimmerhmm/
Exonerate v2.4.0	83	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
Trinity v2.8.5	84	https://github.com/trinityrnaseq/trinityrnaseq/wiki
PASA v2.4.1	85	https://pasa.rti.org/
EVM v1.1.1	86	https://www.msi.umn.edu/sw/evidencemodeler
OrthoFinder v2.2.7	87	https://github.com/davidemms/OrthoFinder
MUSCLE v.3.8.31	88	https://www.drive5.com/muscle/downloads.htm
RAxML v7.7.8	89	https://github.com/stamatak/standard-RAxML
PAML v4.8	90	http://abacus.gene.ucl.ac.uk/software/paml.html
LASTZ v1.02.00	91	http://www.bx.psu.edu/~rsharris/lastz/
MUMmer v4.0.0	92	https://mummer4.github.io/
CD-HIT v4.8.1	93	http://weizhongli-lab.org/cd-hit/
BLAST v2.5.0	94	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Circos v0.69	95	http://circos.ca/
agriGO v2.0	96	http://systemsbiology.cau.edu.cn/agriGOv2/index.php
HMMER v3.1b2-linux-intel-x86_64	97	http://www.hmmer.org/
CD Search	98	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
Bismark v0.22.3	99	https://www.bioinformatics.babraham.ac.uk/projects/bismark/
methylKit v1.12.0	100	https://github.com/al2na/methylKit
GATK v4.1.1	101	https://github.com/broadinstitute/gatk

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GCTA v1.26.0	[102]	https://cnsgenomics.com/software/gcta/#Overview
PHYLIP v3.696	[103]	https://evolution.genetics.washington.edu/phylip.html
MULTIZ v.012109	[104]	https://anaconda.org/bioconda/multiz
PHAST v1.4	[105]	http://compgen.cshl.edu/phast/
SnpEff v4.3	[106]	http://pcingola.github.io/SnpEff/
PROVEAN v1.1.5	[107]	http://provean.jcvi.org/index.php
VCFtools v0.1.13	[108]	https://vcftools.github.io/index.html
fastPHASE v1.4.0	[109]	http://stephenslab.uchicago.edu/software.html#fastphase
REHH v3.0.1	[110]	https://cran.r-project.org/web/packages/rehh/index.html
hapFLK v1.4	[111]	https://pypi.org/project/hapflk/
FactoMineR	[112]	http://factominer.free.fr/
factoextra	R package; Alboukadel Kassambara, Fabian Mundt	http://www.sthda.com/english/rpkgs/factoextra
rrBLUP v4.6.1	[113]	https://cran.r-project.org/web/packages/rrBLUP/index.html
FaST-LMM.207c.Linux	[114]	https://www.microsoft.com/en-us/research/project/fastlmm/?from=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fum%2Fredmond%2Fprojects%2Fmscompbio%2Ffastlmm%2F
lmPerm v2.1.0	R package; Bob Wheeler, Marco Torchiano	https://github.com/mtorchiano/lmPerm

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Qiang Xu (xuqiang@mail.hzau.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Genome assembly data for *P. mira* (v.2), *P. mira* (v.1), *P. mume* 'Tibet', and *P. armeniaca* 'Tibet', RNA-seq data for genome annotation, whole-genome sequencing data for the *Prunus* accessions, and RNA-seq and WGBS data for Tibetan peach and cultivated peach have been deposited at NCBI (https://www.ncbi.nlm.nih.gov/) and the National Genomics Data Center (China National Center for Bioinformation, https://bigd.big.ac.cn/), and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Original metabolomics data for Tibetan peach and cultivated peach have been deposited in figshare (https://figshare.com/) and are publicly available as of the date of publication. The DOI is listed in the key resources table. All original code used in the analysis of methylome has been deposited at figshare and is publicly available as of the date of publication. The DOI is listed in the key resources table. Any additional information required to reanalyze the data reported in this work paper is available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

No experimental models typical in the life sciences were used in this study.

METHOD DETAILS

Materials collection and sequencing

The 321 samples (289 *P. mira*, one *P. davidiana*, 22 *P. avium*, seven *P. armeniaca* 'Tibet', one *P. mume* 'Tibet', and one *P. salicina* 'Tibet') used in this study were collected in the wild from 2017 to 2019 in the Lhasa, Nyingchi, Xigaz and Shannan regions of Tibet. Two *P. ferganensis* accessions were collected from Xinjiang Province, China. In addition, 44 *P. persica* accessions and ten *P. mira* accessions were collected at the Qinghai-Tibet Plateau Fruit Tree Scientific Observation Test Station in Lhasa. Among the 346 peach accessions, *P. persica* and *P. ferganensis* accessions belong to landraces, and the others are wild germplasm. In our field germplasm investigation, we sampled individuals that provided diversity based on the phenotypes of flowers, fruits and leaves. Individuals with similar phenotypes to the individuals that were already sampled were excluded from our collection. The localities of the harvest sites, identities, and characters of the *P. mira* accessions were provided in details (Data S1A), including five, four, and four indices for fruit (weight, length, width, shape, and soluble solid content), nutlets (length, width, thickness, and weight), and tree resistance (shriveled leaves, gummosis, perforations, and scabs), respectively.

Total genomic DNA for each of the accessions was extracted from leaves using the CTAB method.[115] 0.3g of leaves ground with liquid nitrogen were transfered to a centrifuge tube with CTAB lytic solution and incubated in a water bath at 65°C for 2 h. After that, equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added, mixed, and centrifuged for 10 min at 12,000 rpm at room temperature. Then supernatant was transfered to another centrifuge tube, and added an equal volume of chloroform-isoamyl alcohol (24:1). After mixing, it was centrifuged for 10 min at 12,000 rpm at room temperature. The supernatant was transfered to a centrifuge tube, mixed with 3/4 volume of isopropanol, and incubated at −20°C for preciptation. After centrifugation at 12,000 rpm for 10 min, the pellet was washed twice with 75% ethanol. Finally, the pellet was dissolved with RNase and ddH_2O at 37°C for 15 min. The DNA-seq was performed on the Illumina NovaSeq 6000 platform with an average depth of 30-fold genome coverage (Data S1A). In addition, the DNA-seq data from seven wild peach accessions (two *P. kansuensis*, three *P. davidiana*, one *P. tangutica* and one *P. mongolica*), 15 *P. dulcis* and five *P. ledebouriana* from a previously published study[18] were collected for population analysis.

The fruit from *P. mira* used in the transcriptome analysis were collected from 11 different altitudes (2,676 m, 2,855 m, 2,987 m, 2,989 m, 3,144 m, 3,584 m, 3,759 m, 3,806 m, 3,866 m, 3,937 m, and 4,396 m) in Tibet (Data S1AM). Three biological replicates were used for transcriptomic analysis. Total RNA was extracted using the RNAprep Pure Plant Kit (DP441, TIANGEN Biotech). RNA-seq was conducted on the Illumina NovaSeq 6000 platform, and 150-bp paired-end reads were generated (Data S1AM). The transcriptome data from the fruit produced by *P. persica* from a previously published study[116] were also used for transcriptome analysis.

The leaves from the two *P. mira* species used in the methylome analysis were collected from the wild in Tibet. The leaves from the single *P. persica* species used in the methylome analysis were collected at Huazhong Agricultural University, Wuhan, China. Total genomic DNA was extracted from leaves using a modified CTAB DNA extraction protocol. BS-seq library construction and sequencing utilized two biological replicates per sample. Libraries were sequenced on the Illumina NovaSeq 6000 platform (Data S1AN).

De novo assembly of three *Prunus* genomes

For *P. mira*, 54.62 Gb of PacBio data (216.08 × genome coverage), 25.36 Gb of HiC data (101.44 × genome coverage), 78.05 Gb of 10 × data (312.20 × genome coverage), and 13.50 Gb of Illumina data (54.00 × genome coverage) were used (Data S1D). We first used the FALCON pipeline[58] to assemble the PacBio reads into contigs. The primary contigs were then polished using the Quiver software[59] by aligning PacBio reads. The Pilon software[60] was used to perform the error correction with Illumina reads. The corrected contigs were then linked to scaffolds using the fragScaff software[61] with 10 × Genomics reads. Finally, the Lachesis software[62] was used to cluster, order and orient the scaffolds with the Hi-C library. This assembled result was defined as version 1 for *P. mira* and was used in the subsequent analyses (Data S1C).

For *P. mume* 'Tibet' and *P. armeniaca* 'Tibet', 33.05 and 22.67 Gb of Nanopore data (113.18 and 83.65 × genome coverage, respectively), 36.02 and 30.98 Gb of HiC data (123.36 and 114.32 × genome coverage, respectively), and 12.16 and 28.84 Gb of Illumina data (41.64 and 106.42 × genome coverage, respectively) were used (Data S1D). In addition, 15.32 Gb of ultralong reads (61.28 × genome coverage) were generated for *P. mira* to further improve the continuity of the genome (Data S1D).

Raw Nanopore reads were filtered for quality at q7 for all of the species. The filtered reads were used for *de novo* assembly. To obtain the best assembly results, several assembly methods that have been described previously were used.[117] For *P. mira*, the longest 45 × reads were first chosen and assembled using Nextdenovo (https://github.com/Nextomics/NextDenovo). This assembly result was defined as version 2 for *P. mira* (Data S1C). For *P. mume* 'Tibet', the best assembly results were obtained using Necat.[63] For *P. armeniaca* 'Tibet', the raw reads were corrected using Canu.[64] The corrected reads were assembled using SMARTdenovo.[65]

The error correction of contigs was performed using Racon[66] and was iterated three times based on Nanopore reads, followed by two rounds of polishing using NextPolish[67] with Illumina short reads. With the Hi-C library, the error-corrected contigs were anchored to eight superscaffolds using the 3D-DNA pipeline[68] and juicer.[69] The completeness of the assembled genomes was evaluated using the BUSCO software.[70] After variation detection with the Illumina sequencing data using the BWA[71] and SAMtools[72] software, the homologous SNP rate was used as an estimate of the error rate of the genome assembly.

Repeat element and protein-coding gene annotation for three *Prunus* genomes

Miniature inverted-repeat transposable elements (MITEs) were detected by scanning the genome using MITE-Hunter[73] with the parameters '-P 1 -S 12345678'. Long terminal repeat (LTR) libraries were constructed using LTRharvest[74] and LTR_FINDER[75] and integrated using LTR_retriever.[76] The LTR insertion time was calculated using a substitution rate of 7.7E-9 nucleotides per site per year and a generation time of 10 years. The classical SINE library was downloaded from the SINEbase database. The SINE elements were identified using SINE-Finder.[77] Canonical SINEs are distinguished by several structural characteristics.[77,118] The structure of canonical SINEs include a 5′ targeted site duplication (TSD) region containing 40 nucleotides followed by an A motif (RVTGG), a spacer of 25-50 nucleotides, and a box B motif (GTTCRA). The B box motif is followed by a spacer of 20-500 nucleotides, six adenines and a 3′ TSD region of 40 nucleotides. The SINE elements in three *Prunus* genomes were clustered as either canonical SINE elements with classical conserved module constitutions or noncanonical SINE elements with novel sequences. The noncanonical SINE elements were validated by manually checking the presence of reads mapped to the genome using IGV.[78] The remaining genome sequences were used to build a *de novo* TE library using the RepeatModeler software.[79] The TE library was used to identify repeat sequences in particular genomes using RepeatMasker[80] (Data S1J).

Gene models were annotated based on *ab initio* gene predictions, homology searches, and RNA-seq. For *ab initio* gene predictions, AUGUSTUS[81] and GlimmerHMM[82] were employed uisng default parameters. The protein databases were constructed by integrating the amino acid sequences from the Rosaceae databases. Homology searching was then conducted using the exonerate alignment tool.[83] In addition, RNA-seq reads were generated from a mixture of tissues (Data S1AM). The trinity software[84] was utilized to perform genome-guided and *de novo* transcript assembly. The PASA software[85] was used to update the protein-coding gene annotations by incorporating PASA alignment evidence, correcting exon boundaries, adding UTRs, and modeling alternative splicing based on the PASA alignment assemblies. All of the gene structures predicted using the aforementioned methods were combined using the EVM software.[86]

Phylogenetic tree construction and estimation of the divergence times of *Prunus* species

Genomic data from 12 representative Rosaceace species were collected. These species included *P. mira*, *P. persica*,[39] *P. ferganensis*,[21] *P. dulcis*,[40] *P. mume* 'Tibetan', *P. mume*,[41] *P. salicina* (https://www.rosaceae.org/Analysis/9019655), *P. armeniaca* 'Tibetan', *P. armeniaca*,[42] *P. avium*,[43] *Fragaria vesca*,[44] *Rubus occidentalis*.[45] *Vitis vinifera*[46] was used as an outgroup. A total of 2,589 single-copy genes shared by these 13 species were identified using the OrthoFinder software.[87] For each of these genes, the coding region sequences were separately aligned using the MUSCLE software.[88] The aligned amino acid sequences were converted into the corresponding CDS sequences and then concatenated (Data S4). The sequence information for *P. salicina* 'Tibetan' was retrieved using resequencing data from *P. salicina* 'Tibetan'. The maximum likelihood (ML) tree was produced with grape as the outgroup using the substitution model GTR+G+I in the RAxML software.[89] A total of 500 rapid bootstrap inferences were performed.

To estimate the evolutionary timescale, we employed the MCMCTree program from the PAML software[90] based on three fossil-based age constraints, including one fossil of *P. persica* found in Kunming[47] (approximately 2.6 Mya), one fossil of *Rubus*[48] (approximately 41.3 Mya), and the divergence between grape and Rosaceae[49] (approximately 131-107 Mya). Divergence time dating was performed using the *Prunus* phylogenetic tree with the following parameters: burn-in of 5,000,000 iterations, sample frequency of 5,000, and the MCMC process performed 20,000 times.

Comparative genomic analyses of *Prunus*

Ten representative *Prunus* species were selected for comparative genomic analyses, including *P. mira*, *P. persica*, *P. ferganensis*, *P. dulcis*, *P. mume* 'Tibetan', *P. mume*, *P. salicina*, *P. armeniaca* 'Tibetan', *P. armeniaca*, *P. avium*. Pairwise comparisons (*P.mira* versus each of the other species for the construction of the *Prunus* pan-genome; *P. mira* versus *P. persica*, *P. mume* 'Tibetan' versus *P. mume*, and *P. armeniaca* 'Tibetan' versus *P. armeniaca* for comparative genomic analysis between Tibetan and cultivated *Prunus* species) of genomic sequences were performed using LASTZ[91] with the parameters '–notransition–step = 20–ambiguous = iupac–nogapped–format = rdotplot' and the nucmer program in the MUMmer software[92] with the parameters '–maxgap = 500–mincluster = 100 -q -r'. The alignment results were further filtered to retain one-to-one alignment regions using the delta-filter program in the Mummer package. Information on genomic differences was obtained using the show-diff program in the Mummer package. We identified genes in Tibetan peach-specific sequences by identifying collinear blocks between Tibetan peach and cultivated peach genomes. To avoid interference from incomplete genomic annotation, we aligned these Tibetan peach-specific genes with the *P. persica* genome using BLAST[93] and identity threshold of 95% and coverage threshold of 100%. The unaligned sequence ($\geq$500 bp) was extracted. Redundancy was removed using CD-HIT[94] and BLAST. The nonredundant novel sequences were added to the genomic sequences of the *Prunus* pan-genome. The annotated genes in these sequences were added to the total genes of the *Prunus* pan-genome.

SNPs, indels, and PAVs that were identified with high confidence were further filtered and validated with resequencing data. The PAVs identified were verified using the PacBio data from *P. mira* and the Illumina data from *P. mira* and *P. persica*. To avoid interference from repetitive sequences, only the uniquely mapped reads were retained for the verification of PAVs. The criteria for the presence of a genomic region were that the coverage was greater than 90% and that the average depth of coverage of the sequencing

data was greater than 5 × . In contrast, genomic regions with greater than 90% of the sequences not covered by sequencing reads were considered to be absent. Then, PAVs containing less than 50 bp were filtered out. Circos[95] was employed to display the features between genomes.

Genes were clustered using OrthoFinder with default parameters. A gene family was considered to be a core gene family if it was shared by all species. A gene family was considered to be variable if the members of the gene family were absent from at least one of the species. To identify core gene families in *Prunus*, we first clustered the amino acid sequences encoded by a total 312,605 genes in the ten *Prunus* species (25,678 genes in *P. mira*; 26,873 genes in *P. persica*; 54,862 genes in *P. ferganensis*; 27,042 genes in *P. dulcis* var. Texas; 27,481 genes in *P. salicina* var. Zhongli No.6; 31,116 genes in *P. mume* 'Tibet'; 29,705 genes in *P. mume*; 28,973 genes in *P. armeniaca* 'Tibet'; 30,436 genes in *P. armeniaca*; and 30,439 genes in *P. avium*) and obtained 28,404 clusters of homologous genes. The 12,239 clusters contained at least one gene from all of these *Prunus* genomes was defined as the core gene families.

For each of the single-copy orthologous genes shared by these *Prunus* genomes, multiple amino acid sequence alignments were performed using MUSCLE. The alignment was then converted to the corresponding CDS alignment. The selection pressures experienced by *Prunus* were estimated using the codeml program from the PAML software. The free-ratio branch model (model = 2, NSsites = 0) was used to estimate different dN/dS ratios for each branch. The one-ratio branch model (model = 0, NSsites = 0) was used to estimate the identical dN/dS ratio for all of the branches. Based on the likelihood ratio test (LRT), positively selected genes were identified according to the chi-square test ($p < 0.05$, Data S1Z). The enrichment analysis for the positively selected genes was performed using agriGO[96] against the background of the corresponding whole genome.

The gene family members in each genome were identified using the HMMER software[97] on the basis of the domain profiles of 58 transcription factor families collected in the PlantTFDB database. The domains of the proteins encoded by the members of *FAR1/FHY3* gene family were manually checked using the CD-Search.[98]

Methylome analyses

The high-quality reads from the WGBS of the Tibetan peach and cultivated peach were separately mapped to the Tibetan peach genome using the bismark program from the Bismark software.[99] Only the uniquely mapped reads were retained. The duplications caused by PCR amplifications were removed using the deduplicate_bismark program from the Bismark software. The DNA methylation ratios for each cytosine were calculated and extracted using the bismark_methylation_extractor program from the Bismark software and in-house scripts.

Differentially methylated regions (DMRs) between Tibetan peach and cultivated peach were identified using the methylKit package[100] in R which used a sliding-window approach with 10,000-bp windows sliding in 5,000-bp steps. For differential methylation, the criteria for significance were that the difference in methylation levels between Tibetan peach and cultivated peach was greater than 0.25 and that the *P value* from the significance test (Fisher's exact test) and the FDR from multiple-testing were both less than 0.01. Then, the adjacent differentially methylated windows were merged to get the final differentially methylated regions.

To measure the DNA methylation pattern over the 10-kb upstream regions, 10-kb downstream regions, and the TE regions, the average DNA methylation levels in CG, CHG, and CHH contexts in 100-bp bins in the flanking regions and in 10 equally long bins in the TE regions were calculated.

Phylogenetic analyses of *Prunus* populations

A total of 304 Tibetan peach accessions (299 from this study), 56 cultivated peaches (46 from this study), two *P. kansuensis*, four *P. davidiana* (one from this study), one *P. tangutica*, one *P. mongolica*, 15 *P. dulcis* and five *P. ledebouriana* were used for population analyses (Data S1A). For each of these accessions, the high-quality paired-end reads were mapped to the *P. mira* genome using the BWA software. Reduplication was performed using the MarkDuplicates command from the GATK software.[101] SNP calling was conducted using the HaplotypeCaller command from the GATK software. The genotype files for each accession were combined and transformed using the CombineGVCFs package and the GenotypeGVCFs command from the GATK software, respectively. The SNP data were filtered with the following parameters: QD < 2.0 || FS > 60.0 || SOR > 3.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0, and biallelic sites were retained.

Principal component analysis based on the total SNPs was conducted with default settings using the GCTA software.[102] The SNPs at fourfold degenerate sites were used to construct a maximum likelihood phylogenetic tree using the RAxML software and a neighbor-joining phylogenetic tree using the PHYLIP software[103] for peach population (Data S5), Mei accessions (Data S2), and apricot accessions (Data S3).

Identification of deleterious mutations in Tibetan and cultivated peach accessions

Genomic sequences of the Tibetan peach genome were identified by quantifying rejected substitution, which is a natural measure of constraint reflecting the strength of past purifying selection on an element. Comparative genomic analysis was conducted for the 11 species in the Fabids clade, including Tibetan peach, cultivated peach, apple,[119] strawberry[44] (*Fragaria vesca*), jujube[120] (*Ziziphus jujuba*), marijuana[121] (*Cannabis sativa*), cucumber[122] (*Cucumis sativus*), soybean[123] (*Glycine max*), common bean[124] (*Phaseolus vulgaris*), alfalfa[125] (*Medicago truncatula*), and red clover[126] (*Trifolium pratense*). Ten runs of pairwise alignments were performed by aligning particular genome sequences to the Tibetan peach genome sequence using the LASTZ software. An 11-way multiple sequence alignment was generated using the roast command in the MULTIZ software.[104] A phylogenetic model for the 11 species was constructed from fourfold degenerate sites using the phyFit program in the PHAST software.[105] Based on the multiple sequence

alignments and the estimated phylogenetic model, *P value*s using the likelihood ratio test for conservation or acceleration (conservation scores) of individual sites were computed using the phyloP program from the PHAST software. Based on the results of phyloP, genomic sites of Tibetan peach with a score above zero were defined as conserved.

The sequencing data from 15 *P. dulcis* accessions were used as an outgroup to predict the ancestral and derived alleles at the polymorphic sites in the Tibetan and cultivated peach accessions. For each biallelic site, the allelic state was defined as ancestral if 14 to 15 *P. dulcis* accessions had the same homozygous genotype (namely, homozygous reference or homozygous alternative). For each Tibetan and cultivated peach accession, the variant effect of each SNP was predicted using the SnpEff software.[106] The PROVEAN software[107] was used to predict the functional effect of nonsynonymous amino acid substitutions against the nr database at NCBI. For the genomic sites of Tibetan peach that are conserved in multiple species, an amino acid substitution was predicted to be deleterious if the score was < -2.5 and tolerated if the score was ≥ -2.5.

The numbers of derived deleterious alleles present in each of the Tibetan and cultivated peach accessions were counted to identify the homozygous deleterious mutations, the heterozygous deleterious mutations, and the total deleterious mutations. The homozygous deleterious mutations were estimated as the number of derived deleterious alleles existing in the homozygous state. The heterozygous deleterious mutations were estimated as the number of derived deleterious alleles existing in the heterozygous state. The total deleterious mutations were estimated as the number of derived deleterious alleles existing in an accession (2 × homozygous deleterious mutations + heterozygous deleterious mutations).

Detection of selection signatures for high-altitude adaptation

The F_{ST} and XP-EHH statistics were employed to identify genomic differentiation at the single-nucleotide and haplotype levels between the 66 Tibetan peaches from the relatively high altitudes ($\geq$ 3,800 m a.s.l.) and the 67 Tibetan peaches from the relatively low altitudes ($\leq$3,200 m a.s.l.). A sliding-window approach (50-kb windows sliding in 10-kb steps) was applied to the genome-wide SNPs to quantify genetic differentiation (F_{ST}) using the VCFtools software.[108] The haplotype phasing of the high-altitude accessions and low-altitude accessions was conducted separately using the fastPHASE program[109] and the extended haplotype homozygosity across populations (XP-EHH) values, which were calculated between high-altitude and low-altitude populations using the REHH package[110] in R. In addition, the haplotype differentiation between the two populations (hapFLK) was calculated using the hapFLK program[111] with 15 clusters and 10 EM runs.

To search for footprints of selection, regions with significantly high F_{ST} values (in the 5% right tail of the empirical distribution of F_{ST} values) and significantly different XP-EHH values (in the 5% right tail and the 5% left tail of the empirical distribution of XP-EHH values) were considered to be under selection.

Metabolomics profiling and analyses

All chemicals used in this study were of analytical reagent grade. Gradient-grade methanol, acetonitrile and acetic acid were purchased from Merck Company, Germany. The water used in this study was deionized twice using a Milli-Q water purification system (Millipore, Bedford, MA). Standards were purchased from ANPEL (Shanghai, China), BioBioPha Co., Ltd., and Sigma-Aldrich (USA).

The freeze-dried fruits were powdered using a mixer mill (MM 400, Retsch) containing zirconia beads for 1.5 min at 30 Hz. One hundred mg of powder was weighed and extracted overnight at 4°C with 1.0 mL of 70% aqueous methanol. Following centrifugation at 10,000 × g for 10 min, the extracts were filtered (SCAA-104, 0.22 μm pore size; ANPEL, Shanghai, China) before LC-MS analysis.

The sample extracts were analyzed using an LC-ESI-MS/MS system (HPLC, Shim-pack UFLC SHIMADZU CBM30A system; MS, Applied Biosystems 6500 Q TRAP). The analytical conditions were as follows: HPLC: column, Waters ACQUITY UPLC HSS T3 C18 (1.8 μm, 2.1 mm × 100 mm); solvent system, water (0.04% acetic acid): acetonitrile (0.04% acetic acid); gradient program, 100:0 V/V at 0 min, 5:95 V/V at 10.0 min, 5:95 V/V at 11.0 min, 95:5 V/V at 11.1 min, 95:5 V/V at 15.0 min; flow rate, 0.4 mL/min; temperature, 40°C; and injection volume, 5 μl. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS system. LIT and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP) using an API 6500 Q TRAP LC/MS/MS system equipped with an ESI Turbo Ion-Spray interface operated in positive ion mode and controlled by the Analyst 1.6.2 software (ABSciex).

A total of 1,768 metabolites with high repeatability were used in our study. The mean values obtained for the two biological replicates were log-transformed before subsequent analysis. Based on the levels of 1,768 metabolites in 275 Tibetan peaches, principal component analysis was separately performed with default parameters using FactoMineR[112] and factoextra (http://www.sthda.com/english/rpkgs/factoextra) in R (Data S1AB).

mGWAS analysis

A total of 275 Tibetan peach accessions were used in the mGWAS. The SNPs with minor allele frequencies above 0.05 and maximum missing rates below 0.1 were filtered using the VCFtools software. A kinship matrix was generated with the A.mat function from the rrBLUP package.[113] The first two principal components from the PCA of 275 accessions were used as covariances to account for the population structure in the mGWAS. The LMM model was used to perform the association analysis using the FaST-LMM software.[114] The genome-wide significance threshold was determined using a Bonferroni correction to adjust all 510,989 SNPs. In addition to the metabolite based GWAS, five environmental factors (altitude, annual mean air pressure, annual average temperature, annual sunshine h, and annual average water vapor pressure) were used as phenotypes to conduct the association analyses.

High-altitude adaptation-related metabolites in the Tibetan peach population

To partition the variation of each metabolite into contributions from genetic and environmental sources, we estimated the polygenic score for each accession using the state matrix as the covariance structure and the rrBLUP package[113] in R. For each metabolite, the regression fit between the breeding values and the altitude was performed. The *P values* for the permutation tests were calculated using the lmp function in the lmPerm package (https://github.com/mtorchiano/lmPerm) of R. Metabolites with adjusted r^2 values greater than 0.25 and *P values* less than 0.05 were retained.

Next, for each metabolite, the narrow-sense heritability (h^2) was calculated as the proportion of the genetic variance of the total phenotypic variance (i.e., the sum of the genetic variance and the error variance). Metabolites that accumulated in the Tibetan peach population with h^2 values greater than 0.1 were retained.

For each metabolite, a Spearman's rank correlation coefficient between altitude and metabolite content was calculated.

Identification of Tibetan peach-specific SINEs

Candidate Tibetan peach-specific SINEs were preliminary identified from the Tibetan peach-specific regions by genomic comparion of Tibetan and cultivated peach. Then the candidate SINEs were confirmed by presence or absence analysis between the Tibetan peach population and the cultivated peach population using genome re-sequencing data. A SINE element was considered as present in an individual if the coverage of the region was greater than 90% and the average depth of the coverage was greater than 5 × . A SINE element was considered as absent in an individual if the coverage of the region was less than 40% and the average depth of the coverage was less than 5 × . Then, at the population level, a SINE was defined as Tibetan peach-specific if the SINE element was present in greater than 90% of the Tibetan peach accessions and absent from 90% of the cultivated peach accessions.

Experimental validation for SINE insertions

The primer pairs for validation of the insertion of SINE1 (chr3: 16623756-16624715), SINE2 (chr3: 16623030-16623781), and SINE3 (chr3: 16582770-16584091) were separately designed (Data S1AO). PCR-based experiments were conducted in 50-μl reaction volumes containing 25 μl of 2 × Phanta Max Buffer (Vazyme Biotech), 1 μl of dNTP Mix (10 mM each) (Vazyme Biotech), 1 μl of Phanta Max Super-Fidelity DNA Polymerase (Vazyme Biotech), 100 ng of genome template, and 2 μl of each forward and reverse primer. The amplification was conducted at 95°C for 5 min followed by 34 cycles of 15 s at 95°C, 30 s at 58°C, 58°C and 56°C for SINE1, SINE2 and SINE3, respectively, and 40 s, 120 s and 75 s at 72°C for SINE1, SINE2 and SINE3, respectively, and finally 5 min at 72°C. Agarose gel electrophoresis was performed to separate the PCR products, and Sanger sequencing of longer and shorter PCR products with target length was performed to confirm the SINE insertions. After confirmation, agarose gel electrophoresis of the PCR product was conducted to investigate whether SINE insertions occurred in the tested peach accessions.

The neighbor joining (NJ) tree of the candidate gene *Pmira3g006670* was constructed with the amino acid sequences of 96 *NAC* genes from *Arabidopsis thaliana* and *Pmira3g006670* with 1,000 bootstraps using the PHYLIP software (Data S6).

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of the statistics applied are provided in the figures and the corresponding legends. Statistical analyses were performed in R 3.6.2.[127] We used the Fisher's exact test to conduct the Gene Ontology enrichment analysis of the target genes relative to the background of the entire genome using the agriGO program.[96] The *P values* from the tests mentioned above were adjusted using the FDR correction (BH method) for multiple testing. The filtering criteria were $p < 0.05$ and $FDR < 0.05$. Statistically significant differences were determined using the Student's t test. Correlations were quantified by calculating Spearman's rank correlation coefficients.